Proceedings of the
25th European Conference on e-Learning
Hosted by Lund University School of Economics and Management, Lund University, Sweden
22-23 October 2026



Edited by
Nam Aghaee and Paul Pierce

A conference managed by ACPI, UK

Proceedings of the
25th European Conference on
e-Learning

ECEL 2026

22-23 October 2026

Hosted By

Lund University School of Economics and Management, Lund, Sweden

Edited by

Nam Aghaee and Paul Pierce

Copyright the authors, 2026. All Rights Reserved.

No reproduction, copy or transmission may be made without written permission from the individual authors.

Review Process
Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

Ethics and Publication Malpractice Policy
ACIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here:
http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/

Self-Archiving and Paper Repositories
We actively encourage authors of papers in ACIL conference proceedings and journals to upload their published papers to university repositories and research bodies such as ResearchGate and Academic.edu. Full reference to the original publication should be provided.

Conference Proceedings
The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including the Web of Science and Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

Since 2022 all ACIL conference proceedings are open access and freely available to read from www.proceedings.com

The Conference Proceedings for this year and previous years can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-917204-99-6
E-Book ISSN: 2048-8645
Print version ISBN: 978-1-918691-00-9
Print Version ISSN: 2048-8637

Published by Academic Conferences and Publishing International Limited
Reading, UK
+44 (0) 118 324 6938
www.academic-conferences.org
info@academic-conferences.org

Contents

Preface ix

Committee x

Biographies xii

Academic Papers

Fostering Self-Direction and Peer Feedback Skills in a Management Course
Dina Adinda, Caroline R O'Neill, Elise Goiseau, Thomas Petit, Christophe Jeunesse 1-9

Alternative Higher Education Assessment Framework in Generative AI Era: Project Proposal and Pitch Approach
Nam Aghaee, Daniel Lindvall , Christina Keller 10-19

Pre-Service Teachers' Perceptions of Artificial Intelligence within the UTAUT Model
Dilara Arzugül Aksoy, Elif Polat 20-27

From Potential to Practice: Challenges of Integrating Artificial Intelligence Technologies in School Classrooms
Norah Almusharraf, Muhammad Imran 28-36

Beyond AI Use: Understanding University Students' Acceptance of AI-Assisted Academic Misuse
Basil Alzougool, Sumayya Banna 37-45

A Scalable, Work-Integrated E-Learning Model for Rapid Upskilling in Biomanufacturing
Ana Isabel Aranda Perez, Sara Vernocchi, Timo Tekhaus, Matthias Afting 46-57

Integrating ChatGPT into University English Teaching: Linguistic Performance, Autonomy, and Learning Experience
Jeany Argueta, Gisela Espinoza, Víctor Sandoval 58-66

Personalised Formative Assessment in Ukrainian Higher Education: Design and Empirical Evaluation of the SMART-PL Model
Artem Boyarchuk, Daria Shteinbrekher 67-76

Integrating AI-Generated 3D/360° Content into VR-Based Secondary Education
Vasile-Ioan-Grigore Avram, Ovidiu Cosma, Nadia Barkoczi 77-88

Cognitive Barriers in Solving Context-Based Geometry Tasks in an LMS Moodle Environment
Katerina Biernatova, Angelika Schmid, Lilla Korenova 89-98

Structured over Chat-Based Feedback: Comparing User Preferences on Different Presentation Feedback Reports
Noemi Giulia Bontempo, Nina Mouhammad, Daniele Di Mitri 99-108

From Action to Judgement: Experiential Learning Processes in a Business Simulation Environment
Yvonne Costin, Michael P. O'Brien , William Hogan 109-118

Technology-Enhanced Peer Assessment via FeedbackFruits: A Multi-Framework Evaluation of Digital Technology Integration in Pedagogical Design
Xiangping Du, Bobby Pothabathula 119-128

Beyond Audio: An Empirical Assessment of AI-Generated Podcasts and Videocasts in Business Education
Gökhan Eğilmez, Karolina Schneider 129-139

Artificial Intelligence as a Mediating Tool in Higher Education: An Exploratory Study in Quantitative Methods
Maria Teresa Garcia Florentino 140-147

Effect of Open and Distance Learning Programmes on Inclusivity in Higher Education
Vishaka Nanayakkara, Savin Gunawardena, Buddhika Karunarathne, Dhanushima Gamage 148-157

An AI-Supported Simulation-Based Serious Game in Digital Health Management Education: Design and Formative Evaluation
Annika Hering, Gülsen Otman, Annika Viedenz, Jan-David Liebe 158-169

Digital Serious Games in Healthcare Management: Comparative Analysis and Design Recommendations
Tobias Höllmer, Annika Hering, Gülsen Otman, Uwe Buddrus, Jan-David Liebe 170-180

Inclusive Immersive E Learning: Participatory Design of Accessible Virtual Reality: Supporting Diverse Learners Through Inclusive Medical Simulation Design
Tina Holmes, Teresa Crea, Xueqing Lu, Graham Hannah 181-188

Student Ambassadors for Sustainable AI EdTech Knowledge Sharing in Higher Education
Hyein Cho, Min Soo Lee, Young Woong Ko, Yong Geun Kim, Jion Kim, Minho Kim, Tae Il Yoon 189-197

Pedagogical Innovation through Generative AI: Benefits, Challenges, and Responsible Use in Higher Education
Muhammad Imran, Norah Almusharraf 198-205

Designing Meaningful AI Learning Activities in Pre-Service Teacher Education
Antonin Jancarik, Kateřina Jančaříková, Martina Šmejkalová 206-213

Pedagogically Asymmetric Elo Models for Multidimensional Adaptive Learning Systems
David Janda, Antonin Jancarik, Jakub Michal, Tomáš Jeřábek 214-224

The Impact of Content Design on the Effectiveness of an Educational Application
Tomas Jerabek, Jakub Michal, David Janda 225-232

Generative Artificial Intelligence Use Among Students: A Cross-National Comparison Between Türkiye and England
Merve Kara, Zeynep Cömert, Kübra Kaymakçı Üstüner 233-241

Digital Transformation Where It Matters: A Four Pillar Maturity Model for Diverse Assessment Practices in Higher Education
Thashmee Karunaratne, Fredrik Enoksson, Richard Lee Davis, Anna Land , Caroline Bern 242-251

Ruwad Al-Mustaqbal System: A Knowledge-Behavior Gap-Aware Framework (KBG-Ruwad) for Personalized Responsible AI Education among Omani Undergraduates
ArunaDevi Karuppasamy, Samah Alzidi, Aseel Al Shizawi 252-259

Cracking the Code: Online Education for Modern Leadership
Amna Khaliq 260-268

Reflex, Refraction and Reflection: A Prism for Dialogue
Thomas Kjaergaard 269-277

Societal Transformation Related to Emerging Technologies in Industry Through Bi-Lateral Collaboration
Marisol Kokko Rico, Mikael Ericsson, Prakruthi Hareesh, Pb Venkataraman, Kerstin Johansen, Lennart Malmskold 278-284

A Redesigned Peer Learning Model for Fair Assessment in Distance Education
Vishaka Nanayakkara, Tharushi Hansana Kurukulasuriya, Buddhika Karunarathne, Jageshkar Sivahar 285-295

Prompt-as-Lesson: Authoring AI-Delivered Lessons for Fidelity and Observability
David Lefevre 296-304

Beyond the Standalone AI: Embedding Educational Chatbots through Scaffolded Pedagogical Design
Colin Loughlin 305-310

From Technology to Teaching Practice: A Didactic Roadmap for Emerging Tech in Education
Gunver Majgaard, Regina Lamscheck-Nielsen, , Maria Skougaard Andersen, Rasmus Ploug 311-319

The Level of Digital Skills of Mining Employees: Their Learning Needs and Interest in Personal Development
Josef Malach, Petr Pyszko, Kateřina Malachová, Dana Vicherková 320-329

Towards a Multidimensional Framework for Educational Data Literacy
Flavio Manganello, Giannangelo Boccuzzi, Giancarlo Masi, Chiara Fante 330-337

Digital Learning Environments and Student Satisfaction: A Learning Analytics Perspective
Sarmīte Mickeviča, Signe Bāliņa 338-346

SMAP-ST: Smart MBTI-Adaptive Public Speaking Tutor
Leonardo Mogianesi, Stelios Kapetanakis, Giacomo Nalli 347-353

From "AI is Supi" to "Democracy is in Danger": Users' Perceptions About AI
Dagmar Monett 354-362

AI and Academic Integrity: Student Perceptions of Plagiarism and Ethical Use at the University of Fort Hare
Bongo Mqukuse, Nathunathi Mvunge 363-370

From Cognitive Overload to Individualized Learning: AI Tutoring in Electrical Engineering
Jens Müller, Annalena Meier, Matthias Heinz, Kilian Göller, Alexander Alfred Zyla 371-380

Artificial Intelligence in Teaching and Learning: Opportunities, Tensions, and Transformation in Higher Education.
Nathunathi Mvunge, Bongo Mqukuse 381-388

Co-Designing AI Guidelines for Learners: Contexts, Principles, and Literacies
David Parsons, Kathryn MacCallum, Stephen Powell 389-397

The Role of Academic Libraries in E-Learning Ecosystems in Higher Education: A Narrative Review
Edmont Pasipamire 398-406

*Towards Human-Centered Learning: AI-Supported Skill Management in a Large Automotive Organization*Svenja Pieczurczyk, Paul Drews 407-414

Curriculum Gap Revisited: LLMs, Adaptive Expertise and PBL for IS
Paul Pierce 415-421

Active Learning Map: A Novel Assimilation Guided Adaptive Learning Platform
Suhan Roy, Jupalli Jaswant Prabhas, Praseeda, Tulika Saha, Prasad Ram, Sushree Behera, Srinath Srinivasa 422-429

Spatial Reasoning Outcomes in Technology-Enhanced Geometry Teacher Education
Angelika Schmid, Tomas Barot, Lilla Korenova, Zsolt Lavicza 430-439

Integrating Generative AI in Assessment and Feedback in Higher Education in Albania
Xhensila Sinanaj, Ervis Sanço, Erion Selimi, Doris Madhi, Elvira Fetahu 440-446

The Prisoner's Dilemma of AI: Cognitive Offloading Among Future Mathematics Teachers
Otto Suchanek, Tomas Javorcik, Tatiana Havlaskova, Magdalena Zavodna 447-455

The Use of Generative AI in Higher Education Student Assessments: A Synthesis of Empirical Studies
Chun Meng Tang 456-463

Modality Matters for Extraversion: A Big Five Meta-Regression in Online Learning Environments
Eisuke Tokiwa 464-471

Design of an Educational Platform for Developing Librarians' Digital Competencies
Hana Ulrichova, Michaela Mrazova, Magdalena Zavodna, Tomas Javorcik 472-482

Transforming Teaching Through Blackboard Ultra Adoption: Lecturer Training, Pedagogical Integration and Sustainable LMS use in Higher Education
Elzette van Niekerk, Adriana Botha 483-493

Factors of non-Fiction Texts Comprehension Among Primary School Pupils in the Czech Republic
Dana Vicherková, Josef Malach, Kateřina Malachová; Martin Kolář;
Veronika Murinová 494-502

AI, Instructor, Observer, and Peer Feedback in Online Microteaching: A Multi-Rater Comparison Study
Veronika Vrábelová, Zuzana Sándorová, Zdena Kráľová, Elena Kováčiková 503-511

How Gaming Activities Shape Lower Secondary School Pupils' Understanding of Artificial Intelligence
Magdalena Zavodna, Hana Ulrichova, Katerina Kostolanyova 512-520

Toward Intersectionally Fair Educational AI for Learning Analytics and Student Guidance
Stefania Zourlidou, Quy Tai Le, Resmin Hossain, Frank Hopfgartner 521-528

Embedding Responsible AI in a Technical Machine Learning Course
Stefania Zourlidou, Frank Hopfgartner 529-536

The Effectiveness of Developing and Optimizing a Self-Learning Course to Promote Chinese Learning
Felix Chao 537-548

PhD Papers
Shifting the Time of Practice: A Digital Environment for Student-Athletes and Remediation
Paul Ronald Klensch, Yves Kreis 551-559

Beyond Engagement: A Human Flourishing Framework for Emotion-Aware AI in Education
Spyridon Kontis, Sofia Anastasiadou 560-567

Digitally Supported Formative Assessment and Achievement Motivation: Evidence from the REFRESH Project
Veronika Murinová, Dana Vicherková 568-577

Work in Progress papers
Challenges and Opportunities for Applying Artificial Intelligence to Learning in Secondary Schools
Ioannis Lappas 579-581

Late Submissions
Beyond the Digital Journal: Epistemic Conversations in Work- Integrated Learning
Ruben Daniels 583-588

PGCE Students' Perceptions of Generative Artificial Intelligence within Learning Ecosystems at a South African Teacher Education Institution
Tunjera Nyarai, Chigona Agnes 589-598

About the 25th European Conference on e-Learning, 2026

The European Conference on e-Learning (ECEL) is a prominent academic conference that serves as a platform for researchers, educators, and practitioners in the field of e-learning to share and discuss their findings and insights. With a history spanning 25 years, the conference has established itself as a vital forum for exploring advancements and innovations in e-learning.

The Proceedings of the 25th European Conference on e-Learning, 2026 includes academic research papers, PhD research papers, Master's Research papers and work-in-progress papers, which have been presented and discussed at the ECEL 2026 conference. The proceedings are of an academic level appropriate to a professional research audience including graduates, post graduates, doctoral and post-doctoral researchers. All papers have been double-blind peer reviewed by members of the Review Committee.

Aims and Scope

The primary aim of ECEL is to foster collaboration and knowledge exchange among those involved in e-learning. It seeks to advance the field by providing a space where latest research, technologies, and pedagogical strategies can be presented. The conference encourages participation from a diverse group of attendees, including academics, industry experts, and policy-makers, reflecting the multifaceted nature of e-learning.

Topics Covered

The call for papers for the ECEL conference asked for contributions that considered the following topic. In addition, the committee welcomed papers on a number of specialist mini-tracks which can be seen at the end of this list.

- **Educational Technologies**: Innovations in digital tools and platforms designed to enhance learning experiences.
- **AI in Education**: Ethical and practical uses of AI in and for education as well as preparing students for the world with AI and cases showing successes and challenges in this area.
- **Blended Learning:** Strategies for effective teaching and learning in online and hybrid environments.
- **Distance Education**: Best practices and challenges in implementing distance learning programs.
- **Learning Analytics**: Use of data to improve educational outcomes and personalize learning experiences.
- **Inclusive Education**: Addressing accessibility and inclusivity in e-learning environments.
- **Mobile Learning:** Exploring the impact of mobile technology on education.
- **Game-Based Learning:** Integration of games and gamification in educational contexts to encourage engagement and motivation.

Mini Tracks

Experts in the field proposed mini tracks on the following topics. Papers accepted on these topics after the double-blind peer review process were presented as mini tracks at the conference and are also published in these proceedings.

- Work-Integrated e-Learning
- New Constellations of PBL in the LLM Era: Liquid Learning, Transfer, and Adaptive Expertise
- Artificial Intelligence in E-Learning: Uses, Implications and Responsible Practice

The programme from the 2026 conference can be downloaded from here and full details about the conference can be seen from the conference website: https://www.academic-conferences.org/conferences/ecel/

ECEL Preface

These proceedings represent the work of contributors to the 25th European Conference on e-Learning (ECEL 2026), hosted by ACPI and the Lund University School of Economics and Management, Lund, Sweden on 22-23 October 2026. The Conference Chair is Dr. Nam Aghaee from the Lund University School of Economics and Management, Lund Sweden and the Programme Chair Dr. Paul Pierce from the Lund University School of Economics and Management, Lund Sweden. ECEL is now a well-established event on the academic research calendar and in its 25th year the key aim remains the opportunity for participants to share ideas and meet the people who hold them. The scope of papers will ensure an interesting two days. The subjects covered illustrate the wide range of topics that fall into this important and ever-growing area of research.

The opening keynote presentation is given by Prof. Stefan Hrastinski from Royal Institute of Technology, Sweden who will discuss *Imagining Hopeful Futures for Education*. We will also have a keynote speech from Md Saifuddin Khalid from (DTU Compute), Technical University of Denmark, Lyngby, Denmark who will talk on the topic of *Reshaping Assessment: How Universities Navigate Generative AI, Regulation, and Innovation.* The second day of the conference will begin with an address by Prof. Amin Jalali from Stockholm University, Sweden who will discuss *Don't Underestimate your LMS Data: Making Invisible Learning Journeys Visible.*

With an initial submission of 146 abstracts, after the double blind, peer review process there are 68 Academic research papers, 3 PhD research papers and 1 work-in-progress papers published in these Conference Proceedings. These papers represent research from, Albania, Australia, Bulgaria, Canada, China, Cyprus, Czech Republic, Denmark, Egypt, El Salvador, Estonia, Finland, France, Germany, Greece, Hungary, India, Ireland, Italy, Japan, Korea, Kuwait, Latvia, Luxembourg, Malta, New Zealand, Norway, Oman, Portugal, Republic of Korea, Romania, Saudi Arabia, Singapore, Slovakia, South Africa, South Korea, Sri Lanka, Sweden, Tanzania, Thailand, Türkiye, UK and USA.

We hope you enjoy the conference.

Dr Nam Aghaee
Lund University School of Economics and Management (LUSEM)

Biographies of Chairs and Keynote Speakers

Conference Chair



Nam Aghaee is a Senior Lecturer in Informatics at Lund University School of Economics and Management in Sweden, specializing in digital learning technologies, user experience design within education and eHealth, and digital transformation. She received her PhD in Computer and Systems Sciences from Stockholm University in 2015, with a focus on Interaction Design and Learning. Dr. Aghaee teaches courses in digital innovation and transformation, designing digitalization, and managing information system; she supervises thesis projects in information system, educational technology, innovation in education and healthcare, and system design. Her current research work frequently examines how digitalization and AI tools can enhance educational settings. She has contributed to projects that integrate design-oriented perspectives, gamification, AI-supported project-based learning, and AI-resilient assessment. She plays an active role in shaping educational initiatives related to digitalization; also contributes to a range of departmental responsibilities at the Department of Informatics.

Programme Chair



Paul Pierce is a Senior Lecturer in Informatics at Lund University School of Economics and Management in Sweden. Dr. Pierce received his PhD in Information Systems in 2013, concentrating on how alliances can enhance organizational ICT capabilities. His research examines digitalization and sustainability, with a particular emphasis on Smart Cities, healthcare systems, and supply chain transparency. He has participated in international and interdisciplinary collaborations that connect technological and organizational viewpoints. At Lund University, he leads the Master's Programme in Information Systems and frequently serves as a track chair at academic conferences. His work integrates research, teaching, and industry engagement to explore how digital technologies shape societal and organizational transformation.

Keynote Speakers



Stefan Hrastinski is Professor, Deputy Head of Department for Research, and Director of Research Education at the Department of Learning at KTH Royal Institute of Technology, Sweden. He leads the Digital Learning research group. His research interests include the design and evaluation of emerging methods and technologies for learning, online collaborative learning, human and artificial tutoring, and education futures.



Md Saifuddin Khalid is Associate Professor at the Department of Applied Mathematics and Computer Science (DTU Compute), Technical University of Denmark, where he leads LearnT - Centre for Digital Learning Technology and teaches introductory statistics and digital learning technology and entrepreneurship. His research concerns the development, integration and adoption of digital systems in education and training, with a current focus on the responsible use of generative AI in assessment and feedback. He coordinates the Erasmus+ KA2 project HEGenAI and contributes to MixAP-EU, and he has previously led or partnered in Danish Innovation Fund's NURSEED, Erasmus+ EdViCon, MeCaMind, \& BLTeae and three Danish DIREC projects. He holds a PhD in Human-Centered Communication and Informatics from Aalborg

University and was elected Teacher of the Year by the students of AAU's Faculty of Humanities in 2016. He has published over 70 peer-reviewed papers, supervised 70+ Master theses, two PhD students as principal supervisor and two as co-supervisor



Amin Jalali is an Associate Professor of Computer and Systems Sciences at Stockholm University, specializing in the development of IT systems for business process modeling, analysis, and management. His research emphasizes business process analysis through model-based and data-driven techniques, with particular focus on process simulation, process mining, event knowledge graphs, and object-centric process mining. He has led numerous research projects across healthcare, education, and finance, bridging academic and practical insights. Alongside his research, he serves as a board member of the department and manages three graduate courses: Business Process Design and Intelligence, Business Process and Case Management, and Data Warehousing. He also brings extensive industry experience in designing and implementing Business Intelligence and Big Data Analytics solutions.

Mini Track Chairs



Dr. Prakruthi Hareesh serves as the Smart Manufacturing Domain Lead at BITS Pilani, Work Integrated Learning Programmes (WILP). She holds a Ph.D. in Mechanical Engineering from the University of Maryland, College Park, USA, with expertise in Mechatronics and MEMS transducer development. With over a decade of academic and industry experience, she has led more than 50 turnkey projects in the semiconductor and manufacturing sectors. At BITS Pilani, she leads Smart and Digital Manufacturing initiatives, developing industry-relevant curriculum with cutting edge experiential learning modules and learning factories in emerging areas like Cyberphysical systems and Digital twins, machine vision, autonomous and collaborative robots and extended reality. Her work bridges applied research and industrial practice, fostering innovation and workforce transformation in next-generation manufacturing.



Prof. Dr. Abbas Fadhil Aljuboori is working currently at Gulf College, Department of Computing Sciences, Muscat, Sultanate of Oman. He is also worked as a Deputy Dean for Scientific Affairs at AlShaab University, Baghdad, Iraq. He has a Ph.D. in Computer Science from Dongguk University, South Korea. Fulbright Visiting scholar – University of Central Oklahoma – Edmond – USA. International Advisory Board Member for CT University, India. He worked as a Vice President for Administrative Affairs and Head of Smart Cities Center at University of Information Technology and Communications, Baghdad, Iraq. Faculty Staff Member at Al Zahra College for Women, Muscat, Oman. Researcher and Manager in the Advanced Institute of Convergence Information Technology (AICIT) – South Korea, Head of Computer Science Department – University of Kerbala. Vice President of Iraqi Universities Accreditation and Quality Assurance Council for Computer Science and IT. His field of Interest are in Data Mining, Big Data, Artificial Intelligence, Cyber Security, Information Systems, Social Media Analysis, and Smart Applications. He is a Member of several of Academic and Professional Societies. He is an Editor-in-Chief, Editorial Board Member and Reviewer of many eminent International Journals and Conferences worldwide.

ECEL Review Committee

ECEL is fortunate to have a significant number of international experts in the field willing to review papers. Care is taken to ensure that a fully double-blind peer review process is followed, and reviewers are not selected if they have any research or organisational connection to the author/s.

Prof Sara A. Abdoh, Faculty of Applied Arts, Benha University, Egypt, Egypt, Dr Adrian Adascalitei, Technical University Gh. Asachi Iasi, Romania, Dr Dina Adinda, Paris Nanterre University, France, Prof Megha Ainapurkar, Padre Concicao College of engineering Goa, India, Prof Hamid Alasadi, Iraq University College, Iraq, Dr Ali Alawneh, Philadelphia University, Jordan, Prof Susana Aldeia, Universidade Portucalense, Portugal, Prof Abbas Aljuboori, University of Information Technology and Communications, Iraq, Prof Stephen Alstrup, Department of Computer Science, University of Copenhagen, Denmark, Prof Lionel Alvarez, University of Fribourg, Switzerland, Paulo Alexandre Vara Alves, Instituto Politécnico de Bragança, Bragança, Portugal, Dr Ammar Alwaheidi, Al-Quds University @ UNRWA, Palestine, Dr Andronic Anca-Olga, Spiru Haret University, România, René Holm Andersen, University College of Northern Denmark, Aalborg, Denmark, Prof Antonios Andreatos, Hellenic Air Force Academy, Greece, Prof Mohamed Arteimi, Libyan Academy of Graduate Studies, Libya, Marc Augier, SKEMA Business School, France, Mr. Utkarsh Avalekar, D. Y. Patil Agriculture & Technical University, Talsande, Kolhapur, MH, India, Prof Lurdes Babo, Porto Accounting and Business School-Polytechnic of Porto, Portugal, Prof Liz Bacon, University of Abertay, Dundee, UK, Prof Alina Badulescu, University of Oradea, Romania, Dr Nimalathasan Balasundaram, University of Jaffna, Sri Lanka, Dr S.R. Balasundaram, NIT TIRUCHIRAPPALLI, India, Dr Joan Ballantine, University of Ulster, UK, Prof Gezani Baloyi, University of South Africa, South Africa, Dr Wendy Barber, University of Ontario Institute of Technology, Canada, Dr Trevor Barker, University of Hertfordshire, UK, Assc Fatima Zohra Belkhir, University of Abou-Bekr Belkaid, Tlemcen, Algeria, Andrea Benn, University of Brighton, UK, Dr Igor Bernik, University of Maribor, Slovenia, Daniel Biella, University of Duisburg-Essen, Germany, Prof Dr Mirela Blaga, Technical University "Gheorghe Asachi", Romania, Dr Ivana Bogdanovic, University of Novi Sad, Serbia, Dr Tharrenos Bratitsis, University of Western Macedonia, Greece, Mie Buhl, Aalborg University Copenhagen, Netherlands, Prof Carsten Busch, University of Applied Sciences HTW Berlin, Germany, Prof Kiymet Caliyurt, Trakya University, Turkey, Dr Sancha Campanella, Instituto Superior de Administração e Línguas, Portugal, Prof Piedade Carvalho, ISEP, Porto, Dr Ivana Cechova, University of Defence, Czech Republic, Maria Celentano, University of Lecce, Italy, Felix Chao, The Chinese University of Hong Kong, New Territories, Hong Kong, Prof Phaik Kin Cheah, Universiti Tunku Abdul Rahman, Malaysia, Satyadhyan Chickerur, B V Bhoomaraddi College of Engineering and Technology, Hubli, India, Dr Goh Chin Fei, Universiti Teknologi Malaysia, Malaysia, Prof Jyotie Choudrie, University of Hertfordshire, UK, Dr Lucian Ciolan, University of Bucharest, Romania, David Comiskey, University of Ulster, Northern Ireland, Dr Eduardo Correia, Christchurch Polytechnic Institute of Technology, New Zealand, Prof Filipe Couto, Escola Superior de Educação de Fafe - ESEF-IEES, Portugal, Dr Caroline Crawford, University of Houston-Clear Lake, USA, Dr Sara Cruz, Polytechnic Institute of Cávado and Ave, Portugal, Dr Sara Cruz, Polytechnic University of Cávado and Ave, Portugal, Ken Currie, CAPDM Ltd, UK, Dr Fisnik Dalipi, Linnaeus University, Sweden, Dr Susanne Dau, University College of Northern Denmark, Denmark, Dr Antonio de Castroavc, Instituto Superior de Engenharia do Porto (ISEP), Portugal, Antonio De Nicola, ENEA, Italy, Prof Dr Carmen De Pablos Heredero, Rey Juan Carlos University, Spain, Dr Souad Demigha, UNIV Paris 1 and Paris11, France, Dr Faiza Derbel, University of Manouba, Tunisia, Dr Varsha Desai, D Y Patil Agriculture and Technical University, Talsande, Kolhapur,

India, Dr Rajiv Dharaskar, GH Raisoni College of Engineering, Nagpur, India, Dr Martina Doolan, University of Hertfordshire, UK, Dr Yanqing Duan, University of Luton, UK, Dr Palitha Edirisingha, University of Leicester, UK, Dr Colin Egan, University of Hertfordshire, Hatfield, UK, Gokhan Egilmez, Lindenwood University, USA, Dr Ramadan Elaiess, University of Benghazi, Libya, ramakanth evani, Independent Technoly Researcher, US, Khristin Fabian, Perth College UHI, UK, Dr Samson Fabiyi, University of Leeds, UK, Dr Kendil Fatima, Université oran2, Algeria, Prof Gert Faustmann, Berlin School of Economics and Law, Germany, Prof Manuela Fernandes, Polithechnic Institute of Tomar, Portugal, Dr Rachel Fitzgerald, University of Queensland, Australia, Dr Panagiotis Fotaris, Brighton University, UK, Dr Michelle French, University of Toronto, Canada, Dr Václav Friedrich, VSB - Technical University of Ostrava, Czech Republic, Dr Vilas Gaikar, Smt. CHM. College, University of Mumbai, India, India, Thomas Goetz, Hokusei Gakuen University, Sapporo, Japan, Prof Maria de Fátima Goulão, Universidade Aberta, Portugal, Dr David Guralnick, Columbia University and Kaleidoscope Learning, New York, USA, Dr Prakruthi Hareesh, Birla Institute of Technology and Science, Pilani, India, Muhammad Said Hasibuan, Institute Business and Informatics Darmajaya, Indonesia, Anita Hiralaal, Durban University of Technology, South Africa, Dr Md. Fokhray Hossain, Daffodil International University (DIU), Bangladesh, Dr Maggie Hutchings, Bournemouth University, UK, Dr Maria Impedovo, Aix-Marseille Universite, France, Dr Sheila Jagannathan, World Bank, USA, Prof Dinesh Chandra Jain, Computer Science & Engineering, S.G.I., India, Assc Antonín Jančařík, Charles University, Faculty of Education, Czech Republic, Prof Aman Jatain, Amity University, India, Kanthi Jayasundera, Simon Fraser University, Canada, Amor Jebali, King Abdulaziz University, Saudi Arabia, Dr Amor Jebali, King Abdulaziz University, Saudi Arabia, Runa Jesmin, Global Heart Forum, UK, Geraldine Jones, University of Bath, UK, Dr Jowati Juhary, National Defence University of Malaysia, Malaysia, Dr Dorela Kaçauni, Fan S. Noli University, Albania, Prof Leila Kajee, University of Johannesburg, South Africa, Dr Michail Kalogiannakis, University of Crete, Faculty of Education, Greece, Dr ArunaDevi Karuppasamy, Gulf College, Oman, Dr Elisabeth Katzlinger, Johannes Kepler University, Austria, Dr Andrea Kelz, University of Applied Sciences Burgenland, Campus Pinkafeld, Austria, Prof Jesuk Ko, Universidad Mayor de San Andres (UMSA), Bolivia, Iain Lambie, Glasgow Caledonian University, UK, Dr Maria Lambrou, University of the Aegean Business School, Greece, Dr Mona Laroussi, Institut National des Sciences Appliquées et de la Technologie, Tnis and Lille, Tunisia, Dr Pinte Lin, University of Reading, UK, Prof Ana Paula Lopes, Polytechnic of Porto (P.Porto)/ISCAP, Portugal, Prof Felipa Lopes dos Reis, Lusófona University, Portugal, Jenny Lorimer, University of Hertfortshire, UK, Ana Loureiro, Politechnic Institute of Santarem - School of Education, Portugal, Prof Sam Lubbe, NWU, South Africa, Dr Robert Lucas, Keylink Computers Ltd, Kenilworth, UK, Dr Magdalena Maciaszczyk, Lublin University of Technology, Polska, Dr Martin Magdin, Constantine the Philosopher University in Nitra, Faculty of Natural Sciences, Slovakia, Prof Abdelhak Mahmoudi, Ecole Normale supérieure, Morocco, Dr Katerina Makri, National and Kapodistrian University of Athens, Greece, Dr Chittaranjan Mandal, Dept of Computer Sc & Engg, IIT Kharagpur, India, Dr Arianit Maraj, AAB College-Faculty of Computer Sciences, Kosovo, Tobi Mark, The Michener Institute of Education at UHN, Toronto, Canada, Dr Bertil P. Marques, GILT/ISEP, Portugal, Dr Jose Alberto Martinez, Universsidad De La Laguna, Spain, Dr Maria J Martinez-Arguelles, Universitat Oberta de Catalunya, Spain, Bente Meyer, Aalborg University Copenhagen, Denmark, Prof Luisa Miranda, Polytechnic Institute of Braganca, Portugal, Dr Ali Moeini, Prague, Czech Republic, Iran, David Moffat, Glasgow Caledonian University, UK, Dr Begoña Montero-Fleta, Universitat Politecnica de Valencia, Spain, Prof Lina Morgado, Universidade Aberta, Portugal, Jolanda Morkel, STADIO Higher Education, South Africa, Dr Antonis Mouhtaropoulos, Metropolitan College, Greece, Greece, Peter Mozelius, Mid-Sweden University, Sweden, Ayanda Msomi, Nelson Mandela University, South Africa,

Dr Minoru Nakayama, Tokoyo Institute of Technology, Japan, Dr Michaela Nettekoven, WU Vienna University of Economics and Business, Austria, Dr Annie W.Y. Ng, City University of Hong Kong, Hong Kong, Prof Dick Ng'ambi, Uiversity of Cape Town, South Africa, Prof Emanuela-Alisa Nica, Center for Ethics and Health Policy and Petre Andrei University from Lasi, Romania, Assc Nanuli Ninashvili, Tbilisi State Medical University, Georgia, Dr Chetsada Noknoi, Thaksin University, Songkhla, Thailand, Prof Jarmila Novotná, Charles University, Czech Republic, Dr Kamila Olsevicova, Univeristy of Hradec Kralove, Czech Republic, Maria Osuna Alarcón, Salamanca University, Spain, Dr Abdul Jalil Othman, Faculty of Education, University of Malaya, Malaysia, Veerabhadram Paduri, Namibia University of Science and Technology, Namibia, Dr Alessandro Pagano, University of Bari, Italy, Vasileios Paliktzoglou, University of eastern Finland, Finland, Masouras Panicos, Cyprus University of Technology, Cyprus, Dr Stamatis Papadakis, University of Crete, Greece, Prof Kyparisia Papanikolaou, School of Pedagogical and Technological Education, Greece, Dr Iraklis Paraskakis, South East European Research Centre (SEERC), Thessaloniki, Greece, Dr Marcello Passarelli, National Research Council of Italy, Italy, Dr Ayyub Patel, King Khalid University College of Medicine Biochemistry, Saudi Arabia, Paul Peachey, University of Glamorgan, Treforest, UK, Dr Arna Peretz, Ben Gurion Univeristy of the Negev, Israel, Dr Carmen Pérez-Sabater, Universitat Politècnica de València, Spain, Dr Beth Perry, Athabasca University, Canada, Dr Donatella Persico, Istituto Tecnologie Didattiochje-Consiglio Nazionale Ricerche, Genova, Italy, Dr Christopher Perumalla, University of Toronto, Canada, Prof Nazanin Pilevari, Islamic Azad University, Iran, Prof Mário Pinto, Polytechnic Institute of Porto, Portugal, Prof Selwyn Piramuthu, University of Florida, Gainesville, USA, Assc Isabel Pitacas, Polytecnic Institute of Tomar, Portugal, Dr Toomas Plank, University of Tartu, Institute of Physics, Estonia, Dr Elif Polat Hopcan, Istanbul University- Cerrahpaşa, Türkiye, Prof Maria Popescu, Carol I National Defense University, Romania, Dr Francesca Pozzi, ITD-CNR, Italy, Susannah Quinsee, City University, London, UK, Dr Bilba Radu, George Bacovia University, Romania, Abdul Rafay, Asia Pacific University College of Technology & Innovation, Malaysia, Dr Prasanna Ramakrisnan, Universiti Teknologi MARA (UiTM), Malaysia, Prof Kemlall Ramsaroop Ramdass, Unisa, South Africa, Prof Pongsak Rattanachaikunsopon, Ubon Ratchathani University, Thailand, Dr Liana Razmerita, Copenhagen Business School, Denmark, Prof Asmaa Retbi, Mohammadia School of Enginners, Mohammed V University in Rabat, Morocco, Hugo Ribeiro, University of Porto, Portugal, Prof Sandra Ribeiro, ISCAP-IPP, Portugal, Prof Cristina Rodrigues, University of Minho, Portugal, Sumowalt Roosevelt, Liberia Initiatives For Fostering Empowerment, Inc., Liberia, Dr Joanna Rosak-Szyrocka, Częstochowa University of Technology, Poland, Dr Eleni Rossiou, Univerisity of Macedonia, Greece, Dr Danguole Rutkauskiene, Kaunas University of Technology, Lithuania, Prof Abdel-Badeeh Salem, Ain Shams University, Cairo, Egypt, David Sammon, Univesity College Cork, Ireland, Prof Vitor Santos, New University of Lisbon, Portugal, Dr Paulus Insap Santosa, Gadjah Mada University, Indonesia, Prof Maya Satratzemi, University of Macedonia, dept of Applied Informatics, Greece, Assistant Prof.Dr. Ousanee Sawagvudcharee, King Mongkut's Institute of Technology Ladkrabang Prince of Chumphon Campus, Thailand, Prof Jeanne Schreurs, Hasselt University, Diepenbeek, Belgium, Dr Jane Secker, London School of Economics, UK, Dr Fabio Serenelli, Università degli Studi Milano Bicocca, Italy, Dr Hazar Shehadeh, University of Arizona Global campus, USA, Dr Petia Sice, University of Northumbria, Newcastle-upon-Tyne, UK, Dr Armando Silva, ESE School of Education of Polytechnic of Porto, Portugal, Dr Paulino Silva, ISCAP / IPP, Portugal, Prof Susana Silva, School of Hospitality and Tourism, Institute Polytechnic of Porto, Portugal, Dr Gurmeet Singh, The University of The South Pacific, Suva, Fiji, Dr Deena Slockett, ADU (Adventist University of Health Sciences), USA, Dr Alisdair Smithies, Imperial College London, UK, Dr Keith Smyth, University of the Highlands and Islands, UK, Prof Filomena Soares, Polytechnic of Porto - P.PORTO / ISCAP / CEOS,

Portugal, Bent Soelberg, South Danish Educationscenter (SDE), Denmark, Dr Yeong-Tae Song, Towson University, Maryland, USA, Dr Michael Sonntag, FIM, Johannes Kepler University, Linz, Austria, Dr Sonia Sousa, Tallinn University, Estonia, Dr Rumen Stainov, University of Applied Sciences, Fulda, Germany, Martin Steinicke, University of Applied Sciences HTW Berlin, Germany, Caroline Stockman, University of Winchester, UK, Miss Baig Sundus, BPP University, United Kingdom, Dr Amanda Sykes, University of Glasgow, UK, Dr Rabail Tahir, Goldsmiths University of London, UK, Dr John Thompson, Buffalo State College, USA, Dr Socaciu Tiberiu, University of Suceava, Romania, Dr Łukasz Tomczyk, Pedagogical University of Cracow, Poland, Prof Cristina Torres, Porto Accounting and Business School – Polytechnic of Porto, Portugal, Dr Tiago Trancoso, IPVC, Portugal, Christopher Turner, UK, Diane Tyczynski, Michener Institute of Education at UHN, Toronto, Canada, Prof Andreas Veglis, Aristotle University of Thessaloniki, Greece, Dr Steven Verjans, Open Universiteit of The Netherlands, The Netherlands, Dr Carla Viana, Polytechnic Institute of Setúbal, Portugal, Prof Isabel VIeira, Polytechnic of Porto, Portugal, Dr Porawat Visutsak, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand, Ass. Professor Tone Vold, Business School, Norway, Elaine Walsh, University of Limerick, Ireland, Dr Philip Wilkinson-Blake, Loughborough University, UK, Dr Sheryl Williams, Loughborough University, UK, Dr Suzie Xiangping, University of Hertfordshire, Hatfield, UK, Dr Justyna Żywiołek, Czestochowa University of Technology, Poland

Fostering Self-Direction and Peer Feedback Skills in Management Course

Dina Adinda[1], Caroline R O'Neill[1], Elise Goiseau[2], Thomas Petit[2] and Christophe Jeunesse[1]
[1]Université Paris Nanterre, Nanterre, France
[2]Excelia Business School Research Lab

dadinda@parisnanterre.fr
oneillc@excelia-group.com
goiseaue@excelia-group.com
petitt@excelia-group.com
cjeunesse@parisnanterre.fr

Abstract: This study examines the implementation and effects of co-assessment within a management course on the development of students' ability to provide feedback, an essential competency for future managers. The size of the cohorts and the number of groups make this type of implementation challenging. Using a dedicated digital tool, therefore, opens a more realistic opportunity for deployment. Particular attention is given to self-direction, a transversal skill that enables learners to produce feedback, especially within the blended learning context. In this study, the ability to produce and provide feedback is assessed in a pre- and post-test, while their self-direction is examined through their self-determination, self-regulation, and self-efficacy in learning. By examining the interplay between co-assessment, peer feedback, and self-direction, this study aims to enhance understanding of peer feedback as both a training activity and an active learning practice. The findings identified a progression in feedback-related skills but show differentiated effects of the implemented learning scenario. The results highlighted a progression in learners' self-direction, notably their perceived self-efficacy in delivering feedback, while improvement in their self-regulation remains moderate. These findings shed light on the intrinsic nature of self-direction as a competency, which requires learners' proactivity in identifying their own action strategies in relation to goals as well as in monitoring their implementation to achieve targeted outcomes, while also inviting reflection on the role of external factors. This study aims to contribute to the ongoing discussion on how to design learning environments that foster both the development of feedback-related competencies and learners' self-direction in learning.

Keywords: Online Co-Assessment, Feedback, Self-Direction, Self-Regulation, Blended Learning

1. Introduction

The increasing integration of digital technologies into higher education has profoundly transformed learning environments and interactions. In management education, where work-study programmes and hybrid course structures are becoming increasingly common, blended learning environments provide new opportunities to diversify learning activities and support learner engagement. These environments combine face-to-face and online learning experiences and often require students to demonstrate greater autonomy in organizing, monitoring, and regulating their learning processes.

Within this context, self-direction in learning has become a central competence for higher education students. Self-direction refers to learners' capacity to take responsibility for their learning process through the articulation of motivational, cognitive, and metacognitive dimensions, including self-determination, self-regulation, and self-efficacy (Carré et al., 2011). While blended learning environments may provide greater flexibility and opportunities for learner agency, previous research suggests that increased flexibility alone does not automatically foster learners' self-direction (Carré, 2003). The design of learning activities and the forms of support provided by lecturers, therefore, remain essential.

Among pedagogical approaches associated with active learning, peer feedback and co-assessment have attracted growing attention in higher education. Peer feedback activities encourage learners to engage critically with both their own work and that of others, while also supporting communication, collaboration, and reflective practices. Research has highlighted the potential of peer assessment to foster learner autonomy, critical thinking, and self-regulation (Shen, Bai and Xue 2020; Girardet, 2020).

This study examines the implementation of a Learning management system platform and the ChallengeMe plugin, designed to support structured online co-assessment activities within a blended learning management module. This master's level module was designed as a blended learning environment integrating online co-assessment activities, including inter-group assessment, intra-group assessment, and meta-feedback activities. The objective of this research is twofold: first, to examine the effects of online co-assessment on learners' self-direction and its dimensions; second, to analyse the evolution of students' feedback-related skills, considered

essential transversal competencies for future managers. By exploring the relationships between blended learning, online co-assessment, peer feedback, and self-direction, this study contributes to ongoing discussions regarding the design of learning environments that support both learner autonomy and the development of feedback-related competencies in higher education.

2. Literature Review

2.1 Learners' Self-Direction in Blended Learning Environments

The development of learners' self-direction has become a central concern in higher education, particularly within blended learning environments that require students to manage learning activities across face-to-face and online contexts and increasingly rely on self-directed learning processes (Luo, Lin and Yang, 2021; Otto and Eickhoff, 2025). The flexibility associated with blended learning may offer opportunities for learner autonomy, but it also places greater responsibility on students to organize, monitor, and regulate their own learning processes.

Self-direction is generally conceptualized as a multidimensional competency involving motivational, cognitive, and metacognitive dimensions (Carré et al., 2011). Within research on adult learning, self-direction refers to learners' capacity to take responsibility for their learning through the articulation of self-determination, self-regulation, and self-efficacy (Carré, 2010). In this perspective, self-direction extends beyond the mere completion of learning tasks and involves learners' proactive engagement in defining goals, mobilizing resources, monitoring progress, and evaluating outcomes.

This conceptualization is closely related to research on self-directed learning, defined by Brockett and Hiemstra (1991) as a process through which learners assume control over the planning, implementation, and evaluation of their learning activities. However, several authors caution against assuming that learner self-direction emerges automatically from flexible or technology-mediated learning environments. Carré (2003) notably argues that increased freedom within educational settings does not necessarily foster one's self-direction, as learning environments remain shaped by instructional structures and institutional constraints. In this regard, Bandura's triadic reciprocal causation (Carré, 2003; Ponton and Carr, 2012) also highlights the influence on learners' behaviours toward learning that stems from their previous learning or personal experiences, and their environment.

Within blended learning environments, learners are therefore confronted with a dual challenge: managing academic expectations while regulating their engagement across multiple learning spaces and temporalities. Previous studies have shown that instructional design plays a critical role in supporting these processes by structuring activities, interactions, and learner support mechanisms (Adinda et al., 2022). In this perspective, blended learning environments should not merely be understood as technological arrangements, but as socio-technical learning environments that may either support or constrain learners' self-direction and self-regulatory processes. The development of self-direction in blended learning therefore appears closely linked to the nature of the proposed learning activities, particularly those encouraging active participation, reflection, collaboration, and feedback practices.

2.2 Online Co-Assessment and the Role of Feedback for Learning

Peer feedback and online co-assessment are increasingly considered as instructional strategies supporting active learning, and learners' engagement (Wu and Schunn, 2023), as they encourage them to adopt a more active role in the learning process while developing analytical, communicational, and interpersonal competencies (Taghizadeh Kerman et al., 2024).

Peer feedback can be understood as a collaborative learning activity through which learners provide evaluative and constructive comments on the work of their peers (Taghizadeh Kerman et al., 2024). Beyond its evaluative dimension, feedback constitutes a learning process that supports reflective thinking skills (Pham et al., 2020). Several studies have also highlighted the potential of peer assessment to foster learner autonomy and self-regulation by encouraging students to compare, justify, and revise their judgments (Shen, Bai and Xue, 2020; Kaplan, 2019).

In online and blended learning environments, digital technologies can further support these processes by structuring interactions, facilitating asynchronous collaboration, and providing traceability of learning activities. Online co-assessment environments may therefore create conditions favourable to reflective learning practices

by extending opportunities for interaction beyond face-to-face sessions. Previous research has shown that collaborative online activities and problem-based learning situations may contribute to the development of learners' self-direction in blended learning contexts (Adinda et al., 2022).

From an instructional design perspective, online co-assessment activities also provide opportunities for learners to engage with feedback at multiple levels. By allowing learners time to understand and interpret assessment criteria before responding or providing feedback for peers, online environments may encourage the production of more thoughtful and elaborated feedback. This process may contribute to the development of feedback-related skills, including critical thinking (Zhan, 2021), communication, and interpersonal skills (Lai et al., 2020), all of which are also essential leadership skills in workplace environments (Mumford, Campion and Morgeson, 2007). In this regard, Henderson et al. (2019) also add the ability to use appropriate resources for the completion of the activity as one of key elements describing capacity for feedback.

However, the effects of peer feedback on learners' self-direction remain complex and uneven across its different dimensions. While peer assessment may strengthen learners' perceived self-efficacy and engagement, the development of self-determination and self-regulatory processes may require more explicit instructional support and guidance from lecturers. The role of instructional design therefore appears central in structuring interactions, supporting reflection, and creating conditions favourable to learners' active participation and self-direction in online co-assessment activities. Furthermore, when learning activities require learners to provide feedback, the extent to which the learning environment and instructional design support learners' ability to provide feedback is also of interest.

3. Research Method

3.1 The Instructional Design of the Course and Study Sample

The study was conducted among 142 students in the Master in Management (MiM) at a French Business School. The participants were enrolled in a work-study programme, combining periods of academic study with professional experience or internship in a company. A 15-hours blended learning module called "Management and leadership" was selected for the study for two reasons: providing feedback is one of its key addressed competencies, and at this level, students should have advanced expertise to assess their peers' work.The assessment structure for the module comprised 25% of the total grade allocated to multiple-choice questions graded individually by the lecturer. The remaining 75% was focused on co-assessment, divided into two parts: 25% for intra-group assessment, where each student's performance was marked individually, and 50% for inter-group assessment, assessed collectively for the group as a whole.

The studied blended learning module included a case study activity, through which learners worked in small groups and participated in co-assessment, and peer feedback activities. To develop learners critical thinking skills, collaboration, communication and interpersonal skills, necessary for providing constructive feedback, the co-assessment adopted following stages (Figure 1):

- Anonymous inter-group assessment: each group assessed another group's work on case study using assessment criteria provided by lecturer and delivered written feedback following suggested guidelines. In this initial stage, the focus was on work assessment, which is considered more relevant than assessing individual contributions for training peer feedback skills
- Intra-group assessment: within each group, members assessed the individual contributions of their peers to the group's work on the case study, guided by an assessment grid provided by the lecturer
- Inter-group assessment and peer feedback review: each group review feedback and assessment of their work
- Meta-feedback: the lecturer provides feedback on students' written feedback from inter-group assessment, providing formative comments.

The blended learning environment was designed to support asynchronous learning activities. The online activities included individual preparation for the activities (consulting resources, reviewing the case study prior to the face-to-face session, preparing the inter-group assessment and carrying out intra-group assessment), and group activities (working on the case study after the face-to-face session and submitting works for the case study and for inter-group assessment). Discussion for the inter-group assessment was managed during the face-to-face session, although its preparation, notably the individual preparation for the activity, carried out online.

Figure 1 also shows that the groups were invited to review the inter-group assessment and their peers' feedback on their work. The discussion on this topic took place during the face-to-face learning.

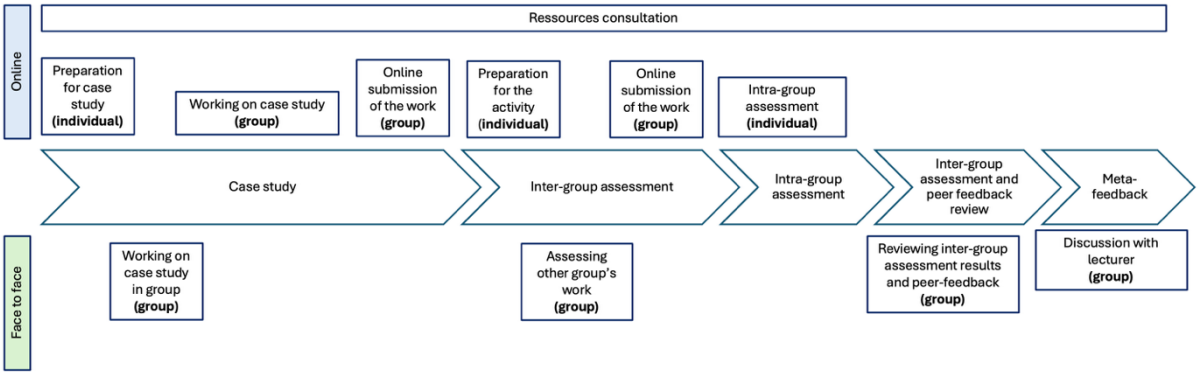


Figure 1: Instructional design of the course

3.2 Data Collection and Analysis Procedures

A questionnaire is often used for surveying a large number of participants. It allows data to be collected at different periods during the semester, enabling the observation of changes or eventual progress over time. Furthermore, as presented in the literature review, learning is a complex process shaped by individual and environmental factors. Given the limitations in observing students' learning experiences, notably the individual part of learning, learners' self-direction and its components (self-determination, self-regulation, and self-efficacy) were quantified through a self-assessment questionnaire with 5 points Likert scale. To understand learners' ability to provide feedback, a self-assessment questionnaire was also used. Feedback-related skills, including critical thinking, communication, interpersonal skills, and their ability to use appropriate resources for the completion of the activity were measured across four dimensions (Table 1), in which critical thinking is integrated into the knowledge of the subject dimension.

Table 1: Measurement model assessment

Constructs	Indicators	Cronbach alpha
Self-direction	Self-determination (SD)	0.843
	Self-efficacy (SE)	
	Self-regulation (SR)	
Feedback Skills (C)	Knowledge of the subject	0.820
	Communication skills	
	Interpersonal skills	
	Knowledge of resources	

Learners were invited to complete the self-assessment questionnaire on two occasions (at the beginning and at the end of the semester). The Cronbach's alpha for the instruments used (Table 1) was tested and revealed a high satisfactory value (Table 1), indicating that the items used are aligned with the objective of the instrument (Midy, 1996). For the analysis of the data related to self-direction and learners' feedback-related skills, by taking into account the Shapiro-Wilk test (Table 2) indicating low p-value and suggesting a violation of normality assumption, a Wilcoxon test was conducted to determine whether there was a statistically significant difference in score between the pre- and post-test.

Table 2: Shapiro-Wilk test results.

Indicators	p
Pre-SD – Post-SD	0.003
Pre-SE – Post-SE	0.14
Pre-SR – Post-SR	0.001
Pre-SC – Post-SC	<.001

4. Results

4.1 Online Co-Assessment and Development of Learners' Feedback-Related Skills

Pre- and post-test administered on the indicators of feedback-related skills (Table 1), pointed out a significant perceived progress in learners' ability to provide structured, relevant and constructive feedback. The mean score of the construct increased from 23.97 in the pre-test to 25.29 in the post-test, representing a significant progression ($p < 0.001$). This result suggested a perceived progress in learners' overall ability to provide feedback, which requires critical thinking, their knowledge of the subject matter and related resources, as well as their communication and interpersonal skills.

As for each indicator of the construct (Table 3), progressions in mean scores were also observed, regardless of their statistical significance. The results of Wilcoxon test indicated that the indicators of interpersonal skills represented the only non-significant progression, despite their higher post-test mean scores:

- "I feel comfortable giving feedback, even in sensitive situations" (C4, p-value = 0.282)
- "I know how to give feedback that encourages the other person to improve" (C5, p-value = 0.441)

These indicators focused on perceived confidence in complex interpersonal skills, which may require longer exposure to evolve significantly, as the reported mean scores of these two indicators do not show a significant increase.

As for learners' perception of the knowledge of the subject (C1), the results of the Wilcoxon test indicated a statistically significant progression ($p < 0.05$). Similar improvement were observed in learners' perception on their communication skills, particularly their ability to produce feedback in a clear and understandable manner (C2, p-value = 0.05), and adapt their feedback to the intended recipient (C3, p-value = 0.10), draw on course content when formulating feedback (C6, p-value = 0.05), and identify tools and external resources that may support them to produce and provide feedback (C7, p-value = <.001).

Table 3: Mean scores of Feedback related skills

Construct	Indicators	Pre-Test	Post-Test
Feedback Skills (C)		23.97	25.29
	Knowledge of the subject		
	C1	3.67	3.86
	Communication skills		
	C2	3.47	3.69
	C3	3.77	3.91
	Interpersonal skills		
	C4	3.31	3.39
	C5	3.81	3.87
	Knowledge of resources		
	C6	3.29	3.49
	C7	2.66	3.08

In sum, although changes in interpersonal skills remain to be further explored, the results indicated that feedback-related skills showed a significant overall improvement from pre- to post-test, with gains in all of the items.

4.2 Learners' Self-Direction in Blended Learning Environment

The results of the Wilcoxon analysis showed that students' scores of self-directions increased significantly between pre- and post-test ($p <.05$), with the mean scores rising from 56.1 to 57.4 (Figure 2). This progression is identified on two of three indicators: self-determination and self-efficacy.

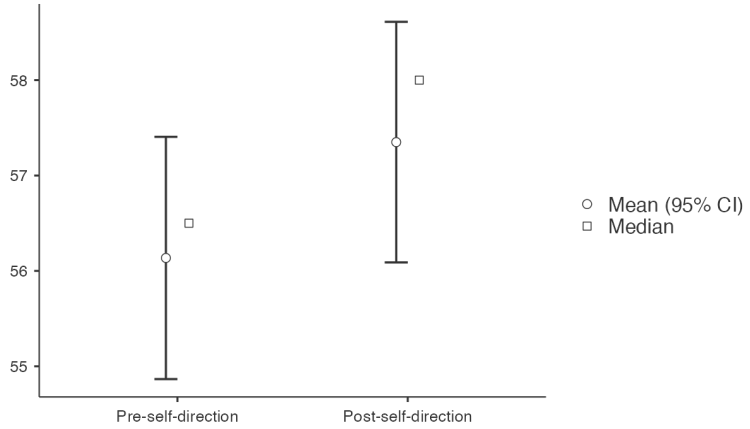


Figure 2: Wilcoxon test results on students' self-direction

The mean score for self-determination rose from 18.0 to 18.3 ($p < 0.1$). This construct encompasses the identification of personal and professional goals, the internalisation of external goals, freedom of action, proactivity, and engagement in learning. At the item level, most items showed increases. However, a small decrease was observed in participants' preference for general instruction allowing initiatives, indicating their need for a more structured one at the end of the program (SD4). Furthermore, the results indicated that the score remained stable at a relatively satisfactory level for participants' ability to identify their needs, strengths, and weaknesses when writing feedback at the beginning and the end of the module (SD3).

As for self-efficacy, the mean score for this construct increased from 18.6 to 19.3 ($p < 0.001$). This refers to participants perceived ability to evaluate their own work and that of others. At the item level, all items related to participants' perception of competence, motivation, and critical thinking showed increases. Only one item, related to participants perceived knowledge for producing feedback (SE4), showed no notable change, but remained at a relatively satisfactory level. At the item level, slight decreases, with a very small variation, were also found in participants' openness to receiving feedback on their errors (SR3), their control of the activity (SR4), and their perseverance to seek additional resources or information to complete a task (SR5).

Despite a minor variation and slight decreases in some items, the findings pointed out generally positive results of the intervention in learners' self-direction, notably in their perceived self-efficacy and self-determination, while improvement in their self-regulation remains moderate.

5. Conclusion

The findings of the study indicated positive outcomes from the experimentation, though the level of statistical significance varied. Perceived feedback-related skills improved significantly from pre- to post-test, with gains observed across almost all items. Only the indicator "interpersonal skills" in giving feedback do not show a significant increase over time. The results also highlighted growth in learners' self-direction, particularly in their perceived self-efficacy when delivering feedback, while improvements in self-regulation were more moderate.

Within self-direction constructs, self-efficacy showed the most improvement, with a statistically significant increase in scores, suggesting that participants' perception of their competency in the activity has improved, except for one item, showing that participants have a positive perception of their knowledge for producing feedback in general contexts since the beginning of the module. This result may show that participants already have some experiences in writing feedback. Nevertheless, the results also showed that they would need more extended exposure to the activity to develop their interpersonal skills significantly in giving feedback. Bandura (1997) argued that self-efficacy is strengthened through iterative activities and accumulated experiences. Indeed, one's perception of their self-efficacy is essential to feeling comfortable providing feedback in sensitive situations. Consequently, these findings also suggest that the ability to provide feedback, notably in sensitive situations requiring interpersonal skills, and the capacity to encourage improvement, are best fostered through repeated exercises, real-world practice, and observations.

In this study, self-regulation appears to be the only component of self-direction that did not demonstrate significant improvement over time. This may be explained by the structured nature of the module's instructional design (Figure 1), which may have limited students' perception of their need to self-regulate their actions: the completion period for the activity was predetermined, and the resources students needed to prepare their

feedback and understand the ones received, such as activity instructions and assessment criteria, were also provided. The co-assessment activities conducted during asynchronous time required students to work autonomously. As such, students' with extrinsic motivation would need to develop an integrated regulation of their actions (Deci & Ryan, 2016).

As for self-determination, the findings highlighted a significant increase. The only identified stable scores are related to participants' ability to identify their own needs, strengths, and weaknesses when producing throughout the module. This stability may be explained by the fact that the intervention did not specifically target participants' ability to identify their own needs, strengths, and weaknesses when producing feedback.

This study aims to contribute to the discussion on how to design learning environments that foster both the development of feedback-related competencies and learners' self-direction in learning. The findings of this work showed students need for more time to develop interpersonal skills. This can be provided in a blended learning environment through repeated exercises that can be conducted during both online and face-to-face activities, as each of them has its own specific potential for learning (Graham, 2022). For instance, online environment can allow repetition without social stress. In contrast, face-to-face exercise provides the opportunity to learn the emotional realism of giving feedback. To support mastery of the skill, role-play exercises, enriched with real-time peer feedback (Luttenberger et al., 2014), might also be beneficial in this context, and to address some slightly decreased skills: openness to receiving feedback on their errors (SR3) and perseverance to seek additional resources or information to complete a task (SR5).

The study also pointed out the need for a learning environment that supports the implementation of integrated regulation. For Deci and Ryan (2016), integrated regulation represents the highest level of internalisation of external learning goals, in which an individual's personal goals are aligned with external learning goals. Consequently, in this level of internalisation, students are able to engage in autonomous actions (Deci and Ryan, 2016). This objective can be targeted by organizing an introductory activity allowing students to recognize various resources provided by lecturer, while also to identify their own needs, strengths, and weaknesses when producing feedback.

In sum, the findings of this study showed that self-direction in learning, encompassing self-determination, self-efficacy and self-regulation, represents an interplay between individual proactivity and learning environment constraints, particularly those related to the learning activities associated with the potentials of a blended learning environment, and the targeted learning skills, confirming Bandura's triadic reciprocal causation (Carré, 2003; Ponton and Carr, 2012).

Finally, this study is limited by its reliance on self-reported measures, which only capture students' perceived progression of self-direction in learning and their feedback-related skills. The absence of a control group also constitutes a limitation of this study. However, for ethical and practical reasons, all students were therefore given access to the same learning activities. Future research could further explore causal inference by comparing alternative feedback-training designs. It could also examine the extent to which the perceived progressions observed vary according to the status of the feedback within the learning objectives. Furthermore, an analysis of the quality of the peer-feedbacks produced, combined with semi-directive interviews, could enrich the present study findings to further investigate students' performance and self-regulatory process, in relation to the role of inter-group and intra-group assessment. The role of the learning environment highlighted in this study also invite further reflection on the role of digital technology as a catalyst for teaching practices to foster learners' self-direction in learning, particularly given the increasing accessibility of generative artificial intelligence that can support students' feedback writing skills.

Ethics Declaration

Participation in this study was voluntary, and all data were fully anonymized.

AI Declaration

An AI tool was used for grammar verification of this manuscript.

References

Adinda, D., Marquet, P., Dalle, J.-M., & Dion, G. (2022) "Effets de situations à problèmes sur l'autodirection des étudiants : le cas d'une formation hybride à l'entrepreneuriat". *Revue internationale des technologies en pédagogie universitaire,* Vol 19, No. 2, pp.160-173. https://doi.org/10.18162/ritpu-2022-v19n2-11

Bandura, A. (1997) *Self-efficacy: The exercise of control.* W. H. Freeman and Company, New York.

Brockett, R. G., & Hiemstra, R. (1991) *Self-direction in adult learning: Perspectives on theory, research, and practice.* Routledge, London.

Carré, P. (2003) "La double dimension de l'apprentissage autodirigé contribution à une théorie du sujet social apprenant". *Canadian journal for the study of adult education,* Vol 17, No. 1, pp.66–91. http://cjsae.library.dal.ca/index.php/cjsae/article/viewFile/1871/1632

Carré, P. (2010) 'L'autodirection des apprentissages'. In P. Carré, A. Moisan, D. Poisson (ed.), *L'autoformation, perspectives de recherche* (1e éd, p. 117-169). Paris: Press Universitaire de France.

Carré, P., Jézégou, A., Kaplan, J., Cyrot, P., & Denoyel, N. (2011) "L'Autoformation: The state of research on self-(directed) learning in France". *International journal of self-directed learning®,* Vol 8, No. 1, pp.7-17. https://6c02e432-3b93-4c90-8218-8b8267d6b37b.filesusr.com/ugd/dfdeaf_b634f9f41e404b5c8a3b2746a0a3f71d.pdf

Deci, E. L., & Ryan, R. M. (2016) 'Favoriser la motivation optimale et la santé mental dans les divers milieux de vie'. In Y. Paquet, N. Carbonneau, et R. J. Vallerand (ed.), *La théorie de l'autodétermination : Aspect théoriques et appliqués* (p. 15-30). Bruxelles: de Boeck Supérieur.

Girardet, C. (2020) "Lorsque les étudiant·es produisent et reçoivent des feedbacks entre pairs à l'université: étude des perceptions d'un dispositif d'évaluation continue facilité par le numérique". *La Revue LEeE,* No. 2. https://doi.org/10.48325/rleee.002.04

Graham, C. R. (2022) 'Exploring definitions, models, frameworks, and theory for blended learning research'. In A. G. Picciano, C. D. Dziuban, C. R. Graham et P. D. Moskal (ed.), *Blended learning: Research perspectives* (Vol. 3). New York: Routledge.

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019) "Conditions that enable effective feedback". *Higher Education Research & Development,* Vol 38, No. 7, pp.1401–1416. https://doi.org/10.1080/07294360.2019.1657807

Kaplan, J. (2019) "Coévaluation entre pairs apprenants comme échafaudage de la régulation du processus de l'apprentissage". *Journal International de recherche en éducation et formation,* Vol 5, No. 1, pp.47-65.

Lai, C-Y., Chen, L-J., Yen, Y-C., & Lin, K-Y. (2020) "Impact of video annotation on undergraduate nursing students' communication performance and commenting behaviour during an online peer-assessment activity". *Australasian Journal of Educational Technology,* Vol *36*, No. 2. 71-88. https://doi.org/10.14742/ajet.4341

Luo, Y., Lin, J., & Yang, Y. (2021) "Students' motivation and continued intention with online self-regulated learning: A self-determination theory perspective". *Zeitschrift für Erziehungswissenschaft*, Vol 24, pp.1379–1399. https://doi.org/10.1007/s11618-021-01042-3

Luttenberger, K., Graessel, E., Simon, C., & Donath, C. (2014) "From board to bedside – training the communication competences of medical students with role plays". *BMC Medical Education,* Vol 14, No. 135. https://doi.org/10.1186/1472-6920-14-135

Midy, F. (1996). *Validité et fiabilité des questionnaires d'évaluation de la qualité de vie : une étude appliquée aux accidents vasculaires cérébraux* (Rapport no 9612). Laboratoire d'analyse et de techniques économiques, Faculté de Science Economique et de Gestion de l'Université de Bourgogne.

Mumford, T. V., Campion, M. A., & Morgeson, F. P. (2007) "The leadership skills strataplex: Leadership skill requirements across organizational levels". *The leadership quarterly,* Vol 18, No. 2, pp.154-166. https://www.sciencedirect.com/science/article/abs/pii/S1048984307000069?via%3Dihub

Otto, D., & Eickhoff, V. (2025) "The rise of private distance universities: A text-mining analysis of student satisfaction through the lens of self-determination theory". *Research in Learning Technology,* No. 33, 3452. https://doi.org/10.25304/rlt.v33.3452

Pham, T. N., Lin, M., Trinh, V. Q., & Bui, L. T. P. (2020) "Electronic peer-feedback, EFL academic writing and reflective thinking: Evidence form a Confucian context". *Sage Open.* https://journals.sagepub.com/doi/10.1177/2158244020914554

Ponton, M.-K., & Carr, P. B. (2012) "Autonomous learning and triadic reciprocal causation : a theoretical discussion". *International journal of self-directed learning,* Vol 9, No. 1, pp.1-10.

Shen, B., Bai, B., & Xue, W. (2020) "The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class". *Studies in Educational Evaluation,* No. 64, pp.1-10. https://doi.org/10.1016/j.stueduc.2019.100821

Taghizadeh Kerman, N., Banihashem, S. K., Karami, M., Er, E., Van Ginkel, S., & Noroozi, O. (2024) "Online peer feedback in higher education: A synthesis of the literature". *Education and Information Technologies,* Vol 27, pp.763-813. https://link.springer.com/article/10.1007/s10639-023-12273-8

Wu, Y., & Schunn, C. D. (2023) "Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback". *Contemporary Educational Psychology,* Vol 73. https://doi.org/10.1016/j.cedpsych.2023.102160

Zhan, Y. (2021) “What matters in design? Cultivating undergraduates’ critical thinking through online peer assessment in a Confucian heritage context”. *Assessment & Evaluation in Higher Education,* Vol 46, No. 4. https://doi.org/10.1080/02602938.2020.1804826

Alternative Higher Education Assessment Framework in the Generative AI Era: Project Proposal and Pitch Approach

Nam Aghaee, Daniel Lindvall and Christina Keller
Lund University, Department of Informatics, Lund, Sweden

nam.aghaee@ics.lu.se
daniel.lindvall@ics.lu.se
christina.keller@ics.lu.se

Abstract: Technological advancements are reshaping higher education (HE) by challenging the traditional non-proctored assignments. These formats have historically supported learning are increasingly vulnerable to Generative Artificial Intelligence (GAI) assisted content generation, raising concerns about authorship, academic integrity, and assessment of higher-order thinking. This study investigates the Project Proposal and Pitch (PPP) approach as an alternative assessment framework, combining a written project proposal with a structured oral pitch. A mixed-method pilot case study was conducted at a university in Sweden. Data sources include design narrative reflection on the assessment design process, and survey responses from undergraduate students with open-ended qualitative reflections. The analysis focuses on how students perceived the PPP format to shape learning practices and support assessment authenticity and students' learning motivation. Findings indicate generally strong student support and satisfactions for the PPP approach. Students report that it enhances idea development, problem-solving, and engagement with course content, while also improving communication and learning organization. The combined proposal and pitch format was perceived as more engaging than traditional project-based assignments and effective for demonstrating learning in a concise and reflective way. The results suggest that PPP shifts GAI use toward supportive, process-oriented engagement rather than output production. The findings contribute to discussions on assessment innovation by demonstrating how the PPP approach can strengthen reasoning, articulation, and situated meaning-making in GAI-augmented learning environments.

Keywords: Project Proposal, Pitch, Assessment Design, Project-Based Learning, Generative AI

1. Introduction

The advancement of Generative Artificial Intelligence (GAI) is reshaping higher education (HE), particularly in teaching, learning, and assessment design, simultaneously challenging established pedagogical and assessment practices (Mohamed Hashim et al., 2022). Project-Based Learning (PBL) offers a foundation for developing process-oriented assessment approaches (Evenddy et al., 2023), which have been shown to improve student learning in HE (Guo et al., 2020). PBL supports meaningful, real-world problem-solving that enables deeper understanding and knowledge retention (Almazroui, 2022; Guo et al., 2020) through inquiry, investigation, and collaborative learning, where students actively construct knowledge by applying theory in practice (Musa et al., 2011). This aligns with current demands for assessment designs that make reasoning and learning processes visible in an AI context where written outputs can be easily generated (Kasneci et al., 2023; Nartgün & Kennedy, 2024b).

PBL engages students in authentic, real-world challenges, supporting the relevance of academic knowledge and employability skills beyond the classroom (Rahman et al., 2022). PBL develops knowledge construction through complex problem-solving and prepares students for professional practice (Sanchez-Muñoz et al., 2020; Evenddy et al., 2023). However, GAI both supports and challenges these processes by assisting with information gathering and idea development while risking reduced engagement in critical thinking and knowledge construction when overused (Dempere et al., 2023; Gopane et al., 2024; Kasneci et al., 2023).

When GAI generates substantial parts of assignments or solutions, opportunities for meaningful learning may be diminished (Dempere et al., 2023; Nartgün & Kennedy, 2024b; Gopane et al., 2024), intensifying concerns about academic integrity, assessment validity, and authentic learning (Kasneci et al., 2023). The need for assessment approaches that emphasize reasoning, reflection, and process-oriented learning rather than final written outputs alone has increased (Awadallah Alkouk & Khlaif, 2024). Consequently, AI should be positioned as a supportive tool for exploration and idea refinement rather than solution generation (Dempere et al., 2023; Kasneci et al., 2023; Karunaratne, Aghaee, & Ferati, 2025).

Within PBL, project proposals function as authentic, real-world assessment artefacts involving project design and planning that mirror professional practice (Herrington, 2005). Complementing this, pitching a project proposal (Faff, 2015) is conceptualised as a structured oral assessment format in which students demonstrate understanding by presenting and defending ideas under time constraints (Lubarda et al., 2023; Luoma, 2004).

Here, knowledge is not merely reproduced but reorganised and communicated, requiring prioritisation of key ideas and adaptation to audience expectations. Accordingly, pitching functions both as a learning activity and as an assessment of conceptual understanding and communicative competence (Faff, 2015; Luoma, 2004).

Assessment focus is shifting toward the articulation of ideas, reasoning, and decision-making processes (Evenddy et al., 2023; Guo et al., 2020). However, project proposals as summative assessment artefacts remain underexplored. Addressing this gap, the Project Proposal and Pitch (PPP) approach combines written project proposals with structured oral pitches, shifting assessment from final products to visible learning processes with real world project ideas (Kasneci et al., 2023; Awadallah Alkouk & Khlaif, 2024). This positions the PPP as a promising assessment approach that enhances motivation and assessment authenticity in AI augmented HE. The approach was refined through three cycles and presented as an educator design narrative in this study. Two research questions guide this study: (1) What are students' perceptions of the PPP assessment approach? and (2) How do students use GAI to support their learning and performance within this assessment format? The purpose of this study is to explore the PPP approach as an alternative assessment format in HE, examining how students perceive it as an alternative in the learning process and addresses GAI related use and challenges through PBL principles, assessment, and process-oriented evaluation.

2. Project-Based Learning and Research Proposal in Higher Education

Thomas (2000) discusses, PBL as an instructional approach that structures learning around projects, which typically involve complex tasks centered on real-world problems or questions. These require students to engage in inquiry, investigation, problem-solving, decision-making, and design. PBL emphasizes student autonomy by allowing learners to work independently or collaboratively over extended periods, culminating in the development and presentation of realistic outcomes. In project proposal writing as an assignment and assessment, PBL enables students to apply theory, research, and project management to develop feasible, real-world solutions that culminate in a formal proposal and presentation (Thomas, 2000; based on Jones, Rasmussen, & Moffitt, 1997; Thomas, Mergendoller, & Michaelson, 1999).

PBL and the development of project proposals (problems and ideas) support learning and encourage learners to take ownership of their educational experiences (Du Toit-Brits, 2019). Students become active participants rather than passive in the learning process. This student-centred approach promotes deeper engagement and greater responsibility for learning outcomes, which is increasingly important as HE seeks to develop learners who can critically evaluate and appropriately use AI-supported technologies (Dempere et al., 2023; Kasneci et al., 2023). PBL and project proposal writing also promote interdisciplinary collaboration and communication, as students from diverse backgrounds work together to address complex problems, share perspectives, and develop teamwork skills essential in contemporary professional environments (Hussein, 2021; Juandi et al., 2021). Through such collaborative engagement, students develop a more comprehensive understanding of multifaceted issues while strengthening their communication, analytical reasoning, and problem-solving abilities (Juandi et al., 2021). These competencies are increasingly emphasized in HE as institutions respond to ongoing digital transformation and changing societal demands (Mohamed Hashim et al., 2022).

Oral assessment formats, including oral examinations and presentations, are widely recognised for their ability to promote deep learning, strengthen students' capacity to articulate reasoning, and develop disciplinary speaking skills (Lubarda et al., 2023; Luoma, 2004). However, traditional oral examinations are often resource-intensive and difficult to scale in HE contexts, particularly in large-enrolment courses where resource constraints pose significant challenges (Lubarda et al., 2023). Pitch-based assessment can therefore be viewed as a scalable adaptation of oral examination principles, retaining the core pedagogical value of spoken knowledge demonstration while addressing some of these practical limitations.

3. Methodological Approach

This study adopted a mixed-methods design. By integrating qualitative and quantitative data, mixed-methods research provides a more comprehensive understanding of the phenomenon under investigation (Creswell, 2013; Creswell & Plano Clark, 2011). In the initial phase, a design narrative approach was employed, in which teachers shared narrative accounts of their experiences with selected courses, including their perceptions of course structure, instructional design, and assessment practices. These narratives established the foundation for the subsequent intervention phase. Drawing on the concept of design narratives (Danko, 2006), the study incorporated an iterative development process to refine and evaluate the pedagogical framework. In the second phase, a pilot survey was conducted to examine students' perceptions of the PPP framework and the extent to

which the use of GAI was also in focus, but to support the validity and development of students' ideas rather than a writing assistant. This study adopted a mixed methods design. By integrating qualitative and quantitative data, mixed methods provided a more comprehensive understanding of the phenomenon under investigation (Creswell, 2013; Creswell & Plano Clark, 2011).

The rationale for this design is to capture measurable patterns in student perceptions and interpretive insights into how the assessment format was experienced and enacted. Following the convergent design logic (Creswell, 2013; Creswell & Plano Clark, 2011), qualitative and quantitative data were collected in parallel, analysed separately, then integrated in the interpretation phase. Enabling the study to move beyond descriptive outcomes explaining how and why the assessment format influenced student learning experiences (Creswell, 2013).

With a pilot survey, the study examines students' perceptions of the PPP framework and GAI support in validating their project ideas. Data were collected concurrently over a two-week period in May 2026 at the end of an Information Systems Management course at a Swedish university, with integration occurring at the interpretation stage to provide a comprehensive understanding of students' perceptions, consistent with Creswell and Plano Clark (2011). The course project required students to develop an innovative digital solution in management context by following the PPP framework, including problem identification, project proposal development, and a final project pitch of one minute per student (five minutes per group). Students received structured instruction on the pedagogical use of GAI in accordance with the provided guidelines at the department of Informatics and the course specific instructions.

Participants consisted of international Bachelor's students, predominantly from Generation Z, representing diverse academic programmes. Detailed demographic information was not collected or reported to preserve participant confidentiality, and demographic factors were outside the scope of this study. Student performance was assessed using the course grading rubric, which evaluated problem identification, the feasibility and originality of the proposed solution, application of the PPP framework, and the quality of the written proposal. Project proposals covered a wide range of digital innovation and management topics addressing real world societal and organisational challenges. Assessment followed the course syllabus and its intended learning outcomes. The pitch was assessed based on clarity of communication, adherence to the NABC pitching criteria, justification of the proposed solution, and the ability to respond to questions. As the study focuses on students' overall perceptions of the PPP framework and GAI support rather than individual project characteristics or student backgrounds, detailed examples of project proposals and the complete assessment rubric are omitted due to conference space limitations but are available from the authors upon request.

3.1 Design Narrative of Iterative Process

In this study, the researchers acted as both educator and designer, reflecting narrative on the iterative development of a project proposal and oral pitch assessment (PPP) across three teaching cycles. The design is understood as evolving through interactions between pedagogical intentions, student engagement, and technological change. Data include longitudinal reflective notes, student work, feedback, and assessment outcomes, analysed iteratively to identify patterns and guide revisions. This process shows how a proposal-based assessment developed into a structured pitch format to better support authentic PBL (Gratchev, 2023).

3.2 Questionnaire Design and Data Collection Process

To evaluate student perceptions of the PPP assessment design, a structured questionnaire was developed following established survey design principles (Creswell, 2013). The instrument included three demographic items, ten Likert-scale questions, and four open-ended questions addressing perceived learning support, engagement, clarity, communication skills development, and use of GAI in learning. The survey was administered to undergraduate students in a management and informatics course. Of 72 students, 41 responses were collected and 38 were valid for analysis, yielding a response rate of 52.7%. This moderate response rate is acknowledged as a limitation of voluntary course-based surveys. Open-ended items were included to capture qualitative reflections on the assessment experience.

3.3 Data Analysis

In line with the exploratory design (Creswell, 2013), quantitative data (Q3–Q10; Table 1) were analysed using descriptive statistics (means, standard deviations, frequency distributions). Item-level patterns were examined

and visualised using bar charts and distribution plots to assess central tendency and dispersion. Question 3 (Q3) and Question 11 (Q11) focused on students' reflections on using GAI in PPP framework, were analysed and visualised in Figures 1 to 5. Qualitative data from Q17–Q21 were analysed using inductive thematic analysis (Braun & Clarke, 2006), involving coding, theme development, and iterative refinement. Q12–Q16 were excluded from the analysis and not used in this study. Selected qualitative responses were also quantified using simple frequency counts to support interpretation. Quantitative analyses were conducted in R and Excel, while thematic analysis was carried out independently by three researchers and refined iteratively to strengthen analytical robustness in line with exploratory research principles (Creswell, 2013).

4. Results and Discussion

4.1 Developing the PPP Assessment Approach

The PPP approach was developed through an iterative design process grounded in PBL. It combines a written project proposal with a structured oral pitch, shifting assessment from product completion toward the evaluation of reasoning, decision-making, and communication. The proposal component supports planning, justification, and problem formulation in authentic, real-world contexts (Herrington, 2005; Gratchev, 2023), while the pitch requires students to articulate and defend ideas orally, demonstrating understanding and communicative competence (Luoma, 2004; Lubarda et al., 2023; Tan et al., 2022; Warda et al., 2024). The following sections present four iterative development cycles.

4.1.1 First Cycle: Initial Project Proposal Implementation

In the first iteration, the project proposal was introduced as a structured but open-ended component in an earlier course. Students were asked to identify a problem, outline a project idea, and present an initial solution structure. The design was grounded in PBL principles emphasizing engagement with authentic problems and idea generation (Du Toit-Brits, 2019; Rahman et al., 2022). Excluding the proposal from the final assessment resulted in varied submission quality, depth, and methodological clarity. Although some students framed problems with coherent reasoning, others produced descriptive, fragmented outputs lacking justification. This highlighted a key design flaw: while students generated ideas, the activity lacked sufficient scaffolding for structured reasoning, argumentation, and alignment with learning outcomes. Ultimately, this cycle revealed a distinct gap between intended pedagogical goals and actual student performance.

4.1.2 Second Developmental Cycle: Refinement and Scaffolding

In response, a second design cycle introduced stronger scaffolding and clearer assessment criteria. Instructions were refined to explicitly include methodological justification and tighter alignment with course learning outcomes. The project proposal was also redesigned as a simulated professional "call for proposals," including branding elements and structured submission guidelines. The proposal remained a formative or preparatory component rather than a final assessment. Based on teachers' assessments, reflective analysis and student performance indicated improved clarity in student reasoning, particularly in problem definition, justification, and linking solutions to context. This adjustment aligns with PBL literature emphasizing that structured support is essential for connecting theory to practice in complex tasks (Guo et al., 2020; Sanchez-Muñoz et al., 2020). From this cycle, based on teachers' experiences a key insight emerged: project proposals function not only as outputs but as process-oriented assessments that reveal and develop students' reasoning and professional thinking.

4.1.3 Final Development Cycle: Integration of Structured PPP Approach

The final evolution introduced a structured PPP model as the summative assessment, designed to counter GAI overreliance and ensure individual understanding (Kasneci et al., 2023; Dempere et al., 2023). Group tasks included a 4,000-word structured project proposal, simulated as a SEK 500,000–1,000,000 funding application for 12–18 months. This culminated in a final seminar where groups of 4–5 students gave a device-free, 5-minute oral pitch to peers acting as investors, followed by a 10-minute defence of their ideas. Guided by Faff (2015) and Lubarda et al. (2023), this device-free presentation evaluated knowledge and reasoning under cognitive constraints, reintroducing epistemic ownership into authentic assessment. Furthermore, as Luoma (2004) notes, speaking without technology assesses a multidimensional construct of fluency, coherence, and audience adaptation, directly reflecting students' true learning and ownership of their work.

4.1.4 Final Synthesis Contribution

The final PPP configuration strengthens assessment authenticity by combining a structured written proposal with an oral pitch, making student thinking visible through a dual-layer system. This design resists GAI-generated substitution by utilizing GAI for idea exploration while requiring students to demonstrate ownership through structured justification and live explanation (Karunaratne, Aghaee, & Ferati, 2025). Evolving from insights on under-scaffolded proposals and the inadequacy of written-only assessment in an AI-saturated context, PPP aligns with shifts toward AI literacy, critical thinking, and evaluative judgment. By integrating iterative writing with an oral defense, it shifts assessment from product completion to process visibility, emphasizing reasoning, reflection, and decision-making (Evenddy et al., 2023; Kasneci et al., 2023; Awadallah Alkouk & Khlaif, 2024). Ultimately, PPP connects theory to practice through authentic problem-solving (Rahman et al., 2022) and extends traditional PBL by verifying individual understanding where written work alone can no longer guarantee authentic learning (Dempere et al., 2023; Kasneci et al., 2023).

4.2 Result from the Empirical Data

The result captured students' perceptions of learning, engagement, communication, and assessment value. The survey indicates positive perceptions of the PPP approach across all dimensions. Participants included first-year and second year/exchange students who completed the same assessment, with study level considered only as contextual information.

4.2.1 Result From the Quantitative Part of the Survey

Survey responses indicate positive perceptions of the PPP approach. Table 1 illustrates statistics of survey responses reflecting on students' perceptions, reflecting questions 4 (Q4) to question 10 (Q10). Different columns in table 1 presents the number of valid responses (N) and missing values for each question, together with the mean scores and standard deviations. The results suggest consistently positive evaluations across the assessed dimensions, although some variation in responses is reflected in the standard deviation values.

As an additional validation step, exploratory Cronbach's alpha and Spearman correlation analyses were conducted across the main questionnaire constructs. No meaningful or unexpected patterns were identified beyond the descriptive findings. Since validating the measurement instrument and examining inter-variable relationships were outside the scope of this design-focused study, these analyses are not reported in detail.

Table 1. Descriptive statistics of survey responses examining students' perceptions of the PPP assessment approach across learning, reflection, communication, and engagement dimensions.

Table 1: Summary of the survey

Variable	N	Missing	Mean	SD
Structuring ideas (Q4)	38	0	3.68	0.933
Critical reflection (Q5)	37	1	3.68	0.973
Learning experience (Q6)	38	0	3.76	1.10
Communicating ideas (Q7)	38	0	3.68	1.25
Deep reflection (Q8)	38	0	3.39	1.35
Core understanding and learning (Q9)	37	1	3.59	1.24
More engaging than traditional formats (Q10)	38	0	3.95	1.21

Table 1 summarises the survey results. Overall, the median score across all items was 4, indicating generally positive perceptions of PPP. All items used the full five-point Likert scale, with responses ranging from 1 to 5. Students reported that the project proposal supported idea structuring (M = 3.68, Figure 1. Q4) and critical reflection (M = 3.68, Figure 2. Q5). The combination of proposal writing and pitching was perceived to enhance overall learning (M = 3.76), and communication benefits were also noted (M = 3.68). The lowest mean concerned pitching as a tool for deeper reflection (M = 3.39), while the highest rating was for engagement, with PPP seen as more engaging than traditional assessment formats (M = 3.95, Figure 3. Q10).

4. The project proposal helped me clearly structure my ideas and understanding of the topic.	Number of responses
1 = Strongly Disagree,	2 (5,3%)
2 = Disagree	1 (2,6%)
3 = Neutral	9 (23,7%)
4 = Agree	21 (55,3%)
5 = Strongly Agree	5 (13,2%)
Total	38 (100,0%)

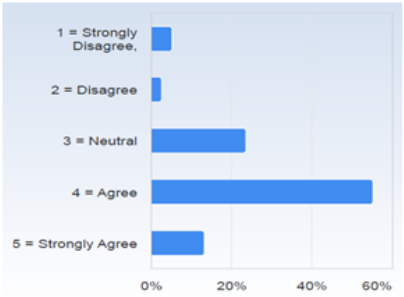


Figure 1: Presenting responses for Question 4.

5. Writing the project proposal improved my ability to reflect critically on course concepts and develop a real project idea.	Number of responses
1 = Strongly Disagree,	2 (5,4%)
2 = Disagree	2 (5,4%)
3 = Neutral	7 (18,9%)
4 = Agree	21 (56,8%)
5 = Strongly Agree	5 (13,5%)
Total	37 (100,0%)

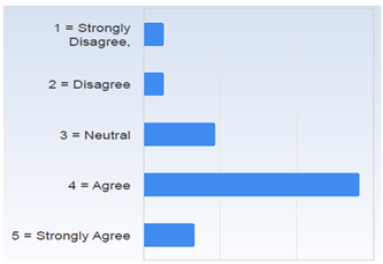


Figure 2: Presenting responses for Question 5.

10. Overall, this assessment method (proposal + pitch) was more engaging than traditional assessment formats (e.g., writing long project reports or essays and presenting in 20-min).	Number of responses
1 = Strongly Disagree,	2 (5,3%)
2 = Disagree	4 (10,5%)
3 = Neutral	4 (10,5%)
4 = Agree	12 (31,6%)
5 = Strongly Agree	16 (42,1%)
Total	38 (100,0%)

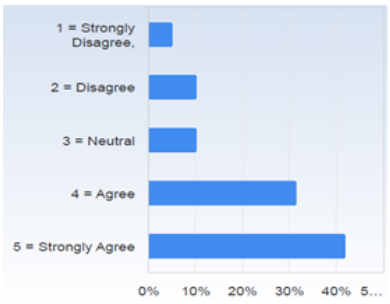


Figure 3: Presenting responses for Question 10.

Based on question 3, students reported frequent use of GAI (Figure 4), mainly for idea validation, structuring, and refinement rather than for initiating or writing proposals. Most used GAI to assess project viability (52.6%), while few used it for writing support (13.2%). This indicates that the PPP assessment design promotes GAI as a tool for validation and critical checking rather than content generation, supporting the view that well-designed assessment can guide GAI use toward learning-oriented engagement rather than substitution of student work (Dempere et al., 2023; Kasneci et al., 2023).

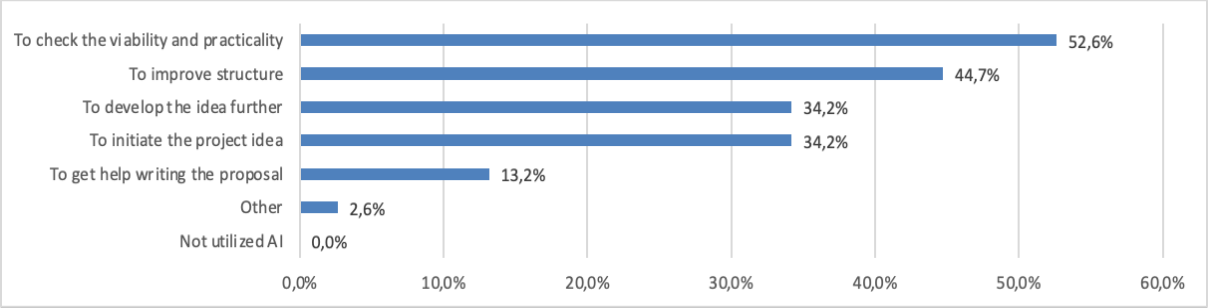


Figure 4: Reported use of GAI during project proposal development

Following Question 3 (Figure 4), Question 11 (Figure 5) examines students' use of GAI in the PPP context. A majority (65%) agreed that GAI supported the development and refinement of project proposal ideas, indicating its primary role as a tool for validation and improvement rather than direct content generation. This contrasts

with earlier studies showing more extensive use of GAI for producing assignments and reports (Dempere et al., 2023; Nartgün & Kennedy, 2024b; Gopane et al., 2024).

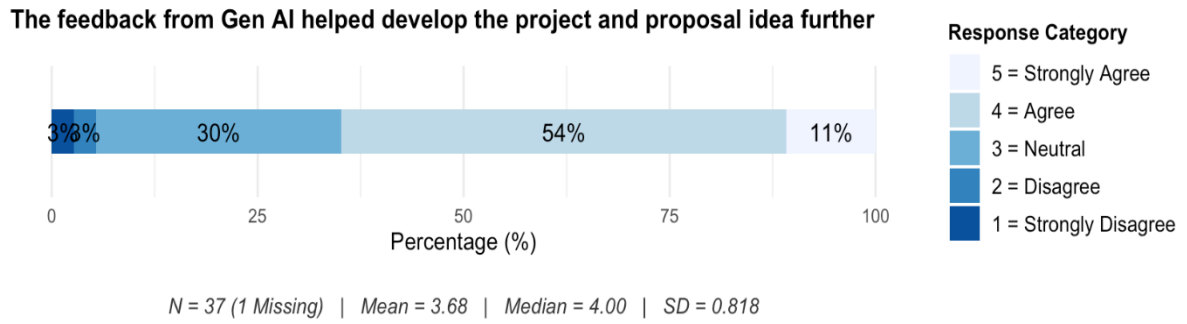


Figure 5: Students' Likert-scale responses on GAI support in developing and refining project and proposal ideas

4.3 Qualitative Reflections and Discussion in Open-Ended Questions

Based on responses to Question17 to –Question 21, the thematic analysis (Table 2) identifies patterns in student experiences regarding authenticity, collaboration, cognitive and skill development, assessment structure, and motivation, interpreted through PBL and assessment frameworks. A prominent theme is authentic, real-world learning, echoing prior studies (Almazroui, 2022; Guo et al., 2020; Musa et al., 2011). Students described the task as closely aligned with professional and entrepreneurial contexts—incorporating market and feasibility analyses. This reflects strong alignment with PBL principles of inquiry in authentic contexts (Thomas, 2000; Guo et al., 2020) and supports employability development through real-world relevance (Rahman et al., 2022; Herrington, 2005).

A second theme is collaborative knowledge construction, where learning is driven by group interaction and idea sharing (e.g., "sharing ideas/knowledge within the group"), consistent with socially constructed learning in PBL (Musa et al., 2011; Juandi et al., 2021). However, coordination challenges indicate the need for stronger self-organisation in student-centred learning (Du Toit-Brits, 2019). A third theme, cognitive development under constraints, concerns synthesis and prioritisation under time pressure (e.g., "only most important ideas in one-minute pitch"), reflecting deeper inquiry and decision-making (Evenddy et al., 2023; Guo et al., 2020). This also aligns with Kasneci et al. (2023), who emphasise making reasoning visible in AI-augmented assessment.

A fourth theme highlights structural tensions, particularly time pressure, unclear instructions, and imbalance between proposal and pitch ("too short pitch", "too long proposal", "vague combination"). While oral and process-based assessment is supported in the literature (Lubarda et al., 2023; Awadallah Alkouk & Khlaif, 2024), findings indicate that clarity and component alignment are critical for effective implementation. A fifth theme concerns skill development and professional preparation, especially communication, pitching, and creativity ("useful in life," "soft skills," "real-life relevant"), consistent with transferable skill development in PBL (Rahman et al., 2022) and communicative competence in oral assessment (Luoma, 2004).

A sixth theme relates to motivational responses, ranging from anxiety to engagement and confidence ("nervous," "fun," "motivated"), reflecting the dual nature of oral assessment as both challenging and developmentally valuable (Luoma, 2004; Kasneci et al., 2023). Overall, the findings show that PPP is perceived as highly authentic, engaging, and skill-developing, strongly aligned with PBL principles (Thomas, 2000; Guo et al., 2020), while also revealing implementation challenges related to structure, clarity, and pacing.

Across all open reflections in the survey, students consistently describe the PPP assessment as highly authentic and closely aligned with real-world practice, particularly through its focus on market analysis, feasibility, and project development in a simulated professional context. This supports PBL principles of learning through authentic inquiry (Thomas, 2000; Guo et al., 2020) and aligns with prior work on applied learning (Rahman et al., 2022; Herrington, 2005). Students also highlight collaborative knowledge construction and skill development, especially in communication, pitching, and critical thinking (Musa et al., 2011; Juandi et al., 2021; Luoma, 2004; Evenddy et al., 2023), although group coordination challenges and cognitive demands under time constraints are also evident.

Table 2: Theme–Subtheme–Theory Mapping Table

Theme	Subthemes	Key Student Evidence (Quotes)	Linked Literature
Real-world learning	Market analysis, feasibility, startup simulation	"real-world situation", "examine the feasibility", "working with real-life problems"	Thomas (2000); Guo et al. (2020); Rahman et al. (2022); Herrington (2005)
Collaboration & knowledge building	Group learning, idea sharing, coordination issues	"sharing the ideas/knowledge within the group"	Musa et al. (2011); Juandi et al. (2021); Du Toit-Brits (2019)
Critical thinking & under constraints	Prioritisation, analysis, decision-making	"only most important ideas", "expand critical skills"	Evenddy et al. (2023); Kasneci et al. (2023)
Assessment structure tensions	Time limits, imbalance, unclear instructions	"too short", "written proposal too long", "vague instructions"	Lubarda et al. (2023); Awadallah Alkouk & Khlaif (2024)
Skill development & employability	Oral skills, pitching, creativity	"useful in life", "training soft skills", "creative"	Rahman et al. (2022); Luoma (2004)
Motivational response	Anxiety, engagement, confidence	"fun and creative", "felt confident"	Kasneci et al. (2023); Luoma (2004)

The analysis reveals clear tensions in assessment design, particularly regarding workload balance, clarity of instructions, and the perceived mismatch between written and oral components, indicating the importance of careful implementation in process-oriented assessment (Lubarda et al., 2023; Awadallah Alkouk & Khlaif, 2024). Challenges regarding workload, grading transparency, and instructional clarity suggest a misalignment between intended process orientation and student experience (Kasneci et al., 2023; Awadallah Alkouk & Khlaif, 2024). Improvements include clearer instructions (n = 7), more explicit rubrics distinguishing proposal, process, and pitch components (n = 26 and n = 24), and structured phases with formative checkpoints to reduce cognitive load while preserving inquiry-based learning (Thomas, 2000; Guo et al., 2020).

4.4 Research Contribution and Take Away

Positive student feedback validates the PPP framework as an authentic, professionally relevant model that develop student's thinking, motivation, and active engagement (Gratchev, 2023). By integrating a structured written proposal with a live oral defence, PPP shifts the focus from mere output production to demonstrated reasoning, developing crucial skills in problem-solving, collaboration, and oral communication. Crucially, this dual-layer design addresses the core challenge of verifying individual authorship in an AI-saturated landscape (Kasneci et al., 2023; Dempere et al., 2023). Rather than substituting student effort, findings show that GAI was channelled toward idea validation, structuring, and refinement. This demonstrates that strategic assessment design and transparent expectations can actively cultivate a learning-oriented, responsible use of AI, contrasting sharply with the heavy reliance on AI-generated outputs reported in current literature (Dempere et al., 2023; Nartgün & Kennedy, 2024b; Gopane et al., 2024).

5. Conclusion and Future Work

In conclusion, the study shows that the PPP evolved through iterative refinement from an initially unstructured project proposal into a structured dual-component assessment consisting of a written justification and a constrained oral pitch. This development was reflected in students' perceptions and experiences, indicating that it represented a viable and engaging shift in assessment logic from product-oriented evaluation in PBL toward a process-visible assessment approach. In this approach, reasoning, justification, and communicative performance are emphasized as central learning processes in AI-augmented HE.

Student feedback indicates strong acceptance of the PPP format, showing it enhances engagement, authenticity, collaboration, and skill development while strengthening ownership of learning. Crucially, students used GAI primarily for ideation and validation rather than full content generation, demonstrating that the assessment design successfully repositions GAI as a tool for cognitive support rather than substitution. However, challenges regarding workload, component balance, and instructional clarity highlight that careful implementation is essential to align design intent with student experience. Overall, this study demonstrates that combining structured written and oral assessments (PPP framework) supports both authentic learning and reflective GAI engagement. As a small, exploratory pilot at a single Swedish university, these findings are not generalizable;

future research should expand this model across diverse contexts, adopt longitudinal designs, refine rubrics, and compare PPP with alternative assessment frameworks.

Acknowledgement

We sincerely thank Stig Nyman for his valuable input in generating and developing the core research idea. We also thank the survey participants for their time and feedback.

Ethics Declaration

The study adhered to institutional ethical guidelines and GDPR regulations. Participants provided voluntary, informed consent prior to data collection after being briefed on the study's purpose, publication potential, and data handling procedures. Participation was fully anonymous, allowing respondents to withdraw at any time without consequence. No sensitive data were collected, and all information was securely processed and reported in aggregated form to ensure confidentiality.

AI Declaration

While no AI tools were used in the core research process, generative AI was utilized solely for language correction, proofreading, and word-count reduction during survey preparation, instructional design, and manuscript writing.

References

Almazroui, K.M. (2023) “Project-based learning for 21st-century skills: An overview and case study of moral education in the UAE”, The Social Studies, 114(3), pp. 125–136.

Awadallah Alkouk, W. and Khlaif, Z.N. (2024) “AI-resistant assessments in higher education: Practical insights from faculty training workshops”, Frontiers in Education, 9, p. 1499495.

Creswell, J.W. and Plano Clark, V.L. (2011) Designing and Conducting Mixed Methods Research, 2nd edn, Sage, Thousand Oaks, CA.

Creswell, J.W. (2013) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage, Thousand Oaks.

Danko, S. (2006), Humanizing Design through Narrative Inquiry. Journal of Interior Design, 31. 10-28. https://doi.org/10.1111/j.1939-1668.2005.tb00408.x

Dempere, J., Modugu, K., Hesham, A. and Ramasamy, L.K. (2023) “The impact of ChatGPT on higher education”, Frontiers in Education, 8, p. 1206936. doi:10.3389/feduc.2023.1206936.

Du Toit-Brits, C. (2019) “A focus on self-directed learning: The role that educators’ expectations play in the enhancement of students’ self-directedness”, South African Journal of Education, 39(2), pp. 1–11.

Evenddy, S.S., Gailea, N. and Syafrizal, S. (2023) “Exploring the benefits and challenges of project-based learning in higher education”, PPSDP International Journal of Education, 2(2), pp. 458–469.

Faff, R.W. (2015) “A simple template for pitching research”, Accounting & Finance, 55(2), pp. 311–336.

Guo, P., Saab, N., Post, L.S. and Admiraal, W. (2020) “A review of project-based learning in higher education: Student outcomes and measures”, International Journal of Educational Research, 102, p. 101586.

Gopane, S., Kotecha, R., Obhan, J. and Pandey, R.K. (2024) “Cheat detection in online examinations using artificial intelligence”, ASEAN Engineering Journal, 14(1), pp. 121–128.

Gratchev, I. (2023) “Replacing exams with project-based assessment: Analysis of students’ performance and experience”, Education Sciences, 13(4), p. 408. doi:10.3390/educsci13040408.

Herrington, J. (2005) Authentic Learning Environments in Higher Education, IGI Global, Information Science Publishing.

Hussein, B. (2021) “Addressing collaboration challenges in project-based learning: The student’s perspective”, Education Sciences, 11(8), p. 434.

Karunaratne, T., Aghaee, N. and Ferati, M. (2025) “Harnessing the Generative AI Wave Towards Fair and Diverse Higher Education Assessments: A Comprehensive Analysis through an Innovative Lens of Students”, in European Conference on e-Learning, Academic Conferences International Limited, pp. 476–484.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E. and Krusche, S. (2023) “ChatGPT for good? On opportunities and challenges of large language models for education”, Learning and Individual Differences, 103, p. 102274.

Lubarda, M.V., Phan, A.M., Ghazinejad, M., Delson, N., Baghdadchi, S., Schurgers, C., Kim, M., Relaford-Doyle, J., Sandoval, C.L. and Qi, H. (2023) “Peer oral exams: A learner-centered authentic assessment approach scalable to large classes”.

Luoma, S. (2004) Assessing Speaking, Cambridge University Press.

Mohamed Hashim, M.A., Tlemsani, I. and Matthews, R. (2022) “Higher education strategy in digital transformation”, Education and Information Technologies, 27(3), pp. 3171–3195. doi:10.1007/s10639-021-10739-1.

Musa, F., Mufti, N., Latiff, R.A. and Amin, M.M. (2011) “Project-based learning: Promoting meaningful language learning for workplace skills”, Procedia-Social and Behavioral Sciences, 18, pp. 187–195.

Nartgün, Z. and Kennedy, E. (2024a) “Cheating in higher education in the age of artificial intelligence”, International Journal on Lifelong Education and Leadership, 10(2), pp. 47–54.

Nartgün, Z. and Kennedy, E. (2024b) “Artificial Intelligence (AI) and education: Contributions opportunities and challenges”, Political Economy and Management of Education, 5(2), pp. 24–46.

Rahman, T., Fitria, N., Nurhidayah, E. and Yuliandani, I. (2023) “Effects of project-based learning on employability skills”, Review of Islamic Studies, 2(1), pp. 1–10.

Sanchez-Muñoz, R., Carrió, M., Rodríguez, G., Pérez, N. and Moyano, E. (2022) “A hybrid strategy to develop real-life competences combining flipped classroom, jigsaw method and project-based learning”, Journal of Biological Education, 56(5), pp. 540–551.

Thomas, J.W. (2000) A Review of Research on Project-Based Learning. Available at: https://tecfa.unige.ch/proj/eteach-net/Thomas_researchreview_PBL.pdf.

Pre-Service Teachers' Perceptions of Artificial Intelligence within the UTAUT Model

Dilara Arzugül Aksoy[1] and Elif Polat[2]
[1] Bayburt University, Bayburt, Türkiye
[2] Istanbul University - Cerrahpasa, İstanbul, Türkiye

dilaraaksoy@bayburt.edu.tr
elif.polat@iuc.edu.tr

Abstract: This study aims to examine pre-service teachers' perceptions of using artificial intelligence (AI) tools in their learning processes and their intentions to utilize these technologies as future teachers within the framework of the Unified Theory of Acceptance and Use of Technology (UTAUT) model. Adopting a qualitative approach, the participants of this research consist of education faculty students studying at a public university in Türkiye. In this context, data were collected from 109 pre-service teachers via written opinion forms and from 6 pre-service teachers through semi-structured interviews. Descriptive findings indicate that all participants reported using AI tools, and half of them reported using these tools a few times a week. Content analysis was employed to derive codes from the data. Codes derived from both written responses and interview data were organized under the themes of performance expectancy, effort expectancy, social influence, facilitating conditions, behavioral intention, and risks and concerns. The findings reveal that students primarily associate the performance expectancy of AI tools with "fast and easy access to information" and "simplified understanding." In terms of effort expectancy, most students found AI tools easy to use, while some students also reported expending effort to verify the accuracy of AI responses. Regarding social influence, participants reported that their AI awareness and usage are shaped significantly by their social environment, particularly by peers and social media. Concerning facilitating conditions, despite the widespread adoption of AI across all fields today, participants emphasized the need for training and curriculum integration to support informed AI use. Although pre-service teachers most frequently intend to use AI for content generation and activity/task design in their future teaching, they expressed significant concerns. These include the potential for AI to promote cognitive laziness, disseminate inaccurate or biased information, negatively affect creativity, and pose risks to data privacy and security. The study concludes that while future educators utilize AI and are willing to integrate it into their practice, they remain cautious, particularly regarding the risk of cognitive laziness. Consequently, it is recommended that AI literacy be integrated into education programs to better prepare future educators for AI-enhanced classrooms.

Keywords: Pre-Service Teacher, Artificial Intelligence, UTAUT, Behavioral Intention, AI in Education

1. Introduction

The adventure of artificial intelligence (AI), which began in 1950 with Alan Turing's seminal question, "Can machines think?", gained official recognition through the "Dartmouth Artificial Intelligence Summer Research Project" led by John McCarthy (Dartmouth, 1956). This journey, initiated by a single inquiry, has evolved into a sophisticated technology capable of perceiving and understanding human language (Bozkurt, 2023) and generating novel content in various formats, such as text, audio, images, or video (Gupta et al., 2024). Due to its "generative" capabilities, AI has begun to integrate into numerous fields, ranging from tourism (Scarpi, 2024) to healthcare (Babu & Boddu, 2024). Education has emerged as one of the sectors undergoing a profound transformation process driven by this technological advancement (Bozkurt, 2023).

Regarding its contributions to education, AI has been shown to offer significant benefits, including content development, providing feedback, personalizing learning, enhancing accessibility, and increasing engagement (Ayeni et al., 2024; Grover, 2024). As an instructional assistant, AI can support educators in tasks such as lesson planning, material preparation, and assessment (Hsu & Ching, 2023). However, AI also brings forth several concerns related to security, data privacy, copyright, and the generation of biased or inaccurate information (UNESCO, 2024). It has been noted that the existing digital divide in accessing and using internet tools may be further exacerbated by AI (Daepp & Counts, 2024). Moreover, the non-transparent operational principles of AI tools, often referred to as the "black box" phenomenon, and the lack of clarity regarding their training datasets (Bozkurt, 2023) heighten user concerns over data privacy and security (Grover, 2024). Additionally, serious apprehensions persist regarding academic integrity, the accuracy of generated outputs, the potential loss of critical thinking skills, and the development of overreliance (Pitts et al., 2025).

Despite the aforementioned concerns, the existing literature indicates that the majority of higher education students maintain a positive perception toward AI (Almassaad et al., 2024; Yakut et al., 2025). According to Chan and Hu (2023), students use AI tools for purposes such as personalized and instant learning support, research and analysis support, assistance with writing and brainstorming, and facilitating visual and auditory multimedia.

However, it was also noted that students are concerned about the accuracy of the output produced, data privacy and security risks, and unclear policies (Chan & Hu, 2023). Furthermore, various studies have investigated the determinants of students' acceptance of AI and indicated that adoption is primarily driven by factors such as perceived benefits, perceived ease of use, and perceived usefulness, alongside behavioral intentions and overall attitudes (Chaamwe, 2025).

The integration of technologies, from television to the internet, into educational settings has been extensively analyzed through the lens of technology acceptance frameworks. Davis (1989) addressed the adoption of information technologies through the Technology Acceptance Model (TAM), focusing on perceived ease of use and perceived usefulness. Building upon this, Venkatesh et al. (2003) developed the Unified Theory of Acceptance and Use of Technology (UTAUT), an integrated model synthesizing core constructs from previous theories. UTAUT posits that performance expectancy, effort expectancy, facilitating conditions, and social influence are direct determinants of behavioral intention and actual usage behavior. Performance expectancy reflects the belief that technology improves job performance, while effort expectancy captures perceived ease of use. Facilitating conditions refer to the necessary infrastructure and resources, whereas social influence denotes the impact of significant others and the environment on adoption.

The "Quality Education" goal of the United Nations Sustainable Development Goals aims to ensure "inclusive and equitable quality education" for all. Similarly, the OECD (2021) report, titled "Building the Future of Education," emphasizes that education is vital for social progress and poverty alleviation. To achieve this, the report stresses the necessity of distributing educational opportunities equitably throughout society. AI holds the transformative potential to bridge these gaps by providing inclusive and accessible quality education through the rapid generation of content in diverse formats tailored to individual student needs. However, the successful realization of this potential depends not only on the technology itself but also on the educators who implement it. Teachers play a critical role in maximizing the opportunities AI offers to education while mitigating its inherent risks. Therefore, understanding the factors that drive or hinder pre-service teachers' adoption of AI is essential for preparing a generation of educators capable of fostering equitable learning environments in an AI-enhanced landscape. Accordingly, this study seeks to address the following research question:

- What factors influence pre-service teachers’ AI adoption intention within the framework of UTAUT?

2. Methodology

This study employs a qualitative research design to gain an in-depth and nuanced understanding of pre-service teachers' perspectives on AI. The study group consists of a total of 109 students selected through purposive sampling from the faculty of education at a leading public university in Istanbul.

Two different qualitative data collection methods were used to ensure data triangulation and increase the reliability of the findings. The written opinion form and interview questions were prepared by the researchers within the framework of UTAUT. First, data was collected from 109 pre-service teachers through structured written interview forms. This allowed for the collection of a wide range of qualitative data by capturing diverse perspectives. The written opinion forms were designed using Google Forms and distributed electronically to the participants. Data collection was conducted on a voluntary basis, ensuring that all pre-service teachers participated in the online platform. Following the analysis of the written forms, in-depth semi-structured interviews were conducted with pre-service teachers from 6 different departments. This stage aimed to examine the themes emerging from the written data in greater depth. The semi-structured interviews were carried out in an online environment. Following the completion of the interviews, the video and audio recordings were systematically transcribed to prepare the qualitative data for content analysis.

Demographic data in written forms were analyzed using descriptive statistical analysis to determine the demographic characteristics of the pre-service teachers (e.g., department, class level) and their usage patterns regarding AI tools (e.g., frequency of use, daily time spent). Open-ended questions in written form and semi-structured interview were analyzed using content analysis. This process involved systematically coding the data and identifying recurring patterns. Codes derived from both written responses and interview data were organized under the themes of performance expectancy, effort expectancy, social influence, facilitating conditions, behavioral intention, and risks and concerns. MAXQDA 2020 was used for content analysis. For written form data analysis, descriptive statistical analysis was employed.

Table 1 presents the demographic characteristics and AI tool usage profiles of the participating pre-service teachers.

Table 1: Demographic characteristics and AI usage patterns of the participants

	Item	f	%
Department	Computer Education and Instructional Technology	33	30.2%
	Primary Education	25	23%
	Early Childhood Education	16	14.7%
	Turkish Language Education	12	11%
	Science Education	9	8.2%
	Guidance and Psychological Counseling	7	6.2%
	Elementary Mathematics Education	4	3.7%
	English Language Teaching	2	1.9%
	Special Education	1	0.9%
Class Level	1	20	18.3%
	2	77	70.1%
	3	8	7.3%
	4	4	3.7%
Have you used AI tools (such as ChatGPT) before?	Yes	109	100%
	No	0	0
How often do you use AI tools?	A few times a month	5	4.6%
	Once a week	3	2.8%
	A few times a week	54	49.6%
	Once a day	8	7.3%
	A few times a day	39	35.8%
How much time do you spend using AI tools daily?	I don't use it at all	2	1.9%
	Less than 1 hour	83	76.1%
	1–3 hours	23	21.1%
	4–6 hours	1	0.9%

According to Table 1, regarding department, the highest participation was from Computer Education and Instructional Technology (30.2%, n = 33), followed by Primary Education (23%, n = 25). In terms of class levels, most of the participants were second-year students (70.1%, n = 77). Crucially, all participants reported having used AI tools such as ChatGPT previously. Regarding the frequency of use, nearly half of the participants (49.6%, n = 54) utilized AI tools a few times a week, while a significant portion (35.8%, n = 39) reported using them a few times a day. Despite this high frequency, 76.1% (n = 83) of pre-service teachers spent less than one hour a day on these platforms. Although all students reported having used AI tools before, the two students who answered "I don't use it at all" likely no longer use them on a daily basis.

3. Findings

This section presents the findings derived from both the written opinion forms and the semi-structured interviews. To ensure participant anonymity and confidentiality in the presentation of the interview data, students were assigned codes such as P1, P2, and P3. The distribution of findings obtained from the written opinion forms is presented in Table 2. These findings have been structured according to the UTAUT framework and additional emerging dimensions.

Table 2: Codes derived from pre-service teachers' written opinions regarding AI acceptance

Theme	Code	f	%
Performance Expectancy	Fast and easy access to information	36	33%
	Simplifying understanding	23	21.1%
	Time saving	20	18.3%
	Homework support	10	9.1%
	Summarization	8	7.3%
	Question generation	7	6.2%
	Research and idea exchange	6	5.5%
	Offer a different perspective	6	5.5%
	Personalized learning	5	4.6%
	Visual and material creation	5	4.6%
	Language learning	3	2.8%
	Others	2	1.9%
Effort Expectancy	Ease of use	89	81.7%
	Effort required to verify accuracy	7	6.2%
	Difficult to use in complex tasks	7	6.2%
Social Influence	Positive social impact	47	43.1%
	Social media effect	15	13.8%
	Limited social influence	9	8.2%
	Guidance from instructors	8	7.3%
	Negative impact of others' intensive ai use	3	2.8%
Facilitating Conditions	Need for training and curriculum integration	45	41.2%
	Prompt writing skill	23	21.1%
	Providing paid ai versions free to students	12	11%
Future Behavioral Intention	Content generation	33	30.2%
	Activity / task design	20	18.3%
	Interactive / advanced design	12	11%
	Lesson plan preparation	7	6.2%
	Personalized learning	5	4.6%
	I do not plan to use it	5	4.6%
	Preparation of assessment	5	4.6%
	Assignment grading and feedback	2	1.9%
	Others	2	1.9%
Risks and Concerns	Risk of cognitive laziness	51	46.8%
	Risk of false or biased information	24	22%
	Concerns about reduced creativity	12	11%
	Data privacy and security	11	10.1%
	Academic integrity	7	6.2%
	Information pollution	2	1.9%

According to interview findings, the content analysis of the data revealed several core factors influencing pre-service teachers' acceptance and future use of AI, as illustrated in Figure 1.

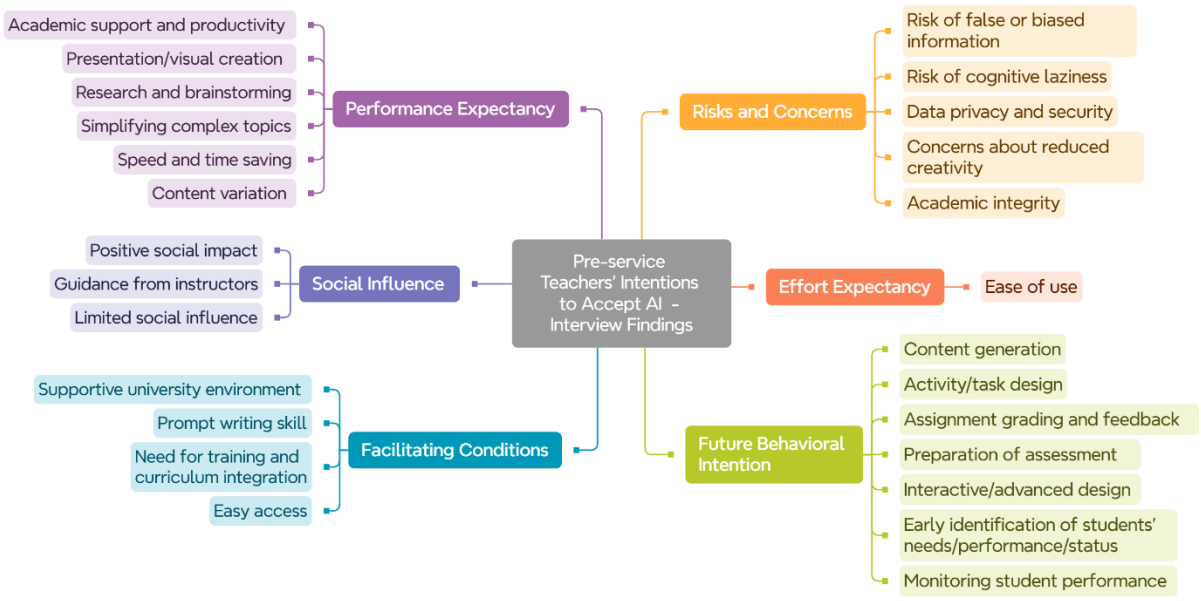


Figure 1: Pre-service teachers' perceptions regarding the intention to adopt AI

The written opinion form (Table 2) and interview data (Figure 1) were described below, categorized by the UTAUT framework and "risks and concerns" theme.

3.1 Performance Expectancy

Table 2 shows the themes that emerged from the written opinion forms, which were organised according to the UTAUT framework and additional emerging dimensions. Within the "performance expectancy" theme, "fast and easy access to information" was the most common response (33%, n = 36), followed by "simplifying understanding" (21.1%, n = 23) and "saving time" (18.3%, n = 20). In addition to these, AI is also used for purposes such as summarization, question generation, research and idea exchange, visual and material creation, and language learning. These results suggest that pre-service teachers primarily value AI for its efficiency in managing information and providing cognitive support. According to interview findings for "performance expectancy", participants emphasized the multifaceted advantages of AI in enhancing academic productivity. The data suggested that AI significantly facilitates learning processes by bridging information gaps and structuring experimental or project phases. Moreover, it streamlines academic tasks such as reporting, correspondence, and material preparation thereby increasing overall efficiency. This capability further supports students' autonomous study processes. Pre-service teachers specifically noted that AI tools facilitate research and brainstorming, simplify intricate topics, and provide significant time and cost savings. Additionally, the capacity of AI to assist with presentation design, visual creation, and content diversification was identified as a primary motivating factor for its integration into the learning process. Within the context of performance expectancy, representative participant responses are as follows:

> *"AI creates flashcards. For example, I learn quickly using double-sided cards, and it's great. It also presents me with more organized data, making learning easier. Besides that, I can tell it to explain a topic in more depth, specifying the features myself. It designs the explanation accordingly, and I learn faster that way." (P3)*

> *"AI is very effective in saving us time because it scans online resources, articles, or books we don't have in physical form much faster and provides quicker results, helping us understand, learn, and apply topics in practice. It allows us to do things in a shorter time." (P5)*

3.2 Effort Expectancy

Regarding "effort expectancy", most participants associated AI tools with "ease of use" (81.7%, n = 89). However, a small number of participants noted concerns regarding the "effort required to verify accuracy" (6.2%, n = 7), suggesting that, while the tools are easy to operate, the workload related to data validation remains a factor.

The interview findings also directly support the written interview form findings. Teacher candidates predominantly associated AI tools with "ease of use," without encountering technical barriers. This suggests that

when AI tools are perceived as user-friendly, the effort required to learn them decreases, thus encouraging their adoption. In this regard, the view of the participant coded as P6 is as follows: "*To learn how to use ChatGPT, I initially used to give commands and things like that. After that, I understood how to use it. I got used to it quickly. It didn't take long." (P6)*

3.3 Social Influence

"Social influence" dimension revealed that "positive social impact" (43.1%, n = 47) is a significant driver, whereas "guidance from instructors"(7.3%, n = 8) appears to play a smaller role in the current adoption phase. Participants reported that their AI awareness and usage are shaped significantly by their social environment, particularly by peers and social media. However, nearly all of the pre-service teachers interviewed emphasized the importance of guidance from instructors rather than peers. Some participants stated that their social environment did not play a significant role in their acceptance of artificial intelligence. Example participant responses are as follows:

> *"There was a time when it became very popular. Seeing it frequently in the news and on social media, I immediately downloaded ChatGPT. After trying it, I liked it." (P3)*

> *"Rather than being influenced by my social environment, I started using artificial intelligence because I saw its benefits firsthand. My surroundings do not have much impact on me in this regard, as I know what kind of outcomes I can achieve when I use AI myself." (P5)*

3.4 Facilitating Conditions

Regarding "facilitating conditions", the "need for training and curriculum integration" was the most frequently cited requirement (41.2%, n = 45). This finding underscores the necessity for context-oriented practical training to ensure the effective and efficient use of these tools, emphasizing how they should be integrated with appropriate pedagogies.

In interview data, "supportive university environment" and "easy access to tools" were identified as critical facilitating conditions. However, as highlighted in the participants' responses, there is also a clear need for formal training on AI. In addition, participants suggested incorporating AI-related courses into the current curriculum to enable them to integrate this technology into education in the future. Some students, rather than advocating for a standalone course, emphasized that AI could be integrated across existing courses. Furthermore, "prompt writing skills" emerged as a specific technical competency necessary for effectively leveraging AI capabilities in educational contexts. Example views are as follows:

> *"Most students use it, and they are already encouraged to do so. However, instructors emphasize that while its use is inevitable as technology advances, students should verify the accuracy of the information and incorporate their own thinking into the process." (P3)*

> *"I think students should be informed about how much they should use AI and in which areas it is appropriate to use it. They should also be taught how to use it effectively when preparing course materials. In addition, it is important to emphasize that AI should not be used for harmful purposes and that students should be guided to avoid using it in ways that undermine skills such as creative thinking and problem-solving." (P6)*

3.5 Behavioral Intention

According to written interview form data, participants' "future behavioral intention" focused heavily on "content generation" (30.2%, n = 33) and "activity/task design" (18.3%, n = 20). This suggests that pre-service teachers envisage AI more as creative assistants for instructional materials than for administrative tasks such as assessment preparation (4.6%, n = 5). The interview data also supported the form findings. All pre-service teachers interviewed expressed behavioral intentions to use AI in education in the future. Participants strongly associated AI with various pedagogical practices. They most frequently reported plans to use AI for "content creation," "activity/task design," and "assignment grading and feedback."

> *"It can assist me in designing infographics. I can prepare more effective materials that are visually oriented, concise, and clear, without taking too much time. Students will naturally tend to choose the easier option, and I think AI is very effective in quickly providing information and enhancing retention of learning." (P3)*

> *"For example, I think that in cases where I cannot directly demonstrate experiments or abstract concepts—such as weather changes or phase transitions—I can make them more understandable for students by visualizing them or using video-based explanations." (P5)*

3.6 Risks and Concerns

Finally, while UTAUT framework provides a basis for technology acceptance, participants in this study also articulated several risks and concerns that fall outside the model's core constructs. Under "risks and concerns", the "risk of cognitive laziness" was the most significant apprehension (46.8%, n = 51), far outweighing concerns about "data privacy and security" (10.1%, n = 11) or "academic integrity" (6.2%, n = 7). This indicates that pre-service teachers are more concerned about the impact of AI on their own and their future students' thinking processes than on ethical or technical security issues. The interview findings were consistent with the form findings. All pre-service teachers expressed concerns regarding the risks of "inaccurate or biased information." In addition, concerns were expressed that AI may negatively affect cognitive processes by encouraging "cognitive laziness" and pose risks related to "data privacy and security." Issues related to creativity skills were also raised in connection with cognitive laziness. Some views focusing on AI accuracy and cognitive laziness are as follows:

> *"I do not fully trust the accuracy of AI, definitely. If it is something very general or common knowledge, I tend to trust it; however, if it is something important, I avoid relying on a single source and cross-check the information. I definitely cannot say that I trust it completely." (P1)*

> *"I think we should be trained on how to use AI correctly. Otherwise, it can really make people very lazy. For example, I use it for my assignments and ask it to explain things to me, but I could also have it do the work directly—and sometimes I do. When I feel lazy, I directly upload the assignment and ask it to complete it. This makes me lazy." (P2)*

4. In Conclusion

The findings demonstrate that all participating pre-service teachers possess a high level of AI awareness and utilize these technologies for diverse academic purposes. However, evidence from both the survey and interview data highlights a critical consciousness regarding the potential risks and ethical concerns associated with AI. These results confirm a significant need for training, emphasizing that pre-service teachers require structured support to effectively integrate these tools into their future pedagogical practices. Moreover, it highlights that AI literacy should also be embedded at earlier stages of education, starting from K–12, to ensure that future generations develop the necessary competencies from an early age. Consequently, this study suggests the integration of AI literacy into teacher education curricula to better equip future educators for the demands of AI-enhanced classrooms.

In future research, it is recommended to examine the contextual integration of AI tools into teaching across different disciplines. Additionally, investigating the impact of AI literacy training programs on the pedagogical competencies of pre-service teachers would provide valuable insights. Furthermore, developing systematic integration models for incorporating AI tools into teacher education programs is suggested.

Acknowledgements

The author's participation in ECEL 2026 was supported by TÜBİTAK under the 2224-A Grant Program for Participation in Scientific Meetings Abroad

Ethics Declaration

Ethics committee approval was obtained from Bayburt University Ethics Committee (No: E-19460208-200-364505).

AI Declaration

During the preparation of this work the authors used DeepL and ChatGPT to translate, improve language clarity and check grammar. After using this tool, the authors reviewed and edited the content and take full responsibility for the content of the published article.

References

Almassaad, A., Alajlan, H., & Alebaikan, R. (2024) "Student perceptions of generative artificial intelligence: Investigating

utilization, benefits, and challenges in higher education", *Systems,* Vol 12, No. 10, pp 385. https://doi.org/10.3390/systems12100385

Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024) "AI in education: A review of personalized learning and educational technology", *GSC Advanced Research and Reviews,* Vol 18, No. 2, pp 261-271

Babu, A., & Boddu, S. B. (2024) "Bert-based medical chatbot: Enhancing healthcare communication through natural language understanding", *Exploratory Research in Clinical And Social Pharmacy,* Vol 13, 100419

Bozkurt, A. (2023) "Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift", *Asian Journal of Distance Education,* Vol 18, No. 1. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718

Chaamwe, N. (2025) "Investigating the factors influencing students' adoption of generative AIs in universities: A case of the Copperbelt University", *Zambia ICT Journal,* Vol 8, No. 1, pp 47–53. https://doi.org/10.33260/zictjournal.v8i1.340

Daepp, M. I., & Counts, S. (2024) "The emerging AI divide in the United States", *ArXiv* preprint arXiv:2404.11988. https://doi.org/10.48550/arXiv.2404.11988.

Dartmouth. (1956) *Artificial intelligence coined at Dartmouth*. https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth

Davis, F. D. (1989) "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly,* Vol 13, *No.* 3, pp 319-340. https://doi.org/10.2307/249008

Grover, D. (2024) "Next-generation education: The impact of generative AI on learning", *Journal of Informatics Education and Research,* Vol 4, No. 2. https://doi.org/10.52783/jier.v4i2.1019.

Gupta, R., Nair, K., Mishra, M., Ibrahim, B., & Bhardwaj, S. (2024) "Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda", *International Journal of Information Management Data Insights,* Vol 4, No. 1, 100232. https://doi.org/10.1016/j.jjimei.2024.100232

Hsu, Y. C., & Ching, Y. H. (2023) "Generative artificial intelligence in education, part one: The dynamic frontier. TechTrends, Vol 67, No. 4, pp 603-607. https://doi.org/10.1007/s11528-023-00863-9

OECD. (2018) *Bridging the rural digital divide (OECD Digital Economy Papers, No. 265).* OECD Publishing. https://doi.org/10.1787/852bd3b9-en

Pitts, G., Marcus, V., & Motamedi, S. (2025) "Student perspectives on the benefits and risks of AI in education", *arXiv* preprint arXiv:2505.02198. https://doi.org/10.48550/arXiv.2505.02198

Scarpi, D. (2024) "Strangers or friends? Examining chatbot adoption in tourism through psychological ownership", *Tourism Management,* Vol 102*,* No. 104873.

UN. (2022) *Sustainable Development Goals (SDGs).* https://sdgs.un.org/goals/goal4

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003) "User acceptance of information technology: Toward a unified view", *MIS Quarterly,* Vol 27, No. 3, pp 425–478. https://doi.org/10.2307/30036540

Yakut, S. G., Aslan, H. K., & Küsen, G. Y. (2025) "Perspective on artificial intelligence: A profile study on the attitudes of university students", *Journal of Awareness,* Vol 10, No. 1, e2684-e2684. https://doi.org/10.26809/joa.2684

From Potential to Practice: Challenges of Integrating Artificial Intelligence Technologies in School Classrooms

Norah Almusharraf[1] and Muhammad Imran[1,2]
[1]Prince Sultan University, Riyadh, Saudi Arabia
[2]Khazar University, Baku, Azerbaijan

nmusharraf@psu.edu.sa

Abstract: This study investigates the integration of Artificial Intelligence Technologies (AIT) into classroom teaching, focusing on three factors: teachers' access to AIT tools, the frequency of their use, and the key challenges and barriers to integrating AIT for e-learning. A quantitative, survey-based approach was employed to address the study's research questions. A 25-item questionnaire was administered to 90 elementary and secondary school teachers, and the data were analysed descriptively and correlationally using SPSS (version 23). The findings reveal that teachers have limited access to various AIT applications, receive inadequate training, and face infrastructural challenges. More than half of the teachers (54%) reported using the Internet to prepare lessons and carry out other teaching activities. Despite recognizing AIT's potential to enhance teaching effectiveness, educators reported attitudinal and systemic barriers to its integration. This study offers original evidence on AIT integration in elementary and secondary classrooms in Pakistan, a context that remains underrepresented in the literature, by examining teachers' actual access, usage patterns, and the systemic barriers they face, and by situating these findings within established technology-adoption frameworks to offer policy-relevant insights.

Keywords: Artificial Intelligence Technologies (AIT); Technology Adoption Barriers; Classroom Technology; Teacher Technical Knowledge; E-learning

1. Introduction

In this technologically advanced era, the integration of Artificial Intelligence Technologies (AIT) in education has emerged as a transformative force, reshaping traditional pedagogical paradigms and redefining the dynamics of teaching and learning (Marengo & Pange, 2024; Bashir et al., 2025; Imran et al., 2024a, 2024b; Maqbool et al., 2024). AIT integration can improve learners' outcomes while also helping to minimize technology access and equity gaps that are necessary to address for success in the 21st century (Berisha Qehaja, 2025; Akinoso, 2023). While developed nations have embraced AI as a cornerstone of modern education, developing countries such as Pakistan face distinct challenges in harnessing its potential (Shoukat, 2024), including inadequate technological infrastructure, overcrowded classrooms, and constrained annual education budgets (Mannuru et al., 2023; Alam et al., 2023). As the application of AIT in teaching and learning expands worldwide, the need to investigate the challenges faced by students and teachers, the adequacy of study materials, and the pedagogical techniques required to build effective, AIT-integrated learning environments has grown correspondingly (Hamzah et al., 2024; Gull et al., 2020).

This research critically examines AIT integration in elementary and secondary school classroom practices for e-learning, exploring how technology can be strategically incorporated into teaching methodologies to bridge educational disparities, enhance student engagement, and foster 21st-century skills. By examining the infrastructural, attitudinal, and policy-related barriers to AIT adoption, this study aims to propose context-sensitive solutions that align with global trends and address the specific needs of the Pakistani educational landscape. In doing so, it aims to contribute to the ongoing discourse on educational innovation by offering a grounded perspective on how AIT can catalyse equitable and sustainable educational reform in resource-constrained settings. This study further assesses teachers' technical knowledge (TK), that is, their ability to apply information and communication technology tools to classroom activities and everyday professional tasks after developing a working understanding of AIT tools.

Specifically, this study addresses the following research questions: (1) To what extent do teachers have access to AIT tools, and what is their level of technical knowledge in using them? (2) How frequently do teachers integrate AIT tools into classroom activities? (3) What key challenges and barriers hinder the integration of AIT in Pakistani school classrooms?

2. Literature Review

A growing body of research demonstrates that the integration of AIT into educational systems has significantly transformed teaching methodologies worldwide (Tam et al., 2017; Gull et al., 2020; Huang et al., 2024; Maqbool et al., 2024; Asad et al., 2021). Over the past five years in particular, institutions have seen a marked increase in

research examining the role of AIT in classrooms, particularly in developing countries such as Pakistan (Butt et al., 2020). For instance, Aldhafeeri and Alotaibi (2023) proposed a new educational model, Digital Education Shifting (DES), to support quality digital learning in K-12 education. Related evidence indicates that AIT can increase student engagement, encourage collaborative learning, and provide access to extensive resources, thereby improving educational outcomes (Zaremohzzabieh et al., 2024).

Similarly, Meace-Williams (2024) and Wilson (2023) argue that effective AIT integration cannot occur without a corresponding shift in pedagogy: teachers can no longer treat digital tools merely as peripheral aids but must come to view them as instruments capable of transforming teaching practice. These studies further emphasize the importance of sustained professional development programs, given the central role that teachers' AIT-related skills play in successful integration.

To support e-learning trends, Maqbool et al. (2024) and Gull et al. (2020) investigated emerging trends in e-learning, with particular attention to teachers' perceptions of technology for collaborative learning without physical boundaries; they found that fuzzy-interface systems could enhance e-learning both within and outside the formal classroom while also promoting collaboration. AIT also offers advantages for accommodating diverse learning styles: technology can enable personalized learning experiences tailored to individual students' needs (Van Deursen et al., 2020). More broadly, it is widely recognized that education is shifting from traditional chalk-and-talk instruction toward interactive, technology-enabled learning models. In a systematic review, Hsu, Ching, and Grabowski (2013) emphasize that technology-integrated learning environments become more interactive, participatory, and student-centred, qualities that are necessary for developing critical thinking and problem-solving skills.

The above-discussed literature demonstrates broad, cross-national interest in AIT integration; however, much of it remains descriptive, cataloging adoption trends or barriers without engaging critically with the theoretical mechanisms that explain why some teachers adopt AIT while others do not. Established technology-adoption frameworks offer a useful lens for interpreting these patterns. The Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) emphasize perceived usefulness, perceived ease of use, and facilitating conditions as key predictors of technology adoption, while the Technological Pedagogical Content Knowledge (TPACK) framework highlights the interaction between teachers' technological, pedagogical, and content knowledge as a precondition for meaningful integration. Few studies (Barakabitze et al., 2019; Alam et al., 2023; Maqbool et al., 2024), however, have applied these frameworks to elementary and secondary school contexts in developing countries, and fewer still have examined Pakistan specifically, where the infrastructural constraints and teacher-training gaps documented by Butt et al. (2020) and Shoukat (2024) may interact with facilitating conditions differently than in the higher-education or developed-country settings that dominate the existing literature. This gap motivates the present study, which draws on TAM, UTAUT, and TPACK to examine teachers' access to AIT tools, their technical knowledge, and the structural and attitudinal barriers that shape AIT integration in Pakistani elementary and secondary classrooms.

3. Methods

This study employed a quantitative, questionnaire-based approach to examine the integration of AIT in classroom teaching at primary and elementary school levels.

3.1 Participants and Sampling

A stratified random sampling technique was used to select participants from ***18 primary and elementary schools*** from both rural and urban locations. Of the 100 teachers initially selected, 110 questionnaires were distributed. Ninety-six questionnaires were returned (87.3% response rate), and 90 were retained for analysis after six incomplete responses were excluded (81.8% usable response rate). Figure 1 presents the participant recruitment flowchart.

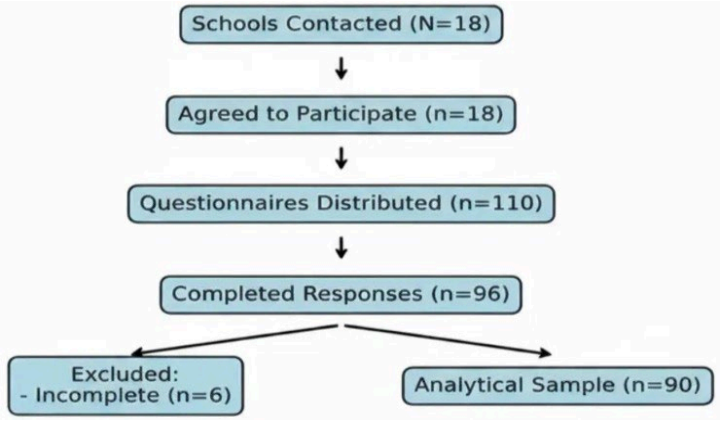


Figure 1: Participants' Recruitment Flowchart

Questionnaires were distributed without discrimination based on gender, qualification, teaching experience, or subject specialization. Nonetheless, as reported in Section 4.1, the achieved sample was heavily skewed toward male respondents (91.11%); this imbalance was not by design and is addressed as a limitation of the study in Section 5.3.

3.2 Instrument Development and Validation

The questionnaire consisted of 25 items organized into three sections: (a) teachers' access to Artificial Intelligence Technologies (AIT) tools and technical knowledge, (b) the frequency of AIT use in classroom activities, and (c) perceived barriers to AIT integration. Responses were measured on a five-point Likert scale ranging from 1 (Never/Strongly Disagree) to 5 (All the Time/Strongly Agree). The questionnaire was developed following an extensive review of the literature and adapted from previous studies on artificial intelligence technologies, educational technology integration, and teachers' technology adoption. The questionnaire was self-developed based on an extensive review of the literature on AIT integration, teacher technology adoption, and digital competence. To establish content validity, the instrument was reviewed by two experts in educational technology and teacher education, who evaluated the relevance, clarity, and comprehensiveness of the items. Their recommendations were incorporated into the final questionnaire before administration. The instrument was also pilot-tested with 10 teachers who were not included in the main study to ensure the clarity, readability, and feasibility of the questionnaire.

3.3 Data Analysis

Data were analysed descriptively using SPSS (version 23), including frequency counts, percentages, means, and standard deviations. Spearman's rank-order correlation was used to examine the strength and direction of association between the five AIT adoption activities reported in Table 2.

3.4 Ethical Considerations

This study was conducted in accordance with the ethical research standards of Prince Sultan University, which granted approval prior to data collection. Participation was voluntary, and all participants provided informed consent after being briefed on the purpose of the study, their right to withdraw at any time without penalty, and the measures taken to protect the anonymity and confidentiality of their responses. No personally identifiable information was collected, and no minors, patients, or animals were involved in this study.

4. Findings and Discussion

Data collected from the 90 respondents were analysed descriptively, using frequencies, percentages, means, and standard deviations, and correlationally, using Spearman's rank-order correlation to examine relationships among AIT adoption activities.

4.1 Descriptive Statistics

The demographic profile of respondents is an important consideration in interpreting the findings of this study. Of the 90 teachers surveyed, 82 (91.11%) were male, and 8 (8.89%) were female. Regarding teaching experience,

43 teachers (48%) reported 6–10 years, 25 (28%) reported more than 10 years, 12 (13%) reported less than 1 year, and 10 (11%) reported 1–5 years. Regarding school location, most teachers (55; 61.11%) were based in urban areas, compared with 35 (38.89%) in rural areas. Most respondents (61; 67.78%) taught at the elementary level, while 29 (32.22%) taught at the secondary level. Regarding highest qualification, 61 teachers (68%) held a bachelor's degree, 10 (11%) held a master's degree, and 19 (21%) held an M.Phil.

4.2 Teachers' Access to AIT Tools and Technical Knowledge (TK)

Table 1 presents the availability of AIT tools and teachers' self-reported level of knowledge in using them. Most teachers reported access to a smartphone with an internet connection (97.78%) and the ability to open email messages (97.78%), while 72.22% reported having a PC or laptop at home. Only 28.33% of teachers reported being able to use an AI tool such as ChatGPT for lesson planning, despite 95.55% reporting that they could search for a topic using ChatGPT, DeepSeek, or other AI chatbots — a gap suggesting that many teachers are comfortable with general-purpose AI search but have not yet adopted AI tools specifically for instructional planning. Similarly, while 75.55% of teachers could install apps from the Play Store and 64.44% could upload files to the internet, only 41.11% reported being able to create a presentation using MS PowerPoint, indicating a specific gap in presentation-software skills relative to more general digital literacy.

Table 1: Accessibility of Participants to AIT Tools and Technical Knowledge

No.	Item	Frequency	Percentage (%)
1	I have a smartphone with an internet connection.	88	97.78
2	I have a PC/laptop at home.	65	72.22
3	I can open my email messages.	88	97.78
4	I can use an AI tool (e.g., ChatGPT) for lesson planning.	25	28.33
5	I can install apps from the Play Store on my phone or computer.	68	75.55
6	I can search for a topic using ChatGPT, DeepSeek, or other AI chatbots.	86	95.55
7	I can write a suitable prompt to create images using AI tools.	50	56.66
8	I can upload files to the internet.	58	64.44
9	I can create a presentation using MS PowerPoint.	37	41.11
10	I can compose a letter using a word processor.	58	64.44

Table 2 reports the frequency of use of specific AI-powered tools and digital platforms for instructional preparation, captured on a five-point Likert scale (Never, Rarely, Sometimes, Often, All the time). The most prevalent activity was "Use of ChatGPT to prepare lessons" (Activity 1), with the largest share of respondents (54.44%) reporting "Sometimes" and a mean frequency of M = 2.77 (SD = 0.86). A similar pattern of moderate, occasional use emerged for "Use of AI applications to prepare a presentation for lessons" (Activity 3), with 60.00% responding "Sometimes" and a nearly identical mean of M = 2.78 (SD = 0.92). Together, these results suggest that ChatGPT and comparable AI presentation tools are becoming embedded, though not yet dominant, components of lesson planning for a substantial share of teachers, with moderate variability in usage intensity across the sample. "Download/upload/browse material from a learning platform" (Activity 4) recorded the highest mean (M = 3.00, SD = 1.10), while "Look for online professional development opportunities" (Activity 5) recorded the lowest (M = 2.60, SD = 0.97), indicating that teachers engage somewhat more consistently with retrieving instructional material than with pursuing their own professional development online.

Table 2: Frequency and Percentage of AIT Integration in Classroom Activities

No.	Activity	Never	Rarely	Sometimes	Often	All the time	Mean	S.D.
1	Use of ChatGPT to prepare lessons	7 (7.77%)	22 (24.44%)	49 (54.44%)	9 (10.00%)	3 (3.33%)	2.77	0.86
2	Use of AI chatbots to collect resources for lessons	15 (16.66%)	10 (11.11%)	44 (48.88%)	14 (15.55%)	7 (7.77%)	2.89	1.21
3	Use of AI applications to prepare a presentation for lessons	14 (15.55%)	9 (10.00%)	54 (60.00%)	9 (10.00%)	4 (4.44%)	2.78	0.92
4	Download/upload/browse material from a learning platform	9 (10.00%)	14 (15.55%)	43 (47.77%)	17 (18.89%)	7 (7.77%)	3.00	1.10
5	Look for online professional development opportunities	12 (13.33%)	25 (27.77%)	35 (38.88%)	14 (15.55%)	4 (4.44%)	2.60	0.97

Figure 2 shows the Spearman's rank correlation coefficients between the five AIT adoption activities. Each coefficient measures the strength and direction of a monotonic relationship between activity pairs, with values close to +1 indicating a very strong positive relationship; diagonal values equal 1.00 by definition, reflecting the perfect self-correlation of each activity. All off-diagonal coefficients exceed 0.86, reflecting consistently strong positive associations among the AIT adoption activities in this sample.

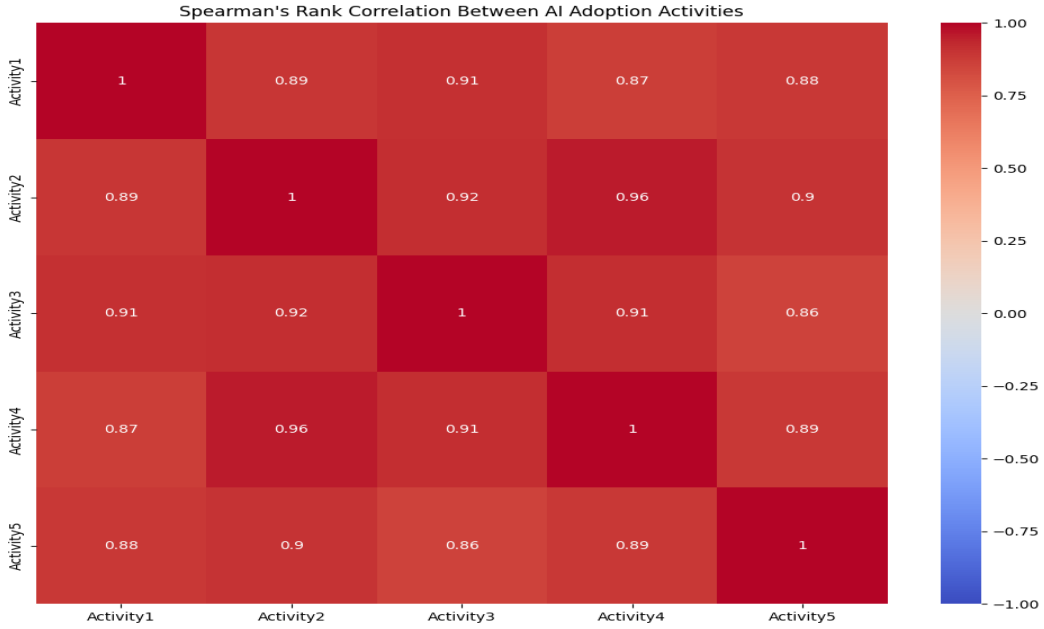


Figure 2: Spearman Correlation Between AIT Adoption Activities

The strongest relationship was between Activity 2 (use of AI chatbots to collect lesson resources) and Activity 4 (downloading/uploading/browsing material from a learning platform), at 0.96. In practical terms, this suggests that these two behaviours are not adopted independently: a teacher who regularly turns to an AI chatbot for resources is, in this sample, almost certain to also be a regular user of online learning platforms, which is consistent with both behaviours drawing on the same underlying digital-resource-seeking habit rather than on distinct skill sets. Activity 3 (AI-assisted presentation preparation) maintained consistently high correlations with all other activities (0.86–0.92), suggesting it functions as a central, integrative behaviour within teachers' broader pattern of AIT use. The lowest observed correlation, between Activity 3 and Activity 5 at 0.86, is still a strong positive association, indicating that even the least closely related activities in this dataset tend to co-occur. Because all five activities are drawn from the same respondents and rated on the same underlying construct of AIT engagement, these uniformly high correlations should be interpreted with some caution: they may partly reflect a general disposition toward AIT use (or toward more positive self-report) rather than five fully independent behaviours, a possibility we return to in the Limitations (Section 5.3).

Beyond adoption patterns, the analysis identifies several barriers that impede AIT integration in schools (Table 3). Organisational constraints were the most strongly endorsed barriers: pressure to prepare students for examinations (M = 2.94, SD = 1.05) and inflexible school schedules (M = 2.85, SD = 0.93) left teachers little room to experiment with AIT tools during instructional time. Infrastructural barriers were also prominent, particularly a lack of internet-connected computers (M = 2.69, SD = 0.91), while the sheer shortage of computers was rated as a comparatively more moderate concern (M = 2.32, SD = 0.79). Teachers were similarly dissatisfied with technical support (M = 2.33, SD = 1.02) and, to an even greater extent, with pedagogical support for AIT integration (M = 2.14, SD = 0.91), suggesting that the absence of in-school pedagogical guidance is felt more acutely than the absence of technical troubleshooting. Attitudinal barriers were comparatively less pronounced: teachers' own attitudes toward AIT (M = 2.02, SD = 0.92) and their perception that AIT offers limited benefit (M = 2.42, SD = 1.28) were rated lower than the structural barriers above, suggesting that teachers are not, in the main, resistant to AIT on personal or attitudinal grounds; rather, structural and organisational constraints appear to be the primary obstacle to integration in this sample.

Table 3: Perceived Barriers to AIT Integration

Barrier	Mean	S.D.
Insufficient number of computers	2.32	0.79
Lack of internet-connected computers	2.69	0.91
Inadequate technical support	2.33	1.02
Inadequate pedagogical support	2.14	0.91
Inflexible school schedules	2.85	0.93
Pressure to prepare students for examinations	2.94	1.05
Negative teacher attitudes toward AIT	2.02	0.92
Teachers' lack of interest in AIT	2.59	1.21
Low perceived benefit of AIT in teaching	2.42	1.28
Low institutional prioritisation of AIT	2.02	0.98

5. Discussion

5.1 Interpreting the Findings

Consistent with Butt et al. (2020) and Shoukat (2024), the findings show that structural and organisational barriers — inflexible schedules, examination pressure, and inconsistent access to internet-connected computers — are rated as more significant obstacles to AIT integration than teachers' own attitudes toward the technology. This pattern aligns with the facilitating-conditions component of UTAUT, which holds that even motivated users will not adopt a technology consistently if the organisational and infrastructural conditions for doing so are absent. In our data, teachers reported comparatively low negative attitudes toward AIT (M = 2.02) alongside comparatively high organisational barriers (M = 2.85–2.94), indicating that the barrier to integration in this sample lies less in teacher willingness and more in facilitating conditions — a distinction with direct implications for policy, since it suggests that attitude-change campaigns alone are unlikely to be as effective as investment in infrastructure and scheduling flexibility.

The moderate-but-inconsistent usage pattern reported in Table 2 (means of 2.60–3.00 on a five-point scale, all clustering around "Sometimes") is also consistent with a TPACK-informed reading of the data: teachers appear to possess sufficient technical knowledge to use AIT tools occasionally, but the absence of dedicated pedagogical support (the lowest-rated item in Table 3, M = 2.14) suggests a gap in the pedagogical and technological-pedagogical knowledge components of TPACK, rather than in technical knowledge alone. This may explain why general-purpose AI search (95.55% able to use ChatGPT/DeepSeek to search a topic, Table 1) is far more

widespread than AIT use specifically for lesson planning (28.33%): teachers appear to have crossed the threshold of basic technical competence but not yet the further threshold of integrating AIT into pedagogical design.

The uniformly strong Spearman correlations among the five adoption activities (all $r > 0.86$) extend, rather than merely replicate, prior descriptive findings on AIT adoption trends (Maqbool et al., 2024; Gull et al., 2020): they indicate that in this sample, AIT adoption behaviours do not develop independently but cluster together, consistent with the view that teachers develop a general disposition toward digital-resource use rather than adopting individual tools in isolation. This is a significant contribution relative to prior studies, which have typically reported adoption rates or barriers item-by-item without examining how these behaviours relate to one another.

5.2 Contribution

This study extends the existing literature in three ways. First, it provides original survey evidence from elementary and secondary school teachers, a population that remains underrepresented relative to the higher-education focus of much prior AIT research (Bashir et al., 2025; Imran et al., 2024a, 2024b). Second, by combining descriptive statistics on access, technical knowledge, and usage with a correlational analysis of how adoption behaviours relate to one another, it moves beyond a simple inventory of barriers toward an account of how these barriers and behaviours are structured. Third, by interpreting these patterns through TAM, UTAUT, and TPACK, it offers a theoretical account of why structural and pedagogical-support barriers, rather than teacher attitudes, appear to be the binding constraint on AIT integration in this context — a finding with direct implications for how limited resources should be allocated by policymakers and school administrators.

5.3 Limitations

This study has several limitations. First, the sample was drawn from the primary and elementary levels and was heavily skewed toward male teachers (91.11%), which limits the generalisability of the findings to other regions and to more gender-balanced teaching populations. Second, the study relies exclusively on self-reported, cross-sectional survey data; actual classroom observation or log data on AIT use were not collected, and no qualitative data (e.g., interviews or classroom observations) were gathered, so the findings should be read as teachers' perceptions of their access, use, and barriers rather than as directly observed behaviour. Third, because all five adoption activities in Table 2 were measured on the same self-report instrument, the uniformly high correlations between them may be partly inflated by common-method variance rather than reflecting five fully independent behaviours. Future research, discussed further in Section 6.2, should address these limitations through multi-region sampling, gender-balanced recruitment, and the incorporation of qualitative or observational data.

6. Conclusion

This study examined the current state of AIT integration in school classrooms. More than half of the teachers surveyed (54%) reported using the Internet to prepare lessons and carry out other teaching activities, yet only 28.33% reported using AI tools specifically for lesson planning, and opportunities for professional development in this area remain limited. The findings reveal moderate but inconsistent engagement with AIT tools, driven less by teachers' attitudes than by inconsistent infrastructural access, insufficient pedagogical support, and organisational constraints such as inflexible schedules and examination pressure. Addressing these gaps, particularly through infrastructure investment, continuous training, and dedicated pedagogical support, has the potential to improve teachers' capacity to integrate AIT and, in turn, to improve educational outcomes in school classrooms.

6.1 Recommendations

The findings of this study suggest that stakeholders should prioritise equitable access to AIT tools, including reliable internet-connected devices and learning platforms, across both urban and rural areas. Mandatory training programmes focused specifically on AI-tool use and broader AIT skills should be introduced, and schools should establish dedicated technical and pedagogical support teams to assist teachers in troubleshooting and integrating AIT tools into classroom activities. Given that examination pressure and inflexible scheduling were rated among the most significant barriers, school administrations should also consider building dedicated time for AIT-supported lesson preparation into teachers' schedules.

6.2 Future Implications

This research highlights several directions for future work. The study could be replicated across other regions, with a more gender-balanced sample, to compare urban–rural disparities and the effects of national policy. A longitudinal design would help assess the long-term effects of AIT integration on student engagement and performance, while a mixed-methods extension incorporating classroom observation or teacher interviews would help validate the self-reported patterns identified here. Future research should also examine the role of school administration and leadership in driving AIT adoption through blended learning models.

Acknowledgments

The authors would like to thank the Educational Research Lab at Prince Sultan University for financial and technical support.

Ethics Declaration

This study followed all applicable ethical principles for research with human participants; no animal subjects were involved. See Section 3.4 for full details of the ethical approval and consent procedures.

AI Declaration

No AI tool has been used to write this draft. However, we used Grammarly Premium for proofreading and language improvement in this study.

Conflict Declaration

The authors declare no conflict of interest.

References

Akinoso, S. O. (2023). Motivation and ICT in secondary school mathematics using unified theory of acceptance and use of technology model. Indonesian Journal of Educational Research and Technology, 3(1), 79-90. https://doi.org/10.17509/ijert.v3i1.47183

Alam, M. J., Hassan, R., & Ogawa, K. (2023). Digitalization of higher education to achieve sustainability: Investigating students' attitudes toward digitalization in Bangladesh. International Journal of Educational Research Open, 5, 100273. https://doi.org/10.1016/j.ijedro.2023.100273

Aldhafeeri, F. M., & Alotaibi, A. A. (2023). Reimagining education for successful and sustainable digital shifting. Sage Open, 13(1), 21582440231154474. https://doi.org/10.1177/21582440231154474

Asad, M. M., Hussain, N., Wadho, M., Khand, Z. H., & Churi, P. P. (2021). Integration of e-learning technologies for interactive teaching and learning process: an empirical study on higher education institutes of Pakistan. Journal of Applied Research in Higher Education, 13(3), 649-663. https://doi.org/10.1108/jarhe-04-2020-0103

Barakabitze, A. A., William-Andey Lazaro, A., Ainea, N., Mkwizu, M. H., Maziku, H., Matofali, A. X., Iddi, A., & Sanga, C. (2019). Transforming African education systems in science, technology, engineering, and mathematics (STEM) using ICTs: Challenges and opportunities. Education Research International, 2019(1), 6946809. https://doi.org/10.1155/2019/6946809

Bashir, A., Aziz, A., Imran, M., & Almusharraf, N. (2025). Effect of CALL-based multimodal pedagogy on learner motivation and willingness to communicate in English: A study from university students' perspective. Contemporary Educational Technology, 17(2). https://doi.org/10.30935/cedtech/15961

Berisha Qehaja, A. (2025). Strategic integration of artificial intelligence solutions to transform teaching practices in higher education. Foresight. https://doi.org/10.1108/fs-04-2024-0079

Butt, R., Siddiqui, H., Soomro, R. A., & Asad, M. M. (2020). Integration of Industrial Revolution 4.0 and IOTs in academia: a state-of-the-art review on the concept of Education 4.0 in Pakistan. Interactive Technology and Smart Education, 17(4), 337-354. https://doi.org/10.1108/itse-02-2020-0022

Gull, A., Imran, M., Yiunas, M., & Afzaal, M. (2020). A descriptive study of challenges faced by English language teachers in integrating information and communication technology (ICT) tools at elementary level in Pakistan. International Journal of Advanced Science and Technology, 29(8), 290-305.

Hamzah, F., Abdullah, A. H., & Ma, W. (2024). Advancing education through Technology Integration, innovative pedagogies and emerging trends: A systematic literature review. Journal of Advanced Research in Applied Sciences and Engineering Technology, 41(1), 44-63. https://doi.org/10.37934/araset.41.1.4463

Hsu, Y.-C., Ching, Y.-H., & Grabowski, B. (2013). A review of Web 2.0 tools' design to promote reciprocal teaching, cognitive presence, and cognitive apprenticeship in online learning environments. Journal of Interactive Learning Research, 24(2), 195–221.

Huang, R., Liu, D., Kanwar, A. S., Zhan, T., Yang, J., Zhuang, R., Liu, M., Li, Z., & Adarkwah, M. A. (2024). Global understanding of smart education in the context of digital transformation. Open Praxis, 16(4), 663-676. https://doi.org/10.55982/openpraxis.16.4.761

Imran, M., Almusharraf, N., & Abbasova, M. Y. (2025). Digital learning transformation: A study of teachers' post-Covid-19 experiences. Social Sciences & Humanities Open, 11, 101228. https://doi.org/10.1016/j.ssaho.2024.101228

Imran, M., Almusharraf, N., Abdellatif, M. S., & Ghaffar, A. (2024a). Teachers' perspectives on effective English language teaching practices at the elementary level: A phenomenological study. Heliyon, 10(8). https://doi.org/10.1016/j.heliyon.2024.e29175

Imran, M., Almusharraf, N., Ahmed, S., & Mansoor, M. I. (2024b). Personalization of E-Learning: Future Trends, Opportunities, and Challenges. International Journal of Interactive Mobile Technologies, 18(10). https://doi.org/10.3991/ijim.v18i10.47053

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., & Galley, J. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. Information Development, 02666669231200628. https://doi.org/10.1177/02666669231200628

Marengo, A., & Pange, P. J. (2024). Envisioning general AI in higher education: transforming learning paradigms and pedagogies. The Learning Ideas Conference. https://doi.org/10.1007/978-3-031-72430-5_28

Maqbool, M. A., Asif, M., Imran, M., Bibi, S., & Almusharraf, N. (2024). Emerging e-learning trends: a study of faculty perceptions and impact of collaborative techniques using fuzzy interface system. Social Sciences & Humanities Open, 10, 101035. https://doi.org/10.1016/j.ssaho.2024.101035

Meace-Williams, S. (2024). Conceptualising Welsh Secondary School Teachers' Pedagogic Use of Technology: A Case Study [Doctoral dissertation, Cardiff Metropolitan University].

Shoukat, R. (2024). Harnessing AI for Educational Transformation: A Comparative Study of China, India and Pakistan. Strategic Studies, 44(1). https://doi.org/10.53532/ss.044.01.00352

Wilson, M. L. (2023). The impact of technology integration courses on preservice teacher attitudes and beliefs: A meta-analysis of teacher education research from 2007–2017. Journal of Research on Technology in Education, 55(2), 252-280. https://doi.org/10.1080/15391523.2021.1950085

Zaremohzzabieh, Z., Ahrari, S., Abdullah, H., Abdullah, R., & Moosivand, M. (2024). Effects of educational technology intervention on creative thinking in educational settings: a meta-analysis. Interactive Technology and Smart Education. https://doi.org/10.1108/itse-11-2023-0224

Beyond AI Use: Understanding University Students' Acceptance of AI-Assisted Academic Misuse

Basil Alzougool and Sumayya Banna
Arab Open University, Kuwait Branch, Kuwait

balzougool@aou.edu.kw
sumayyaa@aou.edu.kw

Abstract: The rapid adoption of generative artificial intelligence (AI) in higher education has created new learning opportunities while raising concerns about academic integrity and students' acceptance of AI-assisted academic misuse. This study examined factors associated with such acceptance among 304 university students in Kuwait using a cross-sectional online survey and non-probability convenience and snowball sampling. The three focal constructs—general academic-integrity awareness, policy and AI-plagiarism awareness, and acceptance of AI-assisted academic misuse—were measured using 15 dichotomous Yes/No items coded 0 = No and 1 = Yes. Acceptance of AI-assisted academic misuse was operationalised as the mean of three misuse-acceptance items (Y12-Y14), producing a composite ranging from 0 to 1 that was analysed as a continuous outcome. Data were analysed using descriptive statistics, principal-axis exploratory factor analysis with varimax rotation, KR-20 reliability analysis, correlations, and ordinary least squares regression with HC3 robust standard errors. The regression model was significant ($F = 2.579$, $p = .010$, $R^2 = .067$). Policy and AI-plagiarism awareness was positively associated with misuse acceptance ($B = .193$, $p = .013$), contrary to the hypothesised negative direction, and Kuwaiti nationality was also associated with higher acceptance ($B = .103$, $p = .039$). General academic-integrity awareness and prior AI use were not significant. Because the study is cross-sectional and the misuse-acceptance scale had modest reliability (KR-20 = .587), the findings are interpreted as associations rather than causal effects. The results support clearer, context-specific AI policies and guidance that distinguishes legitimate AI-supported learning from academically inappropriate use.

Keywords: generative Artificial Intelligence; Academic Integrity; AI-Assisted Academic Misuse; AI Policy Awareness; Responsible AI Use; Higher Education; Kuwait.

1. Introduction

The integration of AI chat assistants into higher education has transformed how students access information, engage with content, and complete academic tasks. These tools support legitimate activities including idea generation, concept explanation, revision, and personalised learning. However, their use has raised concerns about plagiarism, unauthorised content generation, assessment validity, dependence on AI-produced material, and reduced critical thinking. Universities must therefore balance educational benefits with academic standards (Cotton et al., 2024; Sallam et al., 2024).

AI-assisted academic misuse occurs when students use AI in ways that violate academic-integrity principles or institutional assessment rules. Examples include submitting AI-generated work as original, using AI without permission or acknowledgement, disguising plagiarism through AI-assisted paraphrasing, generating assignments, or using AI during prohibited assessments. Acceptability depends on purpose, context, disclosure, and institutional requirements. Students' acceptance matters because those who regard inappropriate use as permissible may be more willing to engage in it when rules are unclear or detection is difficult.

General academic-integrity awareness may guide students' judgements by reinforcing originality, acknowledgement, independent work, and compliance with assessment requirements. However, traditional integrity guidance may not resolve AI-specific practices such as automated text generation, extensive paraphrasing, editing, and undisclosed assistance. Awareness of institutional AI policies and AI-related plagiarism may therefore be especially important for distinguishing legitimate support from misconduct (Lund et al., 2025; Khairullah et al., 2025).

Prior AI use may also shape acceptance: familiarity can clarify appropriate boundaries, but repeated use may normalise dependence or blur distinctions between support and misconduct. Demographic characteristics may reflect differences in educational experience, policy exposure, and technology use, although existing findings are mixed (Al Omari et al., 2024; Deng et al., 2025; Syed et al., 2025).

Three gaps motivate this study. First, Kuwait and Gulf higher-education contexts remain underrepresented. Second, much prior research examines AI adoption, usefulness, or learning benefits rather than acceptance of misuse. Third, few studies consider general academic-integrity awareness, AI-specific policy/plagiarism

awareness, prior AI use, and demographic factors together. This study therefore examines how these factors are associated with university students' acceptance of AI-assisted academic misuse in Kuwait.

2. Literature Background

AI chat assistants such as ChatGPT, Gemini, Microsoft Copilot, Claude, and Perplexity AI use large language models and natural-language processing to generate conversational responses, summaries, explanations, and other content. In higher education, they can support concept explanation, idea generation, writing improvement, literature summarisation, feedback, and programming, potentially improving accessibility, productivity, and personalised learning (Alqahtani et al., 2023; Sajja et al., 2024; Khairullah et al., 2025).

These benefits coexist with academic-integrity risks. Students may use AI to generate assessed work, paraphrase copied material, obtain unauthorised assistance, or submit AI-generated content as their own. Additional concerns include inaccurate information, fabricated references, bias, privacy risks, and overreliance that may weaken independent critical thinking and writing (AlBlooshi, 2026; Alotaibi, 2024; Cotton et al., 2024). AI-assisted academic misuse therefore refers to AI use that conflicts with academic-integrity standards, assessment requirements, or institutional rules.

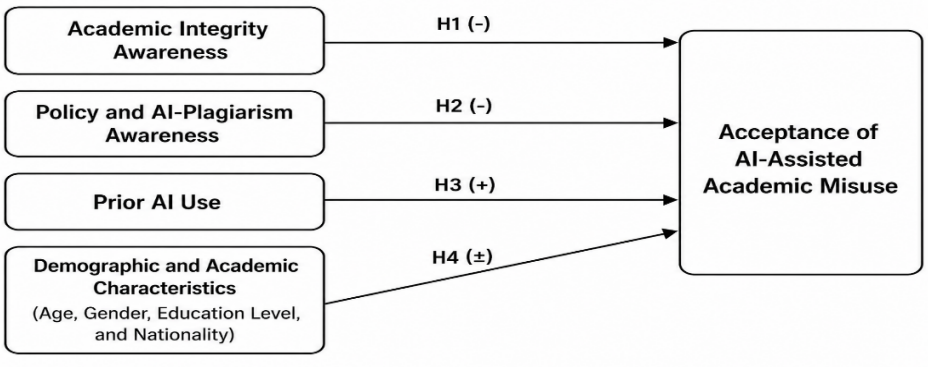
The boundary between legitimate support and misuse is often context dependent. AI may be permitted for explaining concepts, language support, or preliminary ideas but prohibited when it produces assessed work. Course requirements, assessment type, instructor instructions, disclosure expectations, and institutional policy can therefore shape judgements. In this study, acceptance of AI-assisted academic misuse refers to the extent to which students regard inappropriate AI-supported practices as acceptable or justifiable; acceptance is an attitude and does not demonstrate that misconduct occurred.

General academic-integrity awareness emphasises honesty, originality, acknowledgement, and responsibility, but conventional guidance may not fully address AI-generated writing, paraphrasing, editing, and idea generation. AI-specific policy and plagiarism awareness may be more proximal because clear rules can specify permitted, restricted, and prohibited uses, disclosure requirements, and the academic implications of unattributed AI-generated material (Chan, 2025; Khairullah et al., 2025; Alotaibi, 2024).

Prior AI use may either improve understanding of appropriate applications or normalise dependence on AI-generated content (Farhi et al., 2023; Li et al., 2025). Demographic characteristics may also capture differences in educational and institutional experience, although evidence is inconsistent (Koscielniak et al., 2026; Pawlak et al., 2025). Accordingly, the study tests whether general integrity awareness, policy and AI-plagiarism awareness, prior AI use, and demographic characteristics are associated with misuse acceptance in Kuwait.

3. Conceptual Model and Hypothesis Development

Figure 1 presents the conceptual model. Acceptance of AI-assisted academic misuse is the outcome, while general academic-integrity awareness, policy and AI-plagiarism awareness, prior AI use, and demographic/academic characteristics are the explanatory variables. The framework distinguishes broad ethical awareness from AI-specific policy knowledge and contextual characteristics. Because the study is cross-sectional, the hypothesised paths are interpreted as statistical associations rather than causal effects.



Note: (–) indicates a negative relationship; (+) indicates a positive relationship; (±) indicates that the relationship may be positive or negative depending on the group.

Figure 1: Proposed research model

3.1 Academic Integrity Awareness and Acceptance of AI-Assisted Academic Misuse

Students with stronger understanding of academic-integrity principles are generally expected to be less accepting of plagiarism, unauthorised AI-generated work, and related misconduct. Ethical awareness may strengthen judgements concerning originality, acknowledgement, fairness, and responsibility (Lund et al., 2025; Huang et al., 2025; Hanh & Duyen, 2025; Shishakly & Nachouki, 2026; Khairullah et al., 2025). Accordingly:

H1: Academic integrity awareness negatively predicts students' acceptance of AI-assisted academic misuse.

3.2 Policy and AI-Plagiarism Awareness and Acceptance of AI-Assisted Academic Misuse

Awareness of institutional AI policies and AI-related plagiarism should help students distinguish acceptable learning support from prohibited assistance. Clear rules can clarify disclosure, attribution, and assessment-specific boundaries and may therefore reduce tolerance of inappropriate AI use (Chan, 2025; Ali & Qureshi, 2025; Khairullah et al., 2025; AlBlooshi, 2026). Accordingly:

H2: Policy and AI-plagiarism awareness negatively predict students' acceptance of AI-assisted academic misuse.

3.3 AI Use and Acceptance of AI-Assisted Academic Misuse

Familiarity with AI chat assistants may shape students' acceptance of AI-assisted academic misuse. Repeated exposure could normalise AI use in academic tasks and increase tolerance of practices that conflict with integrity requirements; alternatively, experienced users may better understand appropriate boundaries. The expected direction in this study is positive, while recognising that AI use does not itself imply misconduct (Farhi et al., 2023; Li et al., 2025; Hanh & Duyen, 2025). Accordingly:

H3: Students' prior use of AI chat assistants positively predicts their acceptance of AI-assisted academic misuse.

3.4 Demographic Characteristics and Acceptance of AI-Assisted Academic Misuse

Age, gender, education level, and nationality may be associated with students' ethical perceptions through differences in educational experience, exposure to institutional policies, technology use, and assessment expectations. Existing evidence is mixed, so the hypothesis is non-directional across these groups (Al Omari et al., 2024; Koscielniak et al., 2026; Pawlak et al., 2025). Accordingly:

H4: Age, gender, education level, and nationality are significantly associated with students' acceptance of AI-assisted academic misuse.

4. Method

4.1 Research Design, Participants and Data Collection

This study used a quantitative cross-sectional online survey to examine associations with acceptance of AI-assisted academic misuse (Bryman, 2016; Creswell & Creswell, 2017). Participants were university students in Kuwait aged 18 years or older. Convenience and snowball sampling were used because the study sought access to students across institutions; consequently, the sample should not be treated as nationally representative. Data were collected anonymously through Google Forms from January to April 2026 via university mailing lists, student organisations, academic networks, communication channels, and social media. Of the eligible complete questionnaires, 304 were retained. Participation was voluntary, electronic informed consent was obtained, and no personally identifiable information was collected.

4.2 Measures and Scoring

The questionnaire comprised two sections. The first collected demographic and background information, including age, gender, employment status, education level, academic major, nationality, and prior experience with AI chat assistants; additional questions identified the AI tools used and the purposes for which students used them. The second section contained 15 dichotomously scored Yes/No items assessing three constructs.

General academic-integrity awareness comprised Y1, Y2, Y3, Y7, Y9, Y11, and Y15; policy and AI-plagiarism awareness comprised Y4, Y5, Y6, Y8, and Y10; and acceptance of AI-assisted academic misuse comprised Y12, Y13, and Y14. Responses were coded 1 for Yes and 0 for No. Composite scores were calculated as the mean of the constituent items and therefore ranged from 0 to 1, with higher scores indicating greater awareness or greater acceptance of AI-assisted academic misuse, depending on the construct. Prior AI use was also coded 0 = No and 1 = Yes. The questionnaire was reviewed by academics for content validity and pilot-tested for clarity, readability, relevance, structure, and preliminary reliability before administration. Minor wording revisions were made before final distribution. The Kuder-Richardson Formula 20 (KR-20) was used to assess internal consistency for the three multi-item constructs because all constituent items were dichotomously scored.

4.3 Data Analysis

Data were analysed using the Statistical Package for the Social Sciences (SPSS). Before analysis, the dataset was screened for missing values, duplicate responses, outliers, data-entry errors, and violations of relevant statistical assumptions. Frequencies, percentages, means, and standard deviations summarised participants' demographic and academic characteristics and the principal study variables. The 15 dichotomous measurement items were examined using exploratory principal-axis factoring based on a Pearson/phi correlation matrix with varimax rotation. Sampling adequacy and factorability were evaluated using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity. Factor retention and item assignment were guided by the empirical loading pattern, cross-loadings, conceptual relevance, and interpretability of the resulting factors. Because the items were binary and Pearson/phi correlations were used, the factor analysis was treated as exploratory and interpreted cautiously. Internal consistency of the three multi-item dichotomous constructs was assessed using KR-20. Acceptance of AI-assisted academic misuse was operationalised as the mean of Y12-Y14 and therefore formed a bounded continuous composite ranging from 0 to 1 rather than a single binary or ordinal outcome. Ordinary least squares (OLS) regression with HC3 heteroskedasticity-robust standard errors was used to estimate statistical associations with misuse acceptance. The model included general academic-integrity awareness, policy and AI-plagiarism awareness, prior AI use, and selected demographic characteristics. Categorical predictors were dummy coded before analysis: gender as female versus male, age as 18-30 versus over 30, nationality as non-Kuwaiti versus Kuwaiti, prior AI use as No versus Yes, and education as high school/Tawjeehi, two-year college degree, and bachelor's degree, with the two-year college degree as the reference category. Results were reported using unstandardised coefficients, HC3 robust standard errors, model statistics, and p-values. Statistical significance was set at $p < .05$, and coefficients were interpreted as statistical associations rather than causal effects.

5. Findings

The sample included 304 respondents; 68.1% were female, 69.4% were aged 18–30, and 61.8% were non-Kuwaiti. Prior AI use was widespread (92.4%). Among AI users, ChatGPT was the most frequently reported platform (94.7%), followed by Gemini (33.5%), DeepSeek (18.1%), and Copilot (15.3%). The main purposes were understanding difficult concepts (76.2%), summarising materials (44.1%), assignment support (41.6%), research assistance (37.0%), and writing improvement (35.2%). These patterns indicate that AI use in the sample was primarily academic-support oriented rather than evidence of misconduct.

Principal-axis factoring with varimax rotation was conducted on the 15 Yes/No measurement items. Sampling adequacy was satisfactory (KMO = .820), and Bartlett's test of sphericity was statistically significant, $\chi^2(105) = 1002.373$, $p < .001$. Although four initial eigenvalues exceeded 1, the fourth factor was weak and effectively single-item; a three-factor solution was therefore retained on the basis of statistical interpretability, conceptual coherence, and the intended construct structure. The first three initial eigenvalues accounted for 47.91% of the total variance, while the rotated principal-axis solution accounted for 36.11% based on rotated sums of squared loadings. The retained factors represented policy and AI-plagiarism awareness, general academic-integrity awareness, and acceptance of AI-assisted academic misuse. Table 1 reports all three loadings for every item to make cross-loadings transparent and to address the reviewer's request for fuller factor-analysis reporting.

Table 1: Rotated factor loadings for the 15 dichotomous Yes/No measurement items

Item	Statement	Policy / AI-Plagiarism Awareness	Academic Integrity Awareness	AI Misuse Acceptance
Y4	I know that I am not allowed to use AI Chat Assistants for graded assessments.	0.719		
Y5	I am fully aware of my university's policy regarding the usage of AI Chat Assistants.	0.468		
Y6	Using AI Chat Assistants to produce academic writing without proper attribution is considered plagiarism.	0.720		
Y8	Plagiarism from AI Chat Assistants can be detected by instructors using plagiarism detection software.	0.692		
Y10	For academic writing, it is better for students to use their own knowledge, skills, and research than relying on AI Assistants Tools	0.732		
Y1	Plagiarism is a form of academic dishonesty because it involves presenting someone else's work as one's own, without giving credit to the original author or source.		0.656	
Y2	Rewriting or paraphrasing material from any source without citing the original is plagiarism.		0.404	
Y3	Cutting and pasting material from various sources without referencing them is plagiarism.		0.697	
Y7	Relying on AI Chat Assistants discourages critical thinking, problem-solving, and creativity.		0.517	
Y9	AI Chat Assistants can infringe copyright by generating text that may violate the rights of original authors.		0.626	
Y11	AI Assistants Tools should be used to complement my own research and writing and not to complete my assigned homework		0.555	
Y15	Academically weak students are more likely to plagiarise with ChatGPT		0.573	
Y12	Using AI Assistants Tools for plagiarism can be acceptable under certain circumstances (e.g., extreme financial pressures, or low learning value of the assignment)			0.548
Y13	In the absence of university rules, using AI Assistants Tools for plagiarism is acceptable			0.823
Y14	It is no big deal if I submit a homework using AI Assistants tools generated text and with no referencing			0.758

Note: All Y1-Y15 items were dichotomously scored (Yes = 1; No = 0). Extraction method: principal-axis factoring; rotation method: varimax. All factor loadings are shown to make cross-loadings visible. Items with primary loadings of approximately .40 or above were generally retained. Y2 and Y12 had weaker primary loadings but were retained because they represented core conceptual content of their intended constructs. Because binary items were analysed using Pearson/phi correlations, the factor solution is exploratory and should be interpreted cautiously.

Table 2: Descriptive statistics and reliability of study variables

Variable	Items	N	Mean	SD	Min	Max	KR-20
Acceptance of AI-assisted academic misuse	Y12-Y14	304	.538	.361	.000	1.000	.587
General academic-integrity awareness	Y1-Y3, Y7, Y9, Y11, Y15	304	.763	.263	.000	1.000	.735
Policy and AI-plagiarism awareness	Y4-Y6, Y8, Y10	304	.764	.302	.000	1.000	.757
Prior AI chat assistant use	Single binary item	304	.924	.265	.000	1.000	-

Note: The three multi-item constructs are means of dichotomous items coded 0 = No and 1 = Yes; therefore, their composite scores range from 0 to 1. KR-20 is reported for general academic-integrity awareness, policy and AI-plagiarism awareness, and acceptance of AI-assisted academic misuse. Prior AI use is a single binary item and no internal-consistency coefficient is applicable. The misuse-acceptance reliability was modest (KR-20 = .587), which limits measurement precision and requires cautious interpretation.

At the bivariate level, misuse acceptance was positively correlated with general academic-integrity awareness ($r = .161$, $p < .01$) and policy and AI-plagiarism awareness ($r = .188$, $p < .01$), while its correlation with prior AI use was small and non-significant ($r = .082$). These correlations describe association only and do not establish direction or causation.

Table 3: Model summary for OLS regression of acceptance of AI-assisted academic misuse

Statistic	Value
N	304
R^2	.067
Adjusted R^2	.042
F	2.579
p	.010

Note: OLS regression with HC3 robust standard errors; N = 304. The dependent variable is the continuous 0–1 misuse-acceptance composite. Prior AI use tests H3.

Table 4: OLS regression of acceptance of AI-assisted academic misuse

Predictor	B (HC3 robust SE)	β	p
General academic-integrity awareness	.088 (.090)	.064	.327
Policy and AI-plagiarism awareness	.193 (.078)*	.162	.013
Prior AI chat assistant use	.098 (.087)	.072	.262
Gender: male	-.051 (.044)	-.066	.247
Age: over 30	-.051 (.055)	-.065	.358
Education: bachelor's degree	-.041 (.052)	-.054	.432
Education: high school/Tawjeehi	-.011 (.054)	-.015	.832
Nationality: Kuwaiti	.103 (.050)*	.139	.039

*Note: Reference categories are female, age 18–30, college degree (2 years), non-Kuwaiti, and no prior AI use. * p < .05. ** p < .01. *** p < .001.*

Table 5: Multicollinearity diagnostics for the OLS model

Predictor	VIF
General academic-integrity awareness	1.42
Policy and AI-plagiarism awareness	1.40
Prior AI chat assistant use	1.04
Gender: male	1.03
Age: over 30	1.44

Predictor	VIF
Education: bachelor's degree	1.41
Education: high school/Tawjeehi	1.42
Nationality: Kuwaiti	1.46

Note: All VIF values were below 1.50, indicating no material multicollinearity concern among the predictors.

The OLS model was statistically significant, F = 2.579, p = .010, and explained 6.7% of the variance in misuse acceptance (adjusted R^2 = .042). Policy and AI-plagiarism awareness showed a statistically significant positive association (B = .193, p = .013), which was opposite to the negative direction proposed in H2. Kuwaiti nationality was also positively associated with misuse acceptance (B = .103, p = .039). General academic-integrity awareness, prior AI use, gender, age, and education level were not statistically significant. The small explained variance and cross-sectional design require restrained interpretation.

Table 6: Summary of hypothesis testing

Hypothesis	Evidence	Decision
H1	General integrity awareness: B = .088, p = .327	Not supported
H2	Policy/AI-plagiarism awareness: B = .193, p = .013; observed direction positive, hypothesised direction negative	Statistically significant; directional H2 not supported
H3	Prior AI use: B = .098, p = .262	Not supported
H4	Male p = .247; age >30 p = .358; bachelor's p = .432; high school/Tawjeehi p = .832; Kuwaiti: B = .103, p = .039	Partially supported (nationality only)

Note: H2 was statistically significant but not supported in its hypothesised negative direction. "Association" is used throughout to avoid causal interpretation of cross-sectional estimates.

Overall, misuse acceptance was associated most clearly with policy and AI-plagiarism awareness and nationality in the multivariable model. The positive policy-awareness coefficient should not be interpreted as evidence that awareness causes tolerance of misuse; reverse direction, unmeasured contextual factors, or greater exposure to ambiguous AI-use situations remain possible explanations.

6. Discussion

This study examined factors associated with university students' acceptance of AI-assisted academic misuse in Kuwait. AI use was widespread and was concentrated on academic-support activities, particularly concept explanation, summarisation, assignment support, research, and writing. This reinforces the need to distinguish ordinary AI use from academically inappropriate use rather than treating AI users as a uniform integrity-risk group (Cotton et al., 2024; Farhi et al., 2023).

Participants reported relatively high general academic-integrity awareness and policy/AI-plagiarism awareness, yet misuse acceptance was moderate. The pattern indicates that knowledge of integrity principles does not necessarily resolve context-specific questions about AI authorship, disclosure, paraphrasing, and permitted assistance (Lund et al., 2025; Khairullah et al., 2025).

General academic-integrity awareness was not significantly associated with misuse acceptance in the multivariable model (H1 not supported). Students may understand traditional plagiarism while remaining uncertain about AI-generated writing, editing, brainstorming, and language support. Universities should therefore supplement general integrity education with AI-specific examples addressing authorship, attribution, disclosure, and assessment expectations (Huang et al., 2025; Hanh & Duyen, 2025).

Policy and AI-plagiarism awareness was significantly but positively associated with misuse acceptance. Because H2 specified a negative relationship, the directional hypothesis was not supported even though the coefficient was statistically significant. The unexpected positive direction may indicate that students who are more familiar with AI rules also encounter more nuanced or ambiguous use cases; however, the cross-sectional data cannot

determine why this pattern occurred (Lund et al., 2025). Prior AI use was not significant (H3 not supported), suggesting that experience with AI alone does not differentiate students' acceptance of misuse.

Among the demographic variables, only nationality was statistically significant, so H4 received partial support. Kuwaiti respondents reported higher misuse-acceptance scores than non-Kuwaiti respondents after adjustment for the other predictors. This result should not be interpreted as an inherent cultural or national difference. Nationality may proxy differences in educational background, institutional exposure, programme composition, assessment expectations, or other unmeasured contextual factors; replication across institutions and countries is required (Al Omari et al., 2024; Deng et al., 2025; Syed et al., 2025).

Several limitations qualify the findings. The cross-sectional design cannot establish temporal or causal relationships; self-report may be affected by social-desirability bias; convenience and snowball sampling restrict generalisability; and the misuse-acceptance scale showed modest internal consistency (KR-20 = .587), increasing measurement error. The OLS model also explained only a small proportion of variance. Future research should use larger multi-institutional and longitudinal samples, refine the misuse-acceptance measure, and examine additional constructs such as AI literacy, self-efficacy, trust, and institutional support (Bewersdorff et al., 2025).

7. Conclusion Limitations and Implications

This study provides evidence from Kuwait on factors statistically associated with students' acceptance of AI-assisted academic misuse. Policy and AI-plagiarism awareness and nationality were significant in the multivariable model, whereas general academic-integrity awareness and prior AI use were not. Importantly, the policy-awareness association was positive rather than the hypothesised negative direction, so it should be treated as an unexpected association requiring further investigation rather than as support for H2.

The conclusions are bounded by the cross-sectional design, self-reported data, non-probability sampling, modest reliability of the misuse-acceptance scale, and low explained variance. These limitations prevent causal claims and restrict generalisation beyond comparable university settings in Kuwait. The nationality association likewise requires replication before substantive cultural interpretations are made.

The findings nevertheless have practical implications. General plagiarism knowledge alone may be insufficient for emerging practices such as AI-assisted writing, summarising, paraphrasing, and idea generation. Academic-integrity frameworks should clearly distinguish legitimate AI-supported learning from academically inappropriate use and specify purpose, extent, disclosure, attribution, and assessment-specific boundaries (Lund et al., 2025; Khairullah et al., 2025).

Universities should therefore provide clear and consistently applied AI-use policies, practical examples of permitted and prohibited practices, disclosure requirements, and targeted AI-literacy education. Such guidance can support responsible AI-assisted learning while protecting originality, fairness, and assessment validity (Cotton et al., 2024; Sallam et al., 2024).

Ethics Declaration

The study received ethical approval from the Arab Open University–Kuwait Institutional Review Board and complied with all relevant ethical guidelines.

AI Declaration

Generative AI was used for language editing and improving clarity. All suggestions were reviewed and verified by the authors, who take full responsibility for the final manuscript.

References

Al Omari, O., Alshammari, M., Al Jabri, W., Al Yahyaei, A., Aljohani, K. A., Sanad, H. M., Al-Jubouri, M. B., Bashayreh, I., Fawaz, M., ALBashtawy, M., Alkhawaldeh, A., Qaddumi, J., Shalaby, S. A., Abdallah, H. M., AbuSharour, L., Al Qadire, M., & Aljezawi, M. (2024). Demographic factors, knowledge, attitude and perception and their association with nursing students' intention to use artificial intelligence (AI): A multicentre survey across 10 Arab countries. BMC Medical Education, 24, Article 1456. https://doi.org/10.1186/s12909-024-06452-5

AlBlooshi, S. (2026). Artificial intelligence in higher education, opportunities, and challenges: A review. Frontiers in Education, 10, Article 1683968. https://doi.org/10.3389/feduc.2025.1683968

Ali, K., & Qureshi, S. (2025). The impact of AI tools (like ChatGPT) on student learning outcomes and academic integrity: A mixed-methods study. VFAST Transactions on Education and Social Sciences, 31(1), 36–46. https://doi.org/10.21015/vtess.v13i1.2180

Alotaibi, N. S. (2024). The impact of AI and LMS integration on the future of higher education: Opportunities, challenges, and strategies for transformation. Sustainability, 16(23), 10357. https://doi.org/10.3390/su162310357

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K., et al. (2023). The emergent role of artificial intelligence, natural language processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 19(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Bewersdorff, A., Hornberger, M., Nerdel, C., & Schiff, D. S. (2025). AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. Computers and Education: Artificial Intelligence, 8, 100340. https://doi.org/10.1016/j.caeai.2024.100340

Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.

Chan, C. K. Y. (2025). Students' perceptions of "AI-giarism": Investigating changes in understandings of academic misconduct. Education and Information Technologies, 30, 8087–8108. https://doi.org/10.1007/s10639-024-13151-7

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Deng, N., Liu, E. J., & Zhai, X. (2025). Understanding university students' use of generative AI: The roles of demographics and personality traits. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Artificial Intelligence in Education (Vol. 15877, Lecture Notes in Computer Science). Springer. https://doi.org/10.1007/978-3-031-98414-3_20

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., & Slamene, R. (2023). Analyzing students' views, concerns, and perceived ethics about ChatGPT usage. Computers and Education: Artificial Intelligence, 5, 100180. https://doi.org/10.1016/j.caeai.2023.100180

Hanh, N. V., & Duyen, N. T. (2025). AI-assisted academic cheating: A conceptual model based on postgraduate student voices. Frontiers in Computer Science, 7, Article 1682190. https://doi.org/10.3389/fcomp.2025.1682190

Huang, C., Alhur, A. A., Azam, M., & Naeem, S. B. (2025). The challenge of academic integrity in the age of generative artificial intelligence. Libri, 75(4), 339–354. https://doi.org/10.1515/libri-2025-0024

Khairullah, S. A., Harris, S., Hadi, H. J., Sandhu, R. A., Ahmad, N., & Alshara, M. A. (2025). Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in higher education institutions. Frontiers in Education, 10, Article 1548104. https://doi.org/10.3389/feduc.2025.1548104

Koscielniak, M., Chudzicka-Czupała, A., & Gasiorowska, A. (2026). Intentions for AI academic misconduct across 15 countries: The theory of planned behavior and past misconduct perspective. Scientific Reports. Advance online publication. https://doi.org/10.1038/s41598-026-61543-1

Li, Y., Castulo, N. J., & Xu, X. (2025). Embracing or rejecting AI? A mixed-method study on undergraduate students' perceptions of artificial intelligence at a private university in China. Frontiers in Education, 10, Article 1505856. https://doi.org/10.3389/feduc.2025.1505856

Lund, B., Mannuru, N. R., Teel, Z. A., Lee, T. H., Ortega, N. J., Simmons, S., & Ward, E. (2025). Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in higher education. AI in Education, 1(1), Article 2. https://doi.org/10.3390/aieduc1010002

Pawlak, J., Domingues, J. P. T., & Oliveira, O. J. (2025). Demographic and psychological determinants of students' attitudes toward the use of AI tools in education. The TQM Journal, 38(3), 611–632. https://doi.org/10.1108/TQM-12-2024-0520

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. Information, 15(10), 596. https://doi.org/10.3390/info15100596

Sallam, M., Elsayed, W., Al-Shorbagy, M., Barakat, M., El Khatib, S., Ghach, W., Alwan, N., Hallit, S., & Malaeb, D. (2024). ChatGPT usage and attitudes are driven by perceptions of usefulness, ease of use, risks, and psycho-social impact: A study among university students in the UAE. Frontiers in Education, 9, Article 1414758. https://doi.org/10.3389/feduc.2024.1414758

Shishakly, R., & Nachouki, M. (2026). Academic integrity and students' ethical use of ChatGPT in higher education. Journal of Information Technology Education: Research, 25, Article 8. https://doi.org/10.28945/5730

Syed, A. Z., Memon, Z. H., Khan, K., Hameed, I., & Nadeem, M. (2025). Examining the behavioral determinants of AI adoption in higher education: A focus on perceptional factors and demographic differences. On the Horizon, 33(2), 245–264. https://doi.org/10.1108/OTH-02-2025-0019

A Scalable, Work-Integrated E-Learning Model for Rapid Upskilling in Biomanufacturing

Ana Isabel Aranda[1], Sara Vernocchi[1], Timo Tekhaus[2] and Matthias Afting[2]
[1]LUNEX University of Applied Sciences, Differdange, Luxembourg
[2]Cognos International, Cologne, Germany

aaranda@lunex.lu
svernocchi@lunex.lu
ttekhaus@cognos-international.de
mafting@cognos-international.de

Abstract: Africa faces significant challenges in developing a skilled biomanufacturing workforce, including rapid upskilling needs, limited access to high-quality training, and diverse learner backgrounds. These constraints require scalable, adaptable e-learning solutions that align with industry and regulatory demands, particularly for working professionals pursuing continuing education. This paper presents an inclusive, scalable online pedagogical model grounded in constructivist learning theory. Using a flipped-classroom approach, it combines asynchronous, self-paced learning with synchronous, application-oriented sessions within a competency-based framework. Learners engage with multimedia-rich content prior to live online sessions, enabling synchronous time to focus on higher-order skills such as problem-solving, discussion, and contextual application. Gamification and Problem-Based Learning (PBL) support engagement, critical thinking, and collaborative knowledge construction, while feedback from Subject Matter Experts (SMEs) ensures real-time validation of learning. The curriculum is organised into units with clear learning outcomes, aligned with Bloom's taxonomy and supported by question banks to ensure constructive alignment. Formative and summative assessments enable both mastery learning and performance evaluation. Delivered via a digital ecosystem centred on Canvas LMS and interoperable content authoring tools, the model supports scalability, analytics, accessibility, a Bring Your Own Device (BYOD) approach, and multilingual adaptability, facilitating inclusive participation across diverse geographic and professional contexts. Implemented across industrial and higher-education (HE) settings, the model was associated with positive feedback among learners. Its modular and reusable design offers a flexible, transferable approach to e-learning that supports workforce development across disciplines and environments, including resource-limited settings.

Keywords: Flipped-Classroom Model, Problem-Based Learning, Constructivist Pedagogy, Gamification, Biomanufacturing Workforce Development, Work-Integrated Learning.

1. Introduction

In the pharmaceutical and biomanufacturing sectors, professional training is undergoing a major transformation driven by increasing global demand for a skilled workforce, rapid technological advancement, and evolving regulatory expectations. Across the sector, there is growing recognition of the need for workforce development models that support timely, scalable, and competency-based education and training aligned with healthcare and industry priorities (International Pharmaceutical Federation, 2017).

The European biomanufacturing and life sciences sectors are experiencing rapid expansion alongside increasing concerns regarding shortages of appropriately skilled personnel. Recent European Commission and industry reports identify workforce capacity and specialised technical expertise as major constraints affecting the continued growth of biotechnology and biomanufacturing industries in Europe (European Commission, 2024; Work in Health Foundation, 2024). These challenges are further amplified in low- and middle-income regions, particularly in Africa, where the development of local pharmaceutical manufacturing capacity has become a strategic public health priority. The African Union has established a target to locally manufacture 60% of vaccines used on the continent by 2040, requiring substantial expansion of the biomanufacturing workforce (Mynhardt et al., 2024). Current estimates suggest that the vaccine manufacturing workforce will need to increase from approximately 2,000–3,000 professionals to between 9,000 and 14,000 by 2040 (Institut Pasteur de Dakar (IPD), Africa CDC and SAMRC, 2023).

Despite this growing demand, existing training ecosystems face structural limitations. Many programmes remain small-scale and prioritise theoretical knowledge over applied competencies, reducing learners' preparedness for operational roles in highly regulated manufacturing environments. Workplace-based learning and continuing professional development opportunities remain insufficient, while face-to-face delivery limits accessibility and scalability for geographically distributed and professionally active learners. As a result, organisations often rely on fragmented internal training systems or external providers that may not support sustainable long-term workforce development (IPD, Africa CDC and SAMRC, 2023).

This paper presents a scalable, work-integrated e-learning model developed to address these challenges within the context of biomanufacturing workforce development. The model originated from a workforce development initiative supporting the expansion of biomanufacturing operations in Africa. It was designed to provide a flexible training solution for personnel with heterogeneous educational and professional backgrounds while maintaining alignment with scientific, industrial, and regulatory requirements. To address these limitations, the instructional model was structured around three interconnected objectives: scalability across geographically distributed learners; compatibility with full-time professional responsibilities; and competency-oriented learning aligned with operational practice.

The model was not conceived as a replacement for practical laboratory or workplace training, which remains essential in biomanufacturing contexts, but rather as a complementary digital learning layer intended to accelerate theoretical preparation, process understanding, and operational readiness prior to workplace integration. The model combines asynchronous self-paced learning, synchronous collaborative activities, multimedia resources, and competency-aligned assessment within an integrated digital learning ecosystem. The instructional design was informed primarily by constructivist learning theory and operationalised through a flipped-classroom structure supported by Problem-Based Learning (PBL) and gamification strategies. These approaches were selected to support active participation, learner engagement, collaborative knowledge construction, and the application of learning to authentic professional contexts.

This paper, therefore, contributes to current discussions on scalable online professional education by presenting a transferable instructional design model for workforce development in resource-constrained and highly regulated environments.

2. Pedagogical Framework

2.1 Constructivist Learning Theory

Constructivism is a major framework for adult and lifelong learning because it views learning as an active process in which individuals construct knowledge through experience, reflection, social interaction, and authentic situations. Learners connect new information with prior knowledge and practical experience, while instructors primarily act as facilitators who support discussion, knowledge sharing, and the application of learning in workplace contexts. However, the successful implementation of constructivist learning environments in industrial settings may require appropriate pedagogical support and instructional design to develop authentic, learner-centred, and collaborative activities, particularly in technology-supported or online environments (Bada, 2015; Chuang, 2021; Gentile et al., 2022; Mohammed and Kinyo, 2020).

2.2 Flipped Classroom Approach

The flipped classroom is a learner-centred teaching approach in which instructional content is studied before class, while synchronous sessions focus on discussion, collaboration, and problem-solving. Strongly aligned with constructivist principles, this model promotes self-directed and active learning (Seifert, 2023; Akti Aslan, 2022).

Research generally indicates that flipped classroom approaches can improve student participation, motivation, interaction, and problem-solving skills, and engagement with educational content (Cuetos, 2023; Akti Aslan, 2022). However, evidence regarding learning achievement is less consistent. For example, Akti Aslan (2022) reported improvement in self-regulation and interaction, but no significant effect on academic achievement, suggesting that simply relocating content outside synchronous sessions is insufficient to guarantee improved learning achievement. Conversely, Thai et al. (2020) found significantly higher learning performance when additional feedback was incorporated into the online preparatory phase, indicating that the effectiveness of flipped learning may depend less on the “flipped” structure itself than on the quality of instructional support embedded within it. In professional training, flipped learning enables employees to complete self-paced preparation before participating in practical sessions focused on case studies and real-world applications. Its effectiveness, however, depends on learner motivation, self-directed learning skills, and clear instructional support (Gentile et al., 2022).

2.3 Problem-Based Learning Integration

PBL is a learner-centred approach in which learners acquire knowledge and skills by analysing real-world problems. Similar to constructivist learning theory and flipped classroom approaches, PBL emphasises

collaboration, inquiry, and application of knowledge in authentic contexts, while instructors act primarily as facilitators (Zeng and Ruannakarn, 2023; Stegeager et al., 2013; Lopez Brown, 2017).

Research has shown that PBL can improve critical thinking, problem-solving abilities, learner engagement, and the transfer of learning to practical situations compared with traditional lecture-based teaching (O'Brien and Hamburg, 2013; Zeng and Ruannakarn, 2023; Stegeager et al., 2013). This makes PBL relevant to professional training, although successful implementation requires authentic case design, appropriate facilitation, and sufficient learner support (O'Brien and Hamburg, 2013; Stegeager et al., 2013).

2.4 Gamification

Although constructivist, flipped classroom, and problem-based approaches support active learning, maintaining learner engagement remains a major challenge in online and workplace education. Gamification, therefore, provides a complementary strategy to increase motivation, participation, and interaction. Gamification refers to the integration of game elements into non-game learning environments. In e-learning, gamification has been increasingly adopted because online learning systems often suffer from low engagement and high dropout rates (Khaldi et al., 2023). Common elements include points, badges, and leaderboards, while more advanced approaches incorporate storytelling, role-play, challenges, and collaborative activities (Khaldi et al., 2023).

Research shows that gamification can positively influence learning outcomes, motivation, collaboration, participation, and learner enjoyment across different educational settings (Diaz and Estoque-Loñez, 2024; Su et al., 2025). However, these findings should be interpreted cautiously, as different game elements may influence different aspects of learning and engagement. Among the various gamified applications evaluated in educational research, Kahoot demonstrated one of the largest positive effects on learner engagement and academic performance (Diaz and Estoque-Loñez, 2024). It is worth noting that Khaldi et al. (2023) identified a gap between gamification theory and practice, with many applied studies lacking a clear theoretical rationale for selecting and combining particular game elements, while many design approaches remain insufficiently validated. Therefore, the effectiveness of gamification should be considered in relation to the educational context, learner characteristics and intended learning outcomes.

2.5 Cognitive Learning Strategies

In addition to active and learner-centred pedagogies, effective professional training requires strategies that support long-term retention of essential knowledge. Cognitive learning strategies support durable learning by encouraging learners to actively engage with content rather than passively review information. Retrieval practice, in particular, has been shown to be a universally effective learning strategy across different learner profiles and cognitive abilities (Bertilsson et al., 2021). Practical tools for cognitive learning strategies, such as flashcards or formative quizzes, can effectively promote active recall, spaced repetition, and self-assessment. A recent study found that students were highly adherent to flashcard initiatives and that increased flashcard use was associated with improved grades (Santos-Ferreira et al., 2024).

Collectively, the literature supports the integration of constructivist, flipped, problem-based, gamified and cognitive learning strategies; however, the evidence does not suggest that any single approach is universally effective. Rather, learning outcomes appear to depend on how these approaches are designed, facilitated, and aligned with instructional objectives and the learning context. This is particularly relevant when translating evidence predominantly derived from academic settings to professional and industrial training environments, where learners must balance training with workplace responsibilities and apply acquired knowledge to operational practice.

3. E-Learning Model Design

3.1 Content Development

The e-learning content was developed by the instructors through a structured co-design process involving Subject Matter Experts (SMEs) from diverse disciplinary and professional backgrounds, including industry specialists and external experts. SMEs contributed iteratively to the development process, providing feedback on scientific accuracy and industrial relevance.

The curriculum comprised five modules covering foundational scientific concepts (M1 and M2), microbiological principles (M3), biomanufacturing processes (M4), and quality control (M5). The present study focuses on the

first three preparatory modules (M1–M3), which were designed to establish a shared foundational knowledge base prior to process-specific training. These modules represented the most scalable and transferable components of the programme and were therefore selected for analysis. The overall programme received external accreditation and comprised 12 ECTS.

The model operationalised the pedagogical framework through a modular structure supporting progressive knowledge acquisition, learner autonomy, and applied learning. The structure was designed to accommodate diverse learner profiles while maintaining alignment with scientific, industrial, and regulatory requirements.

3.2 Technological Environment

The instructional model was implemented through an integrated digital learning ecosystem centred on Canvas LMS and supported by complementary tools including Articulate Rise, BioRender, Microsoft Teams, Kahoot, and Quizlet. The technological environment was designed to support multimodal content delivery, collaborative interaction, formative assessment, and scalable online participation across geographically distributed learners.

Canvas was selected as the LMS because it supports integration of multimedia resources, simulations, SCORM packages, H5P activities, and third-party learning tools within a single environment, reducing cognitive disruption and maintaining continuity throughout the learning process. In addition, Canvas offered intuitive navigation, integrated assessment and feedback tools, discussion forums, progress tracking, and engagement analytics.

These features supported scalable delivery while facilitating a Bring Your Own Device (BYOD) approach compatible with diverse technological contexts. Compatibility with multimedia and interactive resources also supported the implementation of competency-based and learner-centred training approaches in biomanufacturing education.

3.3 Course Architecture

The course architecture was designed to operationalise constructivist and flipped-classroom principles by integrating asynchronous and synchronous active-learning activities. Each module began with an online kick-off session that introduced the module structure, assessment strategy, and expectations, while supporting early learner interaction.

All learning materials were made available from the beginning of the module. Although learners were free to navigate the content at their own pace and according to their preferences, a sequential progression was recommended to support gradual knowledge development and competency acquisition. Interactivity was embedded throughout the modules to support engagement, reflection, and application of knowledge. Within each module, the curriculum was organised into units that progressed from foundational concepts to more advanced, applied topics (Figure 1, left).

Most units began with a discussion forum designed as a PBL activity (Figure 1, top right). PBL activities were intentionally non-graded to encourage open participation and exploratory thinking. By introducing authentic professional scenarios before formal instruction, these activities aimed to stimulate early reflection, contextual reasoning, and asynchronous peer interaction.

Learning materials were delivered through multiple complementary formats, including instructor-developed videos, curated external video resources, written explanatory materials, and interactive digital content. The use of diverse resource types aimed to accommodate varying learning preferences, support self-paced engagement, and provide multiple entry points for understanding complex scientific and technical concepts. Learning materials incorporated multilayered visual and interactive resources, developed using BioRender and Articulate Rise, to support engagement and accommodate learners with diverse backgrounds, allowing them to explore content at varying depths while reaching common competency goals. Introductory visual summaries provided accessible entry points to the content, while embedded exploratory elements, including hotspots (Figure 1, bottom right), expandable sections, and external references, allowed learners to progressively engage with increasingly advanced information as their learning needs evolved.

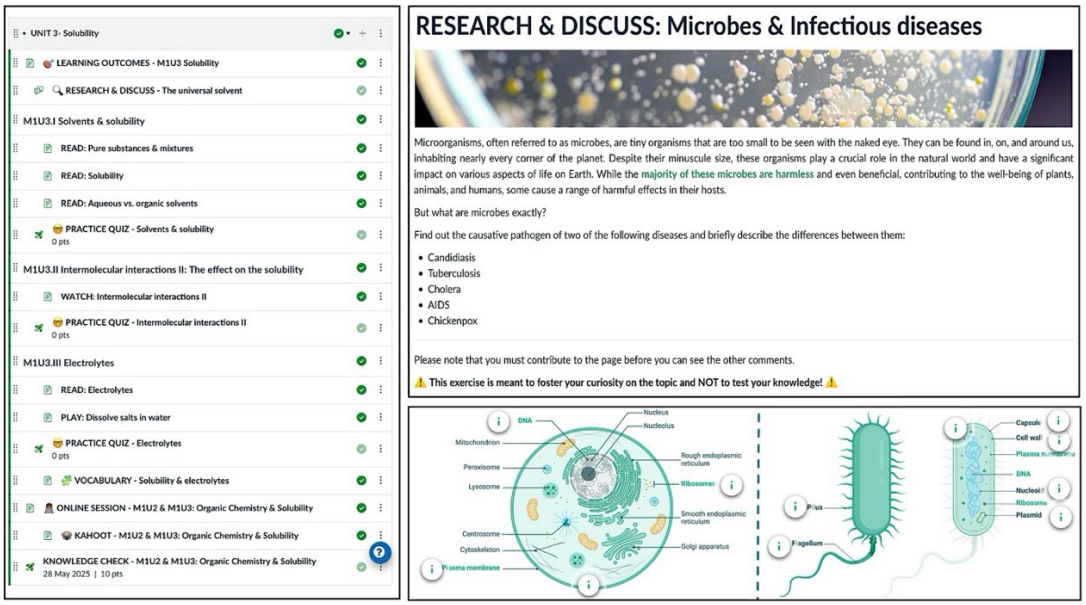


Figure 1: Left: Example of the structure of a unit within a module. Top right: Example of a PBL activity in a discussion forum on Canvas. Bottom right: Example of an interactive picture used in a module.

The learning outcomes were explicitly communicated at the beginning of each unit to support transparency and self-regulated learning. Formative assessment activities were integrated throughout the modules to support continuous self-evaluation and knowledge consolidation prior to summative assessment.

Each unit included flashcard sets that could be used in different game-based study modes (Figure 2). Additional learner-responsive knowledge-check activities were incorporated to reinforce terminology and conceptual understanding (Figure 3).

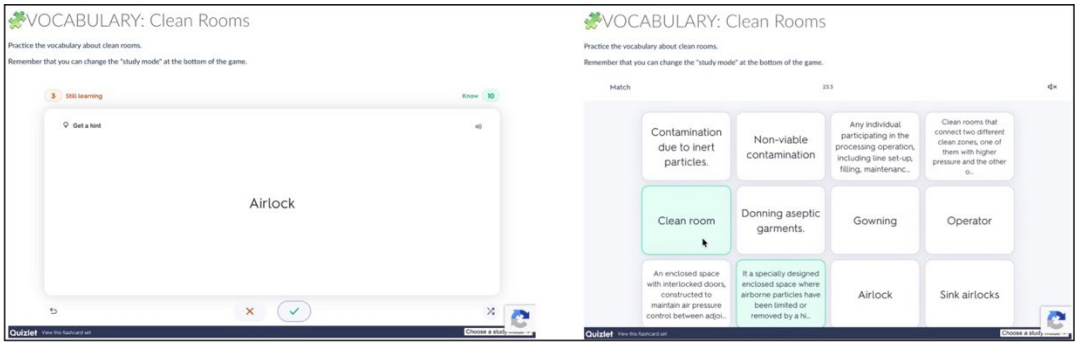


Figure 2: Examples of study modes for vocabulary activities used in the modules.

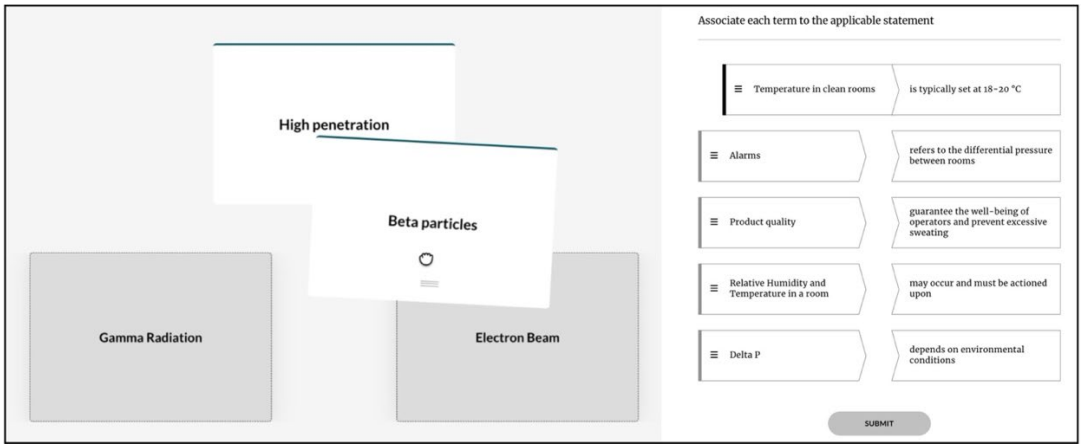


Figure 3: Examples of knowledge-check interactive activities used as formative assessment in the modules.

Simulations and scenario-based activities were incorporated where appropriate to support procedural understanding and application-oriented learning in contexts that would otherwise require laboratory or manufacturing facilities (Figure 4). Such activities were particularly valuable in biomanufacturing contexts, where visualisation, procedural understanding, and process awareness were essential for operational readiness.



Figure 4: Left: Simulation of sterile gowning procedures (NCBioNetwork, 2023). Right: Simulation of the solubility of salts (PhET Interactive Simulations, 2026).

In addition to asynchronous learning activities, regular synchronous online sessions were scheduled throughout each module. These sessions generally covered one or two units and lasted between 1.5 and 2 hours. Learners were encouraged to complete all asynchronous learning materials before attending the corresponding online sessions. These sessions focused on clarifying complex concepts, checking achievement of learning outcomes, elaborating on collaborative reflection, and discussing applied case studies. Kahoot quizzes were used as gamified discussion activities rather than formal assessment tools, promoting participation and peer interaction. Instructors used quiz questions as prompts for discussion and collaborative reflection and continuously encouraged learners to share questions and professional experiences to leverage the diversity of the learner groups. SMEs also participated in selected sessions to support discussion of technical topics and operational case studies.

Collectively, asynchronous self-paced learning, synchronous collaborative activities, multimedia resources, simulations, and gamified interactions created a multimodal active-learning environment in which learners continuously engaged with content, peers, and instructors, and applied their learning to professional scenarios.

From a delivery perspective, the model supported flexible implementation formats adapted to organisational constraints and learner profiles. The part-time format allocated approximately two weeks per module, with an average daily workload of four hours, enabling learners to combine training with their professional responsibilities. An intensive full-time format, comprising one week per module and an estimated daily workload of eight hours, was also proposed to support rapid onboarding and dedicated upskilling. Together, these formats enabled the model to be implemented across different workforce-development scenarios.

3.4 Quiz-Based Assessment Strategy

Formal evaluation was implemented through a structured quiz-based assessment system delivered within Canvas LMS. The assessment architecture was designed by the instructors to support competency-oriented learning, constructive alignment, retrieval practice, and scalable online delivery.

The training was organised into units defined by explicit learning outcomes developed according to the SMART framework and aligned with Bloom's taxonomy. To support alignment among learning objectives and assessment tasks, a dedicated question bank was developed for each learning outcome. Each question bank contained a set of questions targeting the same competency level and cognitive domain associated with the corresponding learning outcome. These question banks enabled consistency across assessments while allowing controlled variability through quiz randomisation mechanisms within Canvas. Assessment activities included multiple-choice and multiple-answer questions, matching exercises, fill-in-the-gap, categorisation tasks, ordering activities, and applied problem-solving exercises. The assessment model combined formative and summative components to support continuous learning and competency evaluation.

Formative assessment quizzes, termed "Practice Quizzes", were distributed throughout each unit to support self-regulated learning, retrieval practice, and iterative knowledge reinforcement. These quizzes contained 4–8

randomly selected questions drawn from the corresponding learning outcome-aligned question banks. Learners were allowed unlimited attempts without time limits, and immediate feedback was provided upon submission.

Summative evaluation was implemented through unit-level "Knowledge Checks" and a cumulative "Final Exam", both derived from the same learning outcome-aligned question bank structure as the Practice Quizzes. The Knowledge Checks and Final Exam were conducted within set time limits and allowed only a limited number of attempts. The pass mark for all summative assessments was set to 50%. Knowledge Checks typically contributed 8–10% each to the final module grade, while the Final Exam accounted for 60% of the total evaluation.

This approach enabled standardised evaluation processes, efficient deployment across multiple course iterations, and competency-aligned assessment with minimal reliance on synchronous examiner involvement. In addition, automated grading and integrated learning analytics enabled continuous monitoring of progress and facilitated the early identification of learners who required additional support.

Overall, the assessment strategy was designed to progressively transition learners from low-stakes knowledge reinforcement towards cumulative competency validation while supporting scalability and consistency within and across cohorts.

4. Methods

The educational programme was developed and delivered as part of routine teaching and professional education activities and was not originally designed as a research study. No learners were prospectively recruited for research purposes. The present study therefore adopted a retrospective case-study approach to describe the instructional design, its application across different educational settings, and learners' perceptions of their experiences.

Learner perceptions were evaluated using anonymous EvaSys (evasys GmbH, Lüneburg, Germany) questionnaires administered after each module. Closed-response items assessed perceived achievement of learning outcomes, quiz–learning outcome alignment, module organisation and navigation, learner interest, adequacy of time allocation, and the contribution of synchronous sessions to consolidating asynchronous learning. Items were rated on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Open-ended questions addressed module organisation, learning-resource preferences, and the teaching approach. Industrial and higher-education (HE) data were analysed separately. As questionnaires were administered after each module, responses represent module-level responses rather than unique learners. Common closed-response items were pooled within each setting and reported as means, standard deviations (SD), and positive responses (ratings 4–5). Pooled estimates were weighted by module-level response counts.

Open-ended responses were analysed using descriptive qualitative content analysis. A substantive response was defined as an interpretable response relevant to the question asked. Non-answers (e.g., "N/A"), uninterpretable responses, and off-topic responses were excluded, whereas explicit statements of no improvement or disliked aspects were retained as substantive responses. Responses were analysed separately for each open-ended question, including paired prompts exploring positively and less positively perceived aspects of module organisation, learning resources, and the teaching approach. Responses expressing similar ideas within each question were grouped into descriptive categories, and multiple categories could be assigned to a single response. The analysis was AI-assisted using ChatGPT (OpenAI, GPT-5.6 Sol) to organise responses, propose descriptive codes and categories, and identify ambiguous or overlapping classifications. Outputs were reviewed against the original responses, with ambiguous classifications manually reviewed and final coding decisions made under author oversight. Frequencies and percentages were calculated using substantive responses to each question as the denominator; percentages were not mutually exclusive because multiple coding was permitted.

5. Results

5.1 Learner Feedback Dataset

The model was implemented across four training deliveries in two educational settings: three industrial deliveries accounting for 120 module completions and one higher-education (HE) delivery accounting for 142 module completions across three preparatory modules (M1–M3). This resulted in 12 module implementations and corresponding evaluations. Anonymous learner feedback was collected at the end of each module through EvaSys by our institution as part of routine quality-assurance procedures. Across the industrial deliveries, 62 questionnaire responses were available from 120 module completions, representing a module-level response

rate of 51.7%. In the HE delivery, 85 responses were available from 142 module completions, representing a module-level response rate of 59.9%. Because feedback was collected after each module, these values represent module-level responses rather than unique respondents. The open-ended items generated 858 item-level responses across all questionnaire prompts—rather than responses from 858 individual learners—comprising 342 from the industrial implementations and 516 from the HE implementation. After blank, uninterpretable and off-topic entries were excluded, 751 responses were retained for qualitative content analysis. Percentages were calculated separately for each question using substantive responses to that question as the denominator.

5.2 Quantitative Learner Perceptions

Overall, learners evaluated the organisation, learning experience, assessment alignment, and synchronous components of the modules positively in both implementation settings (Table 1). Perceived achievement of the stated learning outcomes was rated positively by 87.1% of industrial responses and 90.6% of HE responses, with mean ratings of 4.35 (SD = 0.70) and 4.53 (SD = 0.85), respectively. Alignment between quiz questions and learning outcomes was also evaluated positively in both settings, with mean ratings of 4.55 (SD = 0.67) in industry and 4.75 (SD = 0.65) in HE. Synchronous online sessions were similarly perceived as supporting consolidation of prior asynchronous learning, with 87.1% and 88.3% positive responses in industry and HE, respectively.

Table 1: Selected learner-evaluation outcomes by implementation setting.

Questionnaire Indicator	Industry M (SD)	Industry % 4–5	HE M (SD)	HE %4–5
Content was well organised	4.52 (0.62)	93.3%	4.73 (0.64)	95.3%
Navigation through the module was smooth	4.35 (0.87)	82.3%	4.62 (0.67)	95.3%
Time allocated to the module was appropriate	3.79 (1.16)	66.1%	4.17 (1.14)	76.5%
Interest increased while progressing through the module	4.59 (0.56)	96.8%	4.69 (0.71)	94.1%
Quiz questions were aligned with the learning outcomes	4.55 (0.67)	93.6%	4.75 (0.65)	94.1%
Online sessions were dynamic and supported consolidation	4.44 (0.76)	87.1%	4.52 (0.81)	88.3%
Perceived acquisition of the learning outcomes	4.35 (0.70)	87.1%	4.53 (0.85)	90.6%

Note: Industry: n = 62; HE: n = 85. Items were rated on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree). The column *% 4–5* represents the proportion of positive responses (ratings 4–5).

Ratings concerning organisation and engagement were also high. The organisation of module content received mean ratings of 4.52 (SD = 0.62) in industry and 4.73 (SD = 0.64) in HE. Navigation was somewhat more variable in the industrial implementations (M = 4.35, SD = 0.87; 82.3% positive) than in HE (M = 4.62, SD = 0.67; 95.3% positive). Learners also reported increased interest as they progressed through the modules, with mean ratings of 4.59 (SD = 0.56) in industry and 4.69 (SD = 0.71) in HE; 96.8% and 94.1% of responses were positive, respectively.

In contrast, the adequacy of the time allocated to the modules was the least positively evaluated of the selected dimensions. The pooled rating was 3.79 (SD = 1.16) in industry, with 66.1% of responses selecting 4 or 5, compared with 4.17 (SD = 1.14) and 76.5% positive responses in HE. Module-level inspection indicated that lower ratings of time allocation were not confined to a single module, suggesting that this represented a broader implementation consideration rather than a module-specific issue.

5.3 Qualitative Learner Feedback

Open-ended responses were analysed separately for module organisation, learning resources, and teaching approach. Responses could be assigned to more than one category when multiple aspects were mentioned; therefore, percentages within each question are not mutually exclusive.

Table 2: Key findings from open-ended learner feedback on module organisation, learning resources, and teaching approach.

Question Area	Finding	Industry %	HE %	Total n (%)
Organisation – valued aspects	Clear structure and organisation	35.2	37.0	49 (36.3)
	Variety of learning resources/activities	38.9	34.6	49 (36.3)
Organisation – room for improvement	More time / lower workload	34.1	17.9	30 (23.4)
Resources – preferred	Quizzes	45.7	44.2	59 (44.7)
	Videos	39.1	47.7	59 (44.7)
Resources – least preferred	No disliked resource	34.2	34.6	40 (34.5)
	Videos	13.2	16.7	18 (15.5)
Teaching – valued	Instructor support and responsiveness	21.4	26.8	31 (25.0)
	Gamified/quiz-based interaction	19.0	25.6	29 (23.4)
	Clear explanations	16.7	17.1	21 (16.9)
Teaching – least valued	No disliked teaching aspect	52.8	50.0	59 (50.9)
	Synchronous-session schedule	8.3	15.0	15 (12.9)
	Time/workload pressure	19.4	5.0	11 (9.5)

Note: Percentages use question-specific substantive responses as denominators and are not mutually exclusive because multiple coding was permitted.

Regarding module organisation, the most frequently valued aspects were the variety of learning resources and activities and the clear structure of the modules, each identified in 49 of 135 substantive responses (36.3%). These patterns were similar across settings. The principal organisational concern was related to time and workload, mentioned in 30 of 128 responses (23.4%). This concern was proportionally more frequent in the industrial setting (34.1%) than in HE (17.9%), consistent with the lower quantitative ratings for adequacy of time allocation.

For learning resources, quizzes and videos were the most frequently preferred formats, each identified in 59 of 132 substantive responses (44.7%). Preference for quizzes was similar across the two settings, while videos were more frequently identified as a preferred resource in HE (47.7%) than in industry (39.1%). At the same time, 40 of 116 responses to the least-preferred-resource question (34.5%) explicitly indicated that no resource type was disliked. Among those identifying a specific least-preferred format, videos were mentioned most frequently (18; 15.5%). The appearance of videos among both the most- and least-preferred formats suggests variation in individual resource preferences and supports using complementary learning formats rather than relying on a single resource type.

Regarding the teaching approach, instructor support and responsiveness were the most frequently identified valued feature (25.0%), followed by gamified or quiz-based interaction during synchronous sessions (23.4%) and clear explanations (16.9%). When learners were asked what they liked least about the teaching approach, approximately half of the responses explicitly identified no disliked aspect (50.9%). Among specific concerns, the timing and scheduling of synchronous sessions was most frequently mentioned (12.9%), followed by general time constraints or workload pressure (9.5%).

Overall, the open-ended feedback contextualised the quantitative findings by identifying clear organisation, varied learning resources, instructor support, and interactive teaching as valued components of the learning experience. Synchronous sessions appeared to add value through interactive and gamified activities, although their timing and scheduling represented a practical challenge for some learners. General time and workload constraints also emerged as an implementation consideration, particularly in the industrial setting. Taken together, these findings support the potential transferability of the instructional design and pedagogical structure across industrial and HE contexts, while highlighting the importance of accommodating learners undertaking competency development alongside other responsibilities.

6. Conclusion

This paper presents a scalable, work-integrated e-learning model designed to support competency development across geographically distributed learners while accommodating different educational backgrounds and professional constraints. Implementation in both industrial and HE settings was associated with broadly positive learner perceptions, particularly regarding module organisation, learning resources, assessment alignment, instructor support, and interactive synchronous activities. The consistency of these perceptions across the two settings supports the model's potential applicability across different professional and educational contexts.

The findings also provide insight into specific features of the instructional design. High perceived achievement of learning outcomes, together with positive perceptions of instructor support and responsiveness, are consistent with previous research emphasising the importance of instructional support within flipped-learning environments (Thai et al., 2020; Gentile et al., 2022). Similarly, learners reported increasing interest as the modules progressed and frequently identified gamified and quiz-based synchronous activities as valued features. These findings are consistent with evidence that gamification can support learner motivation, engagement, and participation (Diaz and Estoque-Loñez, 2024; Su et al., 2025). Quizzes were also among the most frequently preferred learning resources, while quiz–learning outcome alignment was rated highly in both settings. The preference for quizzes is consistent with the rationale for incorporating retrieval practice and formative self-assessment into the learning pathway (Bertilsson et al., 2021). In parallel, the high perceived alignment between quiz questions and learning outcomes supports the model's constructively aligned assessment architecture. These associations should nevertheless be interpreted as learner perceptions and do not establish causal relationships between individual design elements and learning outcomes.

Importantly, time allocation was the least positively evaluated of the selected quantitative dimensions, while workload constraints and the timing of synchronous sessions also emerged from the open-ended feedback. This is particularly relevant to the model's objective of enabling learning alongside professional responsibilities. The modular architecture allows delivery to be compressed or extended according to organisational and learner needs; however, the findings suggest that flexibility should encompass not only asynchronous access to learning materials but also the overall pacing of the programme and the scheduling and spacing of synchronous activities.

The present evaluation is limited to self-reported learner perceptions and therefore does not establish objective competency acquisition, long-term learning transfer, or workplace impact. Moreover, the model is intended to complement rather than replace the laboratory and workplace-based training required for practical biomanufacturing competencies. Future research should therefore examine competency development and transfer using objective measures, evaluate longer-term workplace outcomes, and investigate how programme pacing and synchronous-session scheduling influence the experience of learners undertaking professional development alongside other responsibilities.

Acknowledgements

The authors gratefully acknowledge the industry and external SMEs whose expertise supported the scientific accuracy, industrial relevance, and practical applicability of the training materials.

Ethics Declaration

This article is based on a retrospective secondary analysis of educational data originally collected as part of the regular implementation and evaluation of an e-learning programme. At the time of data collection, the programme was not planned as a research study. Therefore, no separate consent for scientific secondary analysis and publication was obtained at that time.

The presented secondary analysis was reviewed prior to resubmission by the LUNEX Research Ethics Committee of LUNEX University of Applied Sciences and, by decision dated 27/08/2026 under file number 26_R_4910137, was approved.

AI Declaration

In developing the e-learning model, the instructors used artificial intelligence (AI) tools to support the development of QBs and English-language proofreading. The QBs were based on educational content originally created by the instructors, and all AI-assisted outputs were carefully reviewed, revised, and validated by the instructors before implementation.

In preparing this manuscript, ChatGPT (OpenAI) was used to support English-language editing and to assist with the descriptive qualitative content analysis, as detailed in the Methods. AI-assisted analytical outputs were reviewed against the original responses, ambiguous classifications were manually reviewed, and final coding decisions remained with the authors. The authors reviewed and edited all AI-assisted outputs and take full responsibility for the analysis, interpretation, and content of the manuscript.

References

Aktı Aslan, S., 2022. The effect of flipped classroom approach on learning achievement, online self-regulation and interaction in synchronous distance education. *Journal of Educational Technology & Online Learning,* 5(3), pp. 535–552. Available at <https://doi.org/10.31681/jetol.1094205>

Bada, S.O., 2015. Constructivism learning theory: A paradigm for teaching and learning. IOSR *Journal of Research & Method in Education*, 5(6), pp. 66–70. Available at <https://doi.org/10.9790/7388-05616670>

Bertilsson, F., Stenlund, T., Wiklund-Hornqvist, C. and Jonsson, B., 2021. Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning & Teaching*, 20(1), pp. 21–39. Available at <https://doi.org/10.1177/1475725720973494>

Chuang, S., 2021. The applications of constructivist learning theory and social learning theory on adult continuous development. *Performance Improvement*, 60(3), pp. 6–14. Available at https://doi.org/10.1002/pfi.21963

Cuetos, M.J., 2023. Application of the flipped classroom model to stimulate university students' learning with online education. *Journal of Technology and Science Education*, 13(1), pp. 368–380. Available at <https://doi.org/10.3926/jotse.1806>

Diaz, A.F. and Estoque-Loñez, H., 2024. A meta-analysis on the effectiveness of gamification on student learning achievement. *International Journal of Education in Mathematics, Science, and Technology (IJEMST),* 12(5), pp. 1236-1253. Available at <https://doi.org/10.46328/ijemst.4185>

European Commission, 2024. *Commission to boost biotechnology and biomanufacturing in the EU.* European Commission News*,* 26 June. Available at <https://blue-economy-observatory.ec.europa.eu/news/commission-boost-biotechnology-and-biomanufacturing-eu-2024-06-26_en> [Accessed 11 May 2026].

Gentile, E., Plantamura, P., Roselli, T. and Rossano, V., 2022. Flipped learning in corporate contexts: A case study. *in Proceedings of the 19th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2022).* Lisbon: IADIS Press, pp. 219–225.

Institut Pasteur de Dakar (IPD), Africa Centres for Disease Control and Prevention (Africa CDC) and South African Medical Research Council (SAMRC), 2023. *Biomanufacturing Workforce Development Post-Workshop Report*. Dakar: Institut Pasteur de Dakar. [Accessed 05 May 2026].

International Pharmaceutical Federation (FIP), 2017. *Research, development and evaluation strategies for pharmaceutical education and the workforce: A global report*. The Hague: International Pharmaceutical Federation. Available at <https://www.fip.org/educationreports> [Accessed 05 May 2026].

Khaldi, A., Bouzidi, R. and Nader, F., 2023. Gamification of e-learning in higher education: a systematic literature review. *Smart Learning Environments*, 10(10). Available at <https://doi.org/10.1186/s40561-023-00227-z>

Lopez Brown, P., 2017. Theme: The 21st century adult learner. *Educational Research and Reviews*, 12(8), pp. 540–548. Available at <https://doi.org/10.5897/ERR2016.2928>

Mohammed, S. and Kinyo, L., 2020. Constructivist theory as a foundation for the utilization of digital technology in the lifelong learning process. *Turkish Online Journal of Distance Education*, 21(4), pp. 90–109.

Mynhardt, M., Mwila, C., Habtemariam, M.K., Tshangela, A., Martinez, M., Ngongo, N., Kaseya, J. and Ndembi, N., 2024. Empowering Africa's healthcare future: The crucial role of human capital development in bio- and pharmaceutical manufacturing. *Journal of Public Health in Africa*. Available at <https://doi.org/10.4081/jphia.2024.2866>

NCBioNetwork, 2023. *Aseptic Gowning.* Available at <https://www.ncbionetwork.org/iet/sterilegowning/> [Accessed 08 May 2026].

O'Brien, E. and Hamburg, I., 2013. Organisational problem based learning and social communities for SMEs. *European Journal of Open, Distance and e-Learning*, 16(2), pp. 50–60.

PhET Interactive Simulations, 2026. *Concentration*. Available at <https://phet.colorado.edu/en/simulations/concentration> [Accessed 08 May 2026].

Santos-Ferreira, D., Guimarães, B., Ladeiras-Lopes, R., Gonçalves-Teixeira, P., Diaz, S.O., Ferreira, P., Gonçalves, F., Cardoso, R.G., Ferreira, M.A., Chaves, P.C., Fontes-Carvalho, R. and Leite-Moreira, A., 2024. Digital flashcards and medical physiology performance: a dose-dependent effect. *Advances in Physiology Education*, 48(1), pp. 80–87. Available at <https://doi.org/10.1152/advan.00138.2023>

Seifert, T., 2023. Fostering collaborative and constructivist learning through the flipped classroom during COVID-19, *Journal of Educators Online*.

Stegeager, N., Thomassen, A.O. and Laursen, E., 2013. Problem Based Learning in Continuing Education – Challenges and Opportunities. *Journal of Problem Based Learning in Higher Education*, 1(1), pp. 151–175.

Su, C.Y., Guo, Y., and Shen, J., 2025. The effects of gamified components on learning motivation, engagement, and teamwork in an online university course. *International Journal of Technology in Education (IJTE*), 8(2), pp. 467-481. Available at <https://doi.org/10.46328/ijte.1072>

Thai, N.T.T., De Wever, B. and Valcke, M., 2020. Feedback: an important key in the online environment of a flipped classroom setting. *Interactive Learning Environments*. Available at <https://doi.org/10.1080/10494820.2020.1815218>

Widodo, W., 2022. Online flipped classroom: Developing postgraduate science education students' critical thinking skills. *Journal of Science Learning*, 5(3), pp. 469–477. Available at <https://doi.org/10.17509/jsl.v5i3.43107>

Work in Health Foundation, 2024. *Bridging the biomanufacturing skills gap in Europe: Insights from experts.* Available at <https://workinhealth-foundation.org/biomanufacturing-skills-gap-europe/> [Accessed 05 May 2026].

Zeng, X. and Ruannakarn, P., 2023. Development of problem-based learning management activities to enhance the knowledge, skills, and interests of students. *Higher Education Studies*, 13(4), pp. 149–160. Available at <https://doi.org/10.5539/hes.v13n4p149>

Integrating ChatGPT into University English Teaching: Linguistic Performance, Autonomy, and Learning Experience

Jeany Argueta[1], Gisela Espinoza[2] and Víctor Sandoval[2]
[1]Investigación, Universidad Gerardo Barrios, San Miguel, El Salvador
[2]Desarrollo e Innovación, Universidad Gerardo Barrios, San Miguel, El Salvador

jlis22@ugb.edu.sv
ggarcia@ugb.edu.sv
victorsandoval@ugb.edu.sv

Abstract: This study examines how the planned integration of ChatGPT influences English language development, student autonomy, and the overall learning experience in a university English as a Foreign Language (EFL) Intermediate English II course in El Salvador. The aim is to analyze both quantifiable changes in language proficiency and students' interactions with artificial intelligence as a learning support tool, particularly in relation to self-regulation and independent learning processes. Furthermore, the study sought to understand how students perceive the usefulness and limitations of AI in academic contexts. A convergent mixed-methods case study design was implemented with twenty-three undergraduate students enrolled in an Intermediate English II course. Quantitative data were collected through TOEFL ITP simulation tests administered before and after the intervention to assess changes in language proficiency. Qualitative data were collected through a focus group of four participants, where students shared their perceptions, attitudes, and experiences interacting with ChatGPT during the learning process. Both datasets were analyzed separately and then integrated to provide a comprehensive interpretation of the findings. The quantitative results reveal a positive, though not statistically significant, trend in TOEFL ITP scores after the intervention. While quantifiable improvements in test performance were limited, the qualitative findings offer a clearer understanding of the learning process. Students reported using ChatGPT as an accessible and ongoing learning partner to practice language skills, clarify doubts, generate examples, and receive immediate feedback. This consistent interaction contributed to increased confidence, greater autonomy, and the development of self-regulated learning strategies. However, participants also expressed concerns about potential over-reliance on AI and occasional inaccuracies in the generated responses, underscoring the importance of maintaining critical awareness. This standalone study highlights the pedagogical potential of AI tools when intentionally integrated into language teaching and underscores the need to foster critical AI literacy so that such technologies enhance, rather than hinder, meaningful learning.

Keywords: TOEFL ITP, Artificial Intelligence, Higher Education, Learner Autonomy, Second Language Acquisition

1. Introduction

There are many debates on the rapid development of generative artificial intelligence (AI) and the digital transformation of higher education. Dempere et al. (2023) identify key benefits of ChatGPT in higher education, such as research assistance, automated grading, and improved human computer interaction, as well as concerns regarding academic integrity, online testing security, job displacement, digital literacy inequalities, and AI induced anxiety.

The integration of AI tools such as ChatGPT raises important theoretical considerations related to learner autonomy. Conceptualized by Holec (1981) as the learner´s ability to take charge of decisions concerning their learning. Autonomy has since been expanded by (Little, 1991; Benson, 2011) to include socially mediated and technologically supported dimensions

Recent research shows the potential of generative artificial intelligence tools to support language learning through feedback, personalized learning, practice activities, and facilitate interactive use, potentially enhancing students' engagement and independent learning (Kasneci *et al.*, 2023; Kohnke, Moorhouse and Zou, 2023) In the context of English as a Foreign Language (EFL), emerging research suggests that AI may support vocabulary development, writing practice, and grammar awareness by providing on-demand explanations and personalized language input (Yan, 2025). However, empirical evidence regarding the impact of these tools on broader language proficiency and learner autonomy remains limited. In addition, much of the existing research has been conducted in Asian and European educational contexts, leaving a gap in empirical studies examining the pedagogical integration of generative AI in Latin American universities and its relationship to measurable language development and student learning experiences.

TOEFL reports show that El Salvador ranked 47th globally in 2025 English proficiency (Education First, 2025), highlighting systemic challenges in English acquisition. At the institutional level, Universidad Gerardo Barrios reported a 3.6% failure rate in intermediate and advanced English courses, suggesting that a portion of students

continue to experience difficulties in achieving expected linguistic competencies. These factors point to the need for innovative yet pedagogically grounded interventions.

This study investigates the linguistic and experiential dimensions of ChatGPT with the following questions:

- What changes are observed in the group's linguistic performance throughout the semester?
- How do students perceive the usefulness, limitations, and academic ethics of using ChatGPT?
- What evidence of autonomy appears, and how is it related to the use of ChatGPT?

2. Background

The integration of Artificial Intelligence (AI) into higher education has generated growing interest in its potential to enhance language learning outcomes. In the context of English as a Foreign Language (EFL), technological tools are increasingly being used to support skill development, provide personalized feedback, and promote student autonomy.

2.1 English as a Foreign Language (EFL) in Higher Education

English has become the dominant language of academic communication, scientific publishing, and professional interaction worldwide (Arquez Muñoz, Tobón, and Cely Betancourt, 2025). In higher education institutions, English proficiency is increasingly positioned as a prerequisite for meeting institutional standards related to teaching quality, research productivity, and internationalization, particularly in non-English speaking contexts where mobility, access to academic literature, publication expectations, and accreditation processes often operate through English (Naka and Spahija, 2022).

Learning English as a Foreign Language (EFL) presents specific challenges, particularly in contexts where exposure to authentic language input is limited (Atrianus Toni, 2026). Within higher education, this global dominance increases the demands placed on students' academic writing and speaking, as EFL learners often face not only linguistic gaps but also cognitive and strategic difficulties, as well as affective barriers such as anxiety, factors that can limit performance even when learners understand course content (Pu and Chang, 2025; Wijaya, 2026). Empirical evidence from TOEFL ITP contexts further demonstrates that students' listening difficulties are influenced by both external factors, such as speaker accent, speed of delivery, and complex sentence structures, and internal factors, including limited vocabulary, lack of practice, concentration problems, and test-related anxiety, all of which contribute to persistently low proficiency scores in standardized assessments (Nuraida, 2025).

Consequently, although English proficiency is considered essential for academic success and institutional development, many students of English as a foreign language in higher education continue to face structural, linguistic, and affective constraints that prevent them from fully participating in academic practices.

2.2 TOEFL ITP as a Standardized Measure of English Proficiency

The TOEFL ITP (Institutional Testing Program), developed by the Educational Testing Service (ETS), is a widely used standardized assessment designed to measure English proficiency in academic contexts (ETS, 2026). The test evaluates three core components: Listening Comprehension, Structure and Written Expression, and Reading Comprehension. Scores range from 310 to 677 and provide a general indicator of language proficiency.

Standardized assessments such as TOEFL are designed to ensure reliability and validity (Khan, Imran, and Javed, 2025). Reliability refers to the consistency of scores, while validity refers to the degree to which the test measures academic English proficiency. Given its recognized reliability, validity, and predictive power, the TOEFL ITP is a standardized tool for assessing academic English proficiency and for conducting pre- and post-test assessments in the field of educational research.

2.3 Artificial Intelligence in Education

Artificial intelligence (AI) refers to a field of computer science focused on creating systems that can perform tasks typically associated with human intelligence, such as learning from experience, reasoning, solving problems, perceiving information, and processing language (Allen, 2020; Long and Magerko, 2020; Stryker and Kavlakoglu, 2025).

Adaptive learning systems adjust the difficulty of content based on student needs, providing immediate feedback and modifying instructional pace according to individual performance (Yang, 2025). Recent research also indicates that, while AI helps improve personalized learning and student engagement, its use must be carefully regulated to avoid overdependence and potential negative effects on critical thinking and academic integrity (Vieriu and Petrea, 2025).

When thoughtfully implemented, AI technologies can enhance learner autonomy and support differentiated instruction, while still requiring human oversight to ensure pedagogical and ethical alignment (Holmes, Bialik and Fadel, 2019). Therefore, the use of IA tools in education should not be understood as a substitute for teaching practices, but rather as a complementary tool to improve teaching and learning.

2.4 Artificial Intelligence in English Language Learning

AI applications in language learning are commonly applied in areas such as automated grammar and writing, speech recognition systems that improve pronunciation, intonation, and fluency, and conversational agents that provide simulated communicative interactions (Haciyeva, 2025). In EFL contexts, these technologies increase exposure to the English language and offer individualized practice opportunities that are often limited in traditional classrooms.

Artificial intelligence offers multiple benefits in foreign language learning, notably its ability to provide immediate feedback, personalize learning, offer continuous practice opportunities, provide adaptive learning experiences, and improve language skills such as grammar, vocabulary, and speaking (AlTwijri and Alghizzi, 2024; Bounif, 2026). They typically provide targeted instruction, automated feedback, and interactive activities that help students practice English more frequently and in a variety of contexts.

While AI technologies offer promising tools to support language development, their effectiveness will largely depend on how they are incorporated into teaching practices and how educators guide them. Therefore, AI should be considered a complementary resource for improving English language learning, rather than a substitute for teacher-led instruction and human interaction in the classroom.

2.5 Learner Autonomy

Learner autonomy was originally conceptualized by Holec, (1981) as “the ability to take charge of one’s own learning”, a definition that has become foundational in language education research. This concept implies that learners are responsible for every key aspect of the learning process, such as setting goals, selecting methods and strategies, and monitoring progress and evaluating outcomes (Little, 1991).

From this perspective, autonomy is not an innate trait, but a capacity that must be developed with pedagogical support. Holec emphasized that the shift of responsibility from the teacher to the learner represents a fundamental transformation in educational practices, positioning learners as active agents in their own learning process rather than passive recipients of knowledge.

In the context of language learning, student autonomy is related to self-directed learning, metacognitive awareness, and reflective practices. Autonomous students demonstrate the ability to make decisions about their own learning, adapt strategies to their needs, and engage in ongoing self-assessment. Consequently, the development of learner autonomy has been linked to greater motivation, more effective learning outcomes, and the promotion of lifelong learning skills.

3. Methodology

The study employed a convergent mixed methods case study designed to examine the integration of ChatGPT in a university Intermediate English II course. The focus was on a single cohort of students enrolled at Gerardo Barrios University in San Miguel, El Salvador.

3.1 Quantitative

The quantitative segment of this research follows a non-experimental, longitudinal, prospective design, involving two measurements throughout the process. The study adopts a relational scope, aimed at comparing the results obtained in the initial phase (pre-test) with those from the final stage (post-test). To achieve this, the quantitative data analysis was structured into three phases. First, an exploratory analysis was conducted, involving the descriptive treatment of the data and graphical representation of the variables (Hernández-

Sampieri and Mendoza, 2018). Second, a correlational analysis was performed to compare input and output results, based on the assumption of a positive trend in final performance (Okoye and Hosseini, 2024). Third, a hypothesis test was carried out using inferential statistics for related samples to determine the statistical significance of the observed differences (Al-kassab, 2022).

3.2 Qualitative

The qualitative element of this study explored students' perceptions of the integration of ChatGPT during their Intermediate English B2 Course. A focus group of four participants was conducted, and the data were analysed using thematic analysis following the approach proposed by Braun and Clarke (2006). The semi-structured interview held at the end of the semester allowed participants to reflect on their language learning experiences, the ways in which the tool supported their autonomy, and ethical considerations related to the use of AI.

The analysis followed these six phases: (1) Familiarization with data; (2) Initial coding; (3) Searching for themes; (4) Reviewing themes; (5) Defining themes; and (6) Producing report (Braun and Clarke, 2006).

Following this procedure, the coding process revealed several categories that were subsequently grouped into broader themes related to students' prior familiarity with AI tools, learner autonomy, and perceived limitations and risks of AI use. The creation of a codebook registered codes, definitions, excerpts, and themes ensures transparency and consistency during the analytical process.

4. Data Analysis

This section presents the findings organized according to the research questions, integrating quantitative and qualitative data. Tables and figures are included to support the interpretation of results and provide a clearer visualization of the findings.

4.1 Changes in Students' Linguistic Performance

RQ1: What changes are observed in the group's linguistic performance throughout the semester?

The quantitative analysis shows a general upward trend in students' linguistic performance across all components of the TOEFL ITP test. Mean scores increased from entry to exit levels in Listening Comprehension (41.4 to 42.8), Structure and Written Expression (39.4 to 41.0), Reading Comprehension (40.5 to 40.7), and the overall TOEFL ITP Level (404.0 to 415.1). These results suggest a tendency toward improvement in students' English proficiency over the course of the semester.

Table 1: Comparative table of descriptive statistics (entry vs. exit)

Statistical Parameter	Listening Comprehension		Structure and Written Expression		Reading Comprehension		TOEFL ITP Level	
	Entry	*Exit*	*Entry*	*Exit*	*Entry*	*Exit*	*Entry*	*Exit*
mean	41.4	42.8	39.4	41.0	40.5	40.7	404.0	415.1
std	7.4	6.7	6.0	5.9	5.9	5.3	54.7	53.2
min	31.0	33.0	32.0	33.0	32.0	33.0	333.0	340.0
25%	35.5	38.0	35.0	37.0	35.5	36.5	367.0	378.3
50%	40.0	42.8	38.0	41.0	39.0	40.7	397.0	415.1
75%	45.0	46.5	41.5	42.5	45.0	44.5	423.5	428.3
max	62.0	65.0	57.0	61.0	56.0	55.0	583.0	603.3

Note: There is a general upward trend in the mean scores across all components from the Entry Level to the Exit Level. This suggests a positive impact of the intervention or the period between the two assessments.

To further examine the distribution of scores, Figure 2 compares entry and exit levels. The findings suggest that although mean scores increased, the variability in performance also changed. This indicates that not all students improved at the same rate, and the distribution of scores at the exit level differs slightly from that at entry.

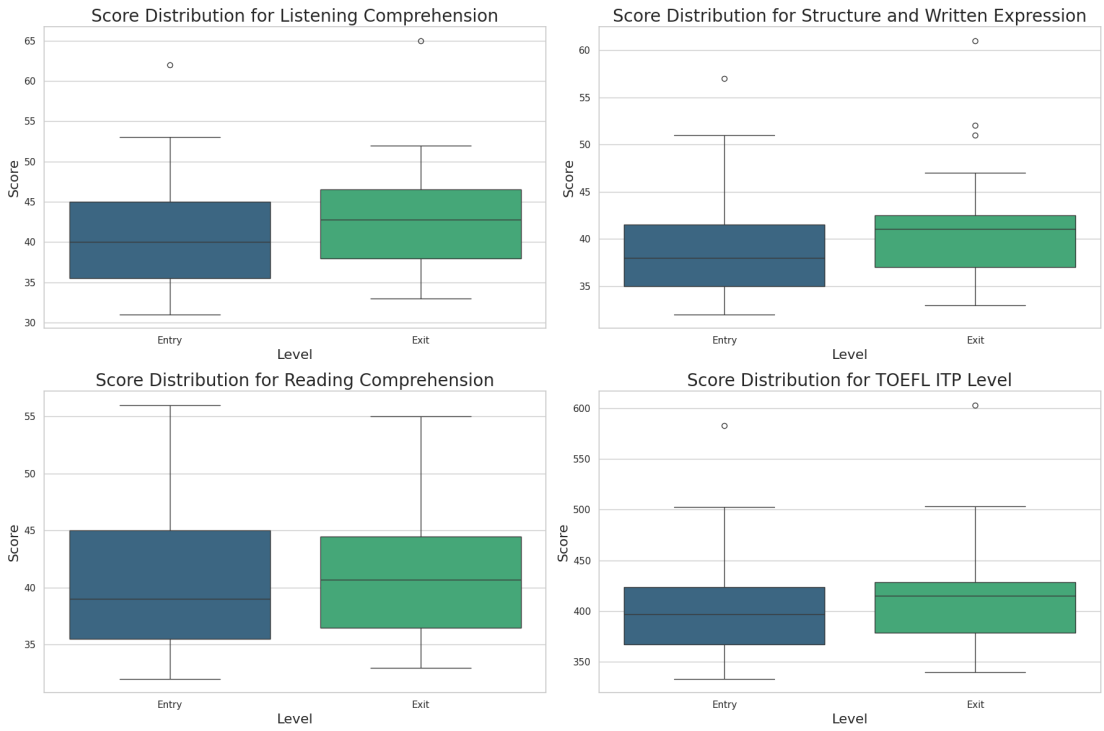


Figure 1: Box plot comparison of entry vs. exit TOEFL ITP scores

Note: For several components, the boxplots suggest that while mean scores increased, the spread (interquartile range) of the Exit scores might have slightly shifted or changed compared to the Entry scores. This means that not all students improved at the same rate, and the distribution of skills at the exit might be different than at entry.

In addition, the correlational analysis depicted in Figure 3 reveals a strong positive relationship between entry and exit scores (r = 0.7635), indicating that students' initial proficiency levels are closely associated with their final performance. This suggests that, although improvement is observed, student progress remains consistent with their initial level.

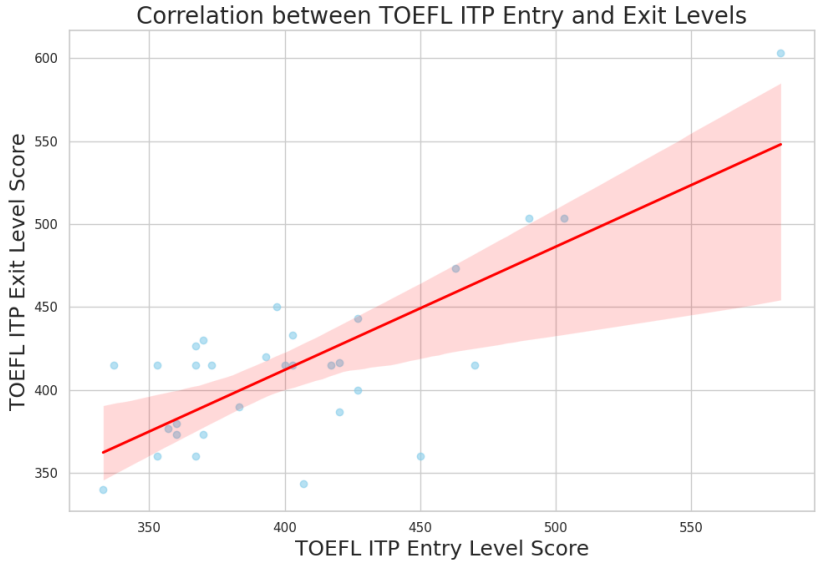


Figure 2: Scatter plot and linear correlation of TOEFL ITP entry vs. exit results

Note: The scatter plot with the trend line shows a positive trend. The points are clustered relatively close to the regression line, which exhibits an upward slope. This visually demonstrates that as the entry score (*X*-axis) increases, the exit score (*Y*-axis) also tends to increase.

To evaluate whether these observed differences were statistically significant, a one-tailed t-test was conducted, given that the study focuses on improvement. The null hypothesis (H_o) established that μ_exit ≤ μ_entry, while the alternative hypothesis (H_1) proposed that μ_exit > μ_entry. A significance level of α = 0.05 was applied. The results of the hypothesis test are presented in Table 2.

Table 2: Statistical significance results of the intervention

Parameter	Value	Observation
T-statistic	1.6652	This value represents the magnitude of the difference between the sample means relative to the variation within the samples. A positive value suggests that the Exit Level mean is higher than the Entry Level mean.
P-value (one-sided)	0.0531	This is the probability of observing a T-statistic as extreme as, or more extreme than, the one calculated, assuming the null hypothesis (μ_Exit ≤ μ_Entry) is true. It indicates the strength of evidence against the null hypothesis.
Significance Level (α)	0.05	This is the threshold for statistical significance. We compare the p-value against this level to make a decision about the null hypothesis. An alpha of 0.05 means we are willing to accept a 5% chance of rejecting a true null hypothesis (Type I error).
Decision	Fail to Reject H_0	Since the one-sided p-value (0.0531) is greater than the significance level (0.05), we do not have enough statistical evidence to reject the null hypothesis. "*The intervention with IA had no positive effect on the scores".*

Note: The test results indicate that there is no statistically significant evidence at the 0.05 level to suggest that the mean TOEFL ITP score at the Exit Level is higher than at the Entry Level. The observed difference in means could be due to random chance. This does not necessarily mean there was no improvement, but that any improvement was not large enough to be statistically distinguishable from chance at our chosen alpha level.

Although descriptive and correlational analyses show a general and consistent improvement in student performance, hypothesis testing indicates that this improvement is not statistically significant at the 95% confidence level and could therefore be attributed to chance. Consequently, despite the positive trend, further qualitative analysis is needed to better understand the impact of the intervention.

4.2 Students' Perceptions of ChatGPT Use

RQ2: How do students perceive the usefulness, limitations, and academic ethics of using ChatGPT?

The thematic analysis of the focus group identified four main themes related to students' experiences with the integration of ChatGPT: prior familiarity with AI, perceived support for language learning, learning practices with AI, and learner autonomy, as well as perceived limitations and risks.

4.2.1 Prior Familiarity with AI

Participants reported varying levels of experience with ChatGPT before the course. Some had already been introduced to the tool in previous academic contexts. For instance, one participant noted, *"In high school they taught us how to use it to improve our English" (Noah)*. Others indicated that their first exposure occurred through peer interaction at university, as reflected in the statement, *"Amelia showed me how to use it" (Emma)*. These comments show that students entered the course with varying degrees of familiarity with AI tools.

4.2.2 Perceived Support for Language Learning

The use of ChatGPT as a support tool for language learning emerged as a central theme. Participants reported using the tool for grammar explanations, vocabulary development, and practice exercises. For example, one student explained that *"If I ask it to explain grammar in detail, it becomes easier to understand" (Emma)*. Similarly, another participant described using the tool to analyze vocabulary in context: *"I can give it ten words and ask for the definition, what type of word it is, and how to use it in a sentence" (Amelia)*. Additionally, students reported using ChatGPT to generate practice materials, as illustrated by the comment, *"I ask it to create exercises for me to practice" (Olivia and Noah)*. The results show that ChatGPT was perceived as an important supplementary resource for language learning.

4.2.3 Learning Practices with AI

Participants described incorporating ChatGPT into their study practices, particularly as a tool for clarification and idea generation. One participant stated, *"It helps me clarify the topic more" (Olivia)*, while another emphasized efficiency, noting that *"It saved time from looking things up in the dictionary" (Amelia)*. Students also viewed ChatGPT as a starting point for generating ideas, as reflected in the statement, *"It is not bad to look for ideas in AI" (Olivia)*. These responses highlight ChatGPT's role as a supplementary reference tool integrated into traditional learning practices.

4.2.4 Perceived Limitations and Risks

Despite its benefits, participants also identified important limitations. Some students noted that the information generated by ChatGPT is not always reliable, as reflected in the statement, *"Sometimes the answer is not correct" (Noah)*. This led to an awareness of the need for verification, with one participant emphasizing that *"You have to check if the information is correct" (Noah)*. Additionally, concerns about overreliance on AI were expressed. One participant stated, *"If ChatGPT does everything, you do not produce the language" (Amelia)*. The students' reflections highlight their critical awareness of both the benefits and the risks associated with the use of AI.

4.3 Evidence of Learner Autonomy

RQ3: What evidence of autonomy appears, and how is it related to the use of ChatGPT?

The analysis also revealed evidence of learner autonomy in students' use of ChatGPT. Participants described using the tool to support independent learning, particularly when they encountered difficulties or needed additional resources.

Some students reported using ChatGPT to clarify concepts they did not understand. For example, one participant explained, *"I use it to understand topics I do not understand" (Noah)*. Others described using the tool to identify further learning materials, as reflected in the statement, *"It gives me websites or books to learn" (Olivia)*. In addition, students demonstrated selective and strategic use of ChatGPT. One participant noted, *"I only use it when I do not understand something" (Emma)*, suggesting a conscious decision to rely on the tool only when necessary. The responses provided suggest that ChatGPT can promote learner autonomy by enabling independent problem-solving, access to resources, and self-directed learning strategies.

5. Discussion

The findings of this study provide a nuanced perspective on the role played by artificial intelligence—specifically ChatGPT in supporting the learning of English as a foreign language (EFL) in higher education.

From a quantitative perspective, the results indicate a consistent upward trend in students' TOEFL ITP scores across all components. Although these improvements did not reach statistical significance, they suggest a positive tendency in language development. This finding can be interpreted in light of previous research highlighting the potential of artificial intelligence tools to enhance language learning through personalized feedback, increased practice opportunities, and adaptive learning environments (AlTwijri and Alghizzi, 2024; Bounif, 2026). However, the lack of statistical significance underscores the complexity of measuring the impact of AI-based interventions, suggesting that factors such as the duration of exposure, intensity of use, and learners' initial proficiency levels may mediate these outcomes.

The qualitative findings provide additional insights that help contextualize these results. Students perceive ChatGPT as a valuable tool for improving grammar, vocabulary, and idea generation, which is consistent with prior studies on AI in education (Yang, 2025; Haciyeva, 2025). However, these findings are interpreted within the framework of the present study and do not reproduce the authors' original wording. At the same time, participants demonstrated a critical awareness of the tool's limitations, particularly regarding the reliability of generated content. This perspective aligns with concerns raised in the literature about the risks of misinformation and the need for guided use of AI technologies in educational contexts (Vieriu and Petrea, 2025).

Importantly, the findings also suggest that ChatGPT may contribute to the development of learner autonomy. Students reported engaging in self-directed practices such as clarifying doubts, seeking additional resources, and using the tool selectively according to their needs. These behaviors reflect key dimensions of learner autonomy as conceptualized by Holec (1981) and Little (1991), particularly in terms of learners' ability to take responsibility

for their learning and regulate their strategies. In this sense, ChatGPT appears to function as a mediating tool that supports autonomous learning practices rather than replacing instructional processes.

The results of this study indicate that ChatGPT's educational value lies not only in its capacity to support language development but also in its potential to foster reflective and self-regulated learning. However, consistent with previous research (Holmes, Bialik, & Fadel, 2019), these benefits depend on its integration into structured learning environments and the presence of pedagogical guidance to ensure its ethical and effective use.

6. Conclusion and Future Work

This study explored the use of ChatGPT as an AI-assisted tool to support English as a Foreign Language (EFL) learning in higher education. Beyond the specific results obtained, the findings contribute to the ongoing discussion on the pedagogical role of generative AI in language education, particularly in contexts where empirical evidence remains limited. The study highlights the relevance of integrating AI tools within structured instructional frameworks, where their use can be aligned with learning objectives and guided by pedagogical criteria.

In this sense, the contribution of ChatGPT should be understood not merely in terms of performance outcomes, but as part of a broader shift toward more flexible, student-centered learning environments. Its use reflects emerging patterns of interaction between learners and technology, where access to immediate support, personalized input, and self-directed exploration plays an increasingly important role in the learning process.

This study has some limitations, including the short duration of the intervention and the specific context of the sample, which may limit the generalizability of the findings. Future research should focus on examining the long-term integration of AI tools in language learning, as well as the conditions under which they can be most effectively incorporated into formal educational settings. In particular, further studies are needed to better understand how pedagogical design, teacher mediation, and contextual factors shape the impact of AI on both language development and learner autonomy.

Ethics Declaration

This study followed the ethical guidelines of Universidad Gerardo Barrios (UGB) and the principles of the Declaration of Helsinki, with informed consent obtained from all participants.

Data were anonymized, securely stored, and participants were free to withdraw at any time without penalty.

AI Declaration

This paper was developed with the support of generative AI tools for language editing and refinement. All ideas, data analysis, and interpretations are the authors' own. AI was not used to generate data or replace critical academic judgment.

Acknowledgments

The authors gratefully acknowledge Alfredo Chévez (lchevez@ugb.edu.sv) for granting permission to conduct this study with his students and for planning the necessary resources for the experimental group. The authors also extend sincere appreciation to the students who participated in this research for their valuable collaboration.

References

Al-kassab, M. (2022) 'The use of one sample t-Test in the real data', *Journal of advances in mathematics*, 21, pp. 134–138. doi:10.24297/jam.v21i.9279.

Allen, G. (2020) Understanding AI Technology. Available at: https://apps.dtic.mil/sti/tr/pdf/AD1099286.pdf (Accessed: 1 March 2026).

AlTwijri, L. and Alghizzi, T.M. (2024) 'Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review', Heliyon, 10(10), p. e31053. doi:10.1016/j.heliyon.2024.e31053.

Arquez Muñoz, D.A., Tobón, M.F. and Cely Betancourt, B.L. (2025) 'The advantages of using English as the medium of instruction for globally oriented professional training. A theoretical review', Ciencia Latina Revista Científica Multidisciplinar, 9(4), pp. 8376–8398. doi:10.37811/cl_rcm.v9i4.19422.

Atrianus Toni (2026) 'Challenges faced by EFL students in learning english as a foreign language', Pedagogic Research-Applied Literacy Journal, 3(1), pp. 28–33. doi:10.70574/hshvd769.

Benson, P. (2011) Teaching and Researching Autonomy. 2nd edn. London: Routledge. Available at: https://www.routledge.com/Teaching-and-Researching-Autonomy-in-Language-Learning/Benson/p/book/9781408205013 (Accessed: 18 March 2026).

Bounif, H. (2026) 'Beneficios de la inteligencia artificial en el aprendizaje de lenguas extranjeras: revisión sistemática', Horizontes. Revista de Investigación en Ciencias de la Educación, 10(41), pp. 235–258. doi:10.33996/revistahorizontes.v10i41.1206.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', Qualitative Research in Psychology, 3(2), pp. 77–101. doi:10.1191/1478088706qp063oa.

Education First (2025) EF EPI EF english proficiency index. Available at: https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2025/ef-epi-2025-english.pdf (Accessed: 17 March 2026).

ETS (2026) The TOEFL ITP® Assessment Series. Available at: https://www.ets.org/toefl/itp.html (Accessed: 1 March 2026).

Haciyeva, A. (2025) 'Integration of artificial intelligence into language teaching', EuroGlobal Journal of Linguistics and Language Education, 2(5), pp. 99–106. doi:10.69760/egjlle.2505008.

Hernández-Sampieri, R. and Mendoza, C. (2018) 'Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta', 10(18), p. 95. doi:10.22201/fesc.20072236e.2019.10.18.6.

Holec, H. (1981) Autonomy and foreign language learning. Oxford: Pergamon Press.

Holmes, W., Bialik, M. and Fadel, C. (2019) Artificial Intelligence in Education. Promise and Implications for Teaching and Learning. Available at: https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning (Accessed: 2 March 2026).

Kasneci, E. et al. (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', Learning and Individual Differences, 103, p. 102274. doi:10.1016/j.lindif.2023.102274.

Khan, A., Imran, M. and Javed, M. (2025) 'Language testing and assessment: Validity and reliability in English proficiency exams', Contemporary journal of social science review, 03(01), pp. 1950–1959. Available at: https://www.researchgate.net/publication/397488435_LANGUAGE_TESTING_AND_ASSESSMENT_VALIDITY_AND_RELIABILITY_IN_ENGLISH_PROFICIENCY_EXAMS (Accessed: 1 March 2026).

Kohnke, L., Moorhouse, B.L. and Zou, D. (2023) 'ChatGPT for Language Teaching and Learning', RELC Journal, 54(2), pp. 537–550. doi:10.1177/00336882231162868.

Little, D. (1991) Learner autonomy. Authentik Language Learning Resources. Available at: https://www.researchgate.net/publication/259874253_Learner_Autonomy_1_Definitions_Issues_and_Problems (Accessed: 17 March 2026).

Long, D. and Magerko, B. (2020) 'What is AI literacy? Competencies and design considerations', in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, pp. 1–16. doi:10.1145/3313831.3376727.

Naka, L. and Spahija, D. (2022) 'Impact of English language as a human capital in the Higher Education Institutions' development strategy', Corporate and Business Strategy Review, 3(2 special issue), pp. 262–272. doi:10.22495/cbsrv3i2siart7.

Nuraida, N. (2025) 'Unpacking listening barriers: an investigation into EFL learners' difficulties in the TOEFL ITP context', GLOBAL: Education Language and Humanity Journal, 3(1), pp. 23–38. Available at: https://journal.pcn.ac.id/index.php/global/article/view/98/67 (Accessed: 1 March 2026).

Okoye, K. and Hosseini, S. (2024) 'Correlation Tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho', in R Programming. Singapore: Springer Nature Singapore, pp. 247–277. doi:10.1007/978-981-97-3385-9_12.

Pu, P. and Chang, D.Y.-S. (2025) 'Effects of different input modes on blended EFL speaking instruction: a quasi-experimental study', Computer Assisted Language Learning, 38(5–6), pp. 1223–1248. doi:10.1080/09588221.2023.2273853.

Stryker, C. and Kavlakoglu, E. (2025) What is artificial intelligence (AI)? Available at: https://www.ibm.com/think/topics/artificial-intelligence?utm_source=chatgpt.com (Accessed: 1 March 2026).

Vieriu, A.M. and Petrea, G. (2025) 'The impact of artificial intelligence (AI) on students' academic development', Education Sciences, 15(3), p. 343. doi:10.3390/educsci15030343.

Wijaya, K.F. (2026) 'Challenges in teaching academic writing to EFL learners in higher education contexts', ELTR Journal, 10(1), pp. 43–57. doi:10.37147/eltr.v10i1.276.

Yang, L. (2025) 'The Potential of AI in Education: Personalizing Learning', International Journal of Artificial Intelligence for Science (IJAI4S), 1(2). doi:10.63619/ijai4s.v1i2.003.

Personalised Formative Assessment in Ukrainian Higher Education: Design and Empirical Evaluation of the SMART-PL Model

Artem Boyarchuk[1] and Daria Shteinbrekher[2]
[1]Tallinn University of Technology, Tallinn, Estonia
[2]National Aerospace University, Kharkiv, Ukraine

artem.boyarchuk@taltech.ee
d.shteinbrekher@gmail.com

Abstract: This paper reports the design, implementation and empirical evaluation of a personalised learning (PL) model developed for Ukrainian higher education during 2022–2025, under the Erasmus+ Capacity Building project SMART-PL – *Students' Personalised Learning Model, Based on the Virtual Learning Environment of Intellectual Tutoring "Learning with No Limits"*. The project was coordinated by Tallinn University of Technology and implemented by six Ukrainian universities. Two contributions are advanced. The first is a four-stage design pattern for consortium-level PL adoption in transitional higher education systems: comparative benchmarking of European and Ukrainian practice; negotiated guidelines specifying PL principles, actors, stakeholders and a piloting roadmap; an explicit metric framework for knowledge acquisition and academic progression; and platform-mediated piloting through an information-diagnostic Moodle-based platform with instructor-authored, rule-mapped feedback blocks. The second is empirical evidence from a one-semester pilot (mid-January to mid-June 2025) across the six Ukrainian universities, engaging 772 students and 64 academic staff across all ten ESCO-aligned ICT professional qualification areas. Survey response rates of 42–71% exceeded thresholds typically required for generalisable findings in educational research. The Improvement Curve metric – relative score change between consecutive attempts on the same competency-aligned test – showed a median gain of 17% (range 9–38%) between the first and second attempt, with upward trajectories sustained across third, fourth and fifth attempts and second-and-subsequent attempts systematically crossing the platform's competency threshold of approximately 8/20. The evidence supports the claim that system-generated feedback tied to identified knowledge gaps produces measurable and consistent learning gains under repeated attempts. The four-stage design pattern is transferable to other Eastern Partnership and accession-track higher education systems.

Keywords: Personalised Formative Assessment, Remediation Feedback, Competency-Based Assessment, Moodle Virtual Learning Environment, Erasmus+ CBHE

1. Introduction

The Ukrainian higher education system entered the period 2022–2025 under a layered set of pressures. The full-scale invasion that began in February 2022 disrupted university infrastructure, displaced students and staff, and reshaped admission cohorts to include returning veterans, internally displaced persons and learners forced into asynchronous or hybrid modes by air-raid alerts (Lugovyi et al, 2023; Londar et al, 2023). Ukraine's candidate-country status and its long-standing Bologna commitments have, at the same time, intensified the alignment work required of universities on quality assurance, student-centred learning and recognition of non-formal credit (Kremen at al, 2021). The combination of these factors is crucial: a HE sector required to deliver flexibility, inclusivity and measurable quality simultaneously, with reduced institutional capacity and a student population.

This paper reports on the design, implementation and empirical evaluation of a personalised learning (PL) model developed under the Erasmus+ CBHE project SMART-PL — *Students' Personalised Learning Model, Based on the Virtual Learning Environment of Intellectual Tutoring "Learning with No Limits"* (2023–2025). The project was coordinated by Tallinn University of Technology and a consortium of six Ukrainian universities, supported by the Institute of Higher Education of the National Academy of Educational Sciences of Ukraine (IHE NAESU) and the National Agency for Higher Education Quality Assurance (NAQA). The model combines a Moodle-based information-diagnostic platform with hybrid co-working centres operating under a *Learning with No Limits* principle of access regardless of age, socio-economic status or modality of participation.

The remainder of the paper presents the model, the data and the limitations of the design, and identifies the elements that are transferable to other Eastern Partnership and accession-track higher education systems facing comparable combinations of pressure.

2. Background and Related Work

2.1 Formative Feedback, Personalised Learning and Adaptive Systems

The empirical case for tightly-coupled feedback as the active ingredient in instructional productivity was made by Black et al (2009), whose synthesis of more than 250 studies reported effect sizes between 0.4 and 0.7 standard deviations with the most pronounced gains accruing to lower-achieving students. Bloom's (1984) "2-sigma" observation – that one-to-one tutoring lifts learners roughly two standard deviations above conventional group instruction – has functioned ever since as the benchmark against which technology-mediated personalisation is measured.

The project does not operationalise personalisation as learner-driven pathway choice, algorithmic content sequencing or adaptive difficulty selection. It operationalises personalisation as *personalised formative assessment and remediation*: a student's item-level response pattern on a competency-aligned diagnostic test determines which instructor-authored feedback block, and which tier of curated remediation resources, that student receives. The unit of personalisation is the diagnosed knowledge gap. Two students sitting the same test on the same competency receive different remediation if their item-level error patterns and score bands differ. This places the model within the data-driven-instruction and mastery-progression strands of the personalised-learning literature (Means and Neisler, 2018; Pane et al, 2017) and it follows directly from the consortium's decision to keep feedback authorship with the responsible instructor for reasons of pedagogical coherence and accreditation traceability.

2.2 The Ukrainian Higher Education Context

Two strands of recent Ukrainian scholarship frame the SMART-PL setting. The first documents the wartime transformation of the sector: institutional displacement, contracted physical capacity, reconfigured admission processes and the rapid normalisation of distance and hybrid delivery (Lugovyi et al, 2023). The second strand concerns the longer-running alignment with the European Higher Education Area, which includes competency-based curricula, learning outcomes, internal quality assurance and individual educational trajectories formalised in the 2014 Law on Higher Education. What is conspicuously thin in this literature is multi-institutional empirical evidence on how a platform-mediated personalised learning model performs under Ukrainian wartime conditions, as distinct from the better-studied Anglo-American and Western European contexts. The project reported in this paper is intended to address that gap.

The paper initiates two main research questions. RQ1: what design pattern emerges when a European–Ukrainian consortium negotiates a personalised learning model for shared implementation? RQ2: does platform-mediated formative feedback, tied directly to identified knowledge gaps and tracked by an Improvement Curve metric, produce consistent learning gains under repeated assessment?

3. Method: the Four-Stage Design

SMART-PL was structured as a sequential design-and-evaluation pipeline. The four stages ran in dependent order: each stage produced an artefact that constrained the next. Stage I conducted a structured comparative benchmarking of European and Ukrainian PL practice across seven institutional dimensions (Section 4). Stage II converted that benchmarking output into consortium-level guidelines specifying PL principles, actors, stakeholders and a piloting roadmap (Section 5). Stage III defined the metric framework – test items, instructor-authored feedback blocks linked to identified gaps, a cut-off threshold and an Improvement Curve metric – and set it up as a Moodle-based platform (Section 6). Stage IV deployed the platform in a one-semester pilot at the six Ukrainian partner universities (Section 7).

The consortium comprised TalTech (coordinator, Estonia), KU Leuven (Belgium), Politechnika Wrocławska (Poland) and six Ukrainian universities – Khmelnytskyi National University (KhNU), V. N. Karazin Kharkiv National University (KKNU), Kryvyi Rih National University (KNU), Ivan Franko National University of Lviv (IFNUL), Odesa National Polytechnic University (ONPU) and Kherson State University (KSU) – supported by IHE NAESU and NAQA.

3.1 Stage I — Comparative Benchmarking of European and Ukrainian PL Practice

3.1.1 Purpose and Design

The first stage was conceived as a structured mapping intended to establish what each partner was actually doing, where the systemic gaps lay and which gaps were tractable within a three-year project. The exercise was led by IHE NAESU and conducted across seven institutional dimensions: (i) key features of the PL model in use; (ii) institutional policies framing PL; (iii) infrastructure, environments, tools and LMS supporting PL; (iv) departments and staff with named responsibility for PL; (v) the integration of PL within the institutional quality assurance system; and (vi) student-facing PL practices, including individual learning pathways, elective choice, formative assessment and tutoring (Kalashnikova et al, 2023). European partners (TalTech, KU Leuven, PWR) supplied institutional self-descriptions across the given dimensions, Ukrainian partners (KhNU, KKNU, KNU, IFNUL, ONPU, KSU) supplied the same self-descriptions and, in addition, conducted staff and student surveys to describe the *current* state.

3.1.2 Findings

Three categories of finding emerged. The first concerns institutional infrastructure for PL. European partners had named, permanently resourced units carrying institutional responsibility for personalised teaching and learning. Ukrainian partners reported diffuse responsibility, typically distributed across deans' offices, methodical councils and individual chairs, and LMS use that was content-delivery oriented. Formative assessment was present in Ukrainian practice but embedded in internal quality assurance cycles, in contrast to European partners where formative cycles were a recognised input to programme-level review.

The second category concerns individual learning trajectories. Ukraine's 2014 Law on Higher Education and subsequent NAQA standards confirm the right of students to individual educational trajectories, and Ukraine is one of the few EHEA member countries to define student-centred learning in national legislation. Survey responses nevertheless indicated that practical implementation was bounded by rigid programme structures, narrow elective windows and weak mechanisms for recognising non-formal learning.

The third category concerns inclusive provision. Ukrainian partners reported more developed institutional services for students with disabilities, internally displaced persons and war veterans than the European partners, reflecting national policy priorities under wartime conditions and predating SMART-PL: dedicated inclusive pedagogy laboratories at KhNU, the Centre for Veteran Development at KKNU, and "Learning Without Limits" centres at ONPU were already in operation when the project began. This finding reframed the consortium's understanding of what was being benchmarked: a mutual mapping in which inclusivity infrastructure flowed in Ukraine – EU direction (Tsybuliak et al, 2024).

3.2 Stage II — Consortium Guidelines: Principles, Actors, Stakeholders, Roadmap

3.2.1 The Guidelines

A consortium-level guideline (document specifying personalised learning principles, actors, stakeholders and a piloting roadmap) should not be read as a translation of European practice into a Ukrainian setting. The principles were those that the six Ukrainian partners could implement within the constraints of Ukrainian programme accreditation procedures, NAQA Sectoral Expert Council expectations, and the 2014 Law on Higher Education (Kalashnikova et al, 2022). The European partners contributed practice and pedagogic argument and the Ukrainian partners contributed regulatory feasibility tests.

3.2.2 The Five Principles

The guidelines settled on five operative principles, each defined sufficiently to act on (Morkun et al, 2024). *Individualisation* requires a documented diagnostic step – a baseline assessment of strengths, weaknesses and learning profile – before personalised pathway decisions are made. *Flexibility* requires that institutions offer multiple modes of access (face-to-face, asynchronous online, hybrid) and alternative assessment formats (portfolios, projects, presentations) so that students with work, caring or military responsibilities are not limited by mode of participation (Tsybuliak et al, 2024). *Engagement* is implemented through active learning techniques, structured tutor–student dialogue and tutor-supported reflection. *Data-driven decision-making* requires institutions to use platform-generated assessment evidence as input to programme-level review. *Continuous*

improvement requires an annual revision cycle for personalised learning materials and feedback blocks, with ethical data handling treated at every step.

3.2.3 Actors and Stakeholders

The actor map distinguishes three layers. The classroom layer is the student–instructor communication level, mediated by the platform. The institutional layer comprises departmental quality assurance officers, methodical councils, deans' offices and dedicated inclusion units (inclusive pedagogy laboratory at KhNU, the Centre for Veteran Development at KKNU and the "Learning Without Limits" centre at ONPU). The system layer comprises NAQA, IHE NAESU and external employer stakeholders, including the IT-cluster cooperation network. The employer presence is important: student–instructor–employer engagement at the Co-working Centres is the mechanism by which curriculum revisions initiated under SMART-PL remain calibrated to labour-market expectations.

3.2.4 The Piloting Roadmap

The roadmap specifies pre-pilot, pilot and post-pilot phases. The pre-pilot phase, completed between September 2024 and January 2025, required each instructor to review and enhance course content, design assessment rubrics and scoring scales aligned with the platform's metric framework, develop a bank of formative and summative test items, and format all resources to platform metadata standards. Technical leads from each Ukrainian partner demonstrated integrated course materials and confirmed launch readiness. The pilot phase is reported in Section 7. The post-pilot phase, ongoing as this paper is written, comprises a second-iteration testing cycle scheduled for the spring 2026 semester and the integration of revised teaching materials into permanent programme accreditation files.

3.3 Stage III — Metric Framework for Knowledge Acquisition and Academic Progression

3.3.1 What is Being Measured

Academic progression is defined as the direction and magnitude of change in a student's measured competency across repeated attempts on the same competency-aligned test, as distinct from the level attained at any single attempt. It is operationalised through two linked quantities. The first is the Improvement Rate (IR) between consecutive attempts, $IR = (S_{n+1} - S_n) / S_n \times 100\%$, where S_n is the score on the n-th attempt; the ordered sequence of IR values across attempts one to five constitutes the student's Improvement Curve. The second is the position of each attempt relative to the competency cut-off of approximately 8/20. A student is recorded as *progressing* when the Improvement Curve trends upward and as having *attained competency* when an attempt crosses the cut-off. The two are reported jointly: an upward curve that never crosses the cut-off evidences movement without mastery, and an above-cut-off score without an upward trajectory evidences attainment without demonstrated progression. The empirical claim of the paper requires both conditions to hold across the pilot population.

3.3.2 Operational Design

Each pilot discipline was aligned to one of ten ESCO Information and Communications Technology professional qualification fields (e.g. 2521.3 Database developer; 2512.4 Software developer; 2511.3 Data analyst), within each field the responsible Ukrainian partner identified the essential skills to be tested. For each skill, instructors authored a bank of test items in single-correct multiple-choice format with one correct answer and three distractors. Each test comprised 20 items scored on a 20-point scale. Alongside the item bank, each instructor authored a parallel bank of feedback blocks: short narrative remediation texts, each mapped to one or more specific test items, prescribing curated learning resources (textbooks, online courses, hands-on practice platforms, research papers) calibrated to the score band achieved.

Students in the lowest score band receive foundational resources such as introductory textbooks and entry-level practice platforms (Codecademy, freeCodeCamp). Mid-band students receive intermediate-level course material, project-based exercises and applied case studies. High-band students (roughly 9–10 out of 10) receive advanced materials – specialised handbooks, peer-reviewed publications, conference proceedings. The tiering is rule-based. The platform applies deterministic mapping rules from score and item-pattern to feedback block. Algorithmically generated feedback was considered and rejected for two reasons: pedagogical coherence between assessment and remediation requires that the same instructor designs both, and Ukrainian programme

accreditation procedures require traceable academic authorship of all assessment-related instructional content (Kremen et al, 2021; Kalashnikova et al, 2022).

3.3.3 The Improvement Curve Metric

The principal outcome metric is the Improvement Rate (IR) between consecutive attempts:

$$I\,R = (S_2 - S_1) / S_1 * 100\%$$

where S_1 is the first-attempt score and S_2 is the second-attempt score on the same competency-aligned test (Boyarchuk et al, 2025). Plotted across attempts 1 through 5 for each student, the resulting trajectory is the Improvement Curve. Two parameters provide criterion-referenced interpretation. The first is the cut-off score, set at approximately 8/20, which represents the platform's threshold for demonstrated competency in the tested skill. The second is the trajectory direction: an uptrending curve is taken as evidence that the remediation loop (test → diagnosed gap → curated resources → re-test) is functioning, while a flat or descending trajectory indicates that either the gap diagnosis or the remediation input has failed for that particular student. The metric is simple: it does not partial out item-level difficulty, test-wiseness or item-familiarity effects, and the constraints this imposes on causal interpretation are returned to in the Discussion.

3.3.4 Platform Used

The metric framework is implemented in a Moodle-based platform developed by Khmelnytskyi National University. The architecture of a SMART-PL platform includes a relational database for users, courses, items, attempts and feedback blocks; a competency-assessment module that scores attempts and resolves the feedback rule-set; a personalised educational diagram module that visualises the Improvement Curve and projected trajectory against the cut-off; a content management subsystem for course materials and a communication subsystem for student–instructor dialogue, configured with data-handling protocols per the consortium's D.3.1 sharing policy (Lysenko et al, 2024).

3.4 Stage IV — Piloting and Empirical Evaluation

3.4.1 Pilot Scope and Participants

This stage addresses the second research question stated in Section 2, refined into the operational questions the pilot was designed to test. RQ2a: Does platform-mediated formative feedback, tied to diagnosed knowledge gaps, produce an upward Improvement Curve across repeated attempts on the same competency-aligned test? RQ2b: Do repeated attempts move students across the competency cut-off of approximately 8/20 - that is, from diagnosed gap to demonstrated competency? RQ2c: Is the pattern consistent across the six partner institutions and the ten ESCO-aligned ICT qualification fields?

The pilot ran for one semester, from mid-January to mid-June 2025, at the six Ukrainian partner universities. Participation was embedded in regular course activity: students enrolled in pilot disciplines accessed the platform as part of their normal study load, and the testing constituted formative-assessment exercises within those courses. A total of 772 students completed at least one assessment over the semester, supported by 64 academic staff who authored test items, feedback blocks and remediation materials. Coverage across the partner institutions was uneven, reflecting the differing scale of the participating departments: 355 students at KhNU, 181 at KKNU, with the remaining 236 distributed in descending order across IFNUL, KSU, ONPU and KNU. The ten ESCO-aligned ICT professional qualification areas covered by the consortium were all represented, from Databases and Algorithmisation through Web Development, Data Analysis, Cybersecurity, Automation Technology and IT Project Management.

Two data sources were collected. The first, platform-generated, comprised the record of each test attempt: item-level responses, item-level correctness, total score on a 20-point scale, and the feedback block delivered. The second, survey-based, comprised two harmonised Google Forms instruments (one for students, one for teachers). There were distributed over a three-week window at the end of the semester. Combined response rates ranged from 42% to 71% across departments.

An illustrative case. Consider a student (referred to here as S-114) enrolled in a Data Analysis discipline mapped to ESCO field 2511.3 (Data analyst) at one partner university. The competency test comprised 20 single-correct multiple-choice items distributed across four skill clusters: data wrangling, descriptive statistics, SQL aggregation and querying, and visualisation. On the first attempt, S-114 scored 6/20 (below the 8/20 competency cut-off).

The item-level record showed competence concentrated in visualisation (4 of 5 items correct) but persistent failure in SQL aggregation (1 of 6) and descriptive statistics (1 of 5). The platform's competency-assessment module resolved this pattern to a low-band feedback block, which delivered a short remediation narrative naming the two weak clusters and prescribing foundational resources: an introductory relational-querying module and guided exercises on a hands-on practice platform.

On the second attempt, taken nine days later after the student engaged the prescribed material, S-114 scored 8/20, an Improvement Rate of 33% [(8 − 6) / 6 × 100], crossing the cut-off into demonstrated competency. The gains were localised to the two diagnosed clusters (SQL aggregation rose to 4 of 6), confirming that the lift tracked the remediation rather than general test familiarity. The third attempt reached 11/20 (IR 37.5%) and triggered a mid-band feedback block with project-based exercises; the fourth and fifth attempts reached 13/20 and 14/20 (IR 18.2% then 7.7%), the declining per-attempt gains characteristic of a curve approaching the upper scale. Plotted in the personalised-educational-diagram module, S-114's Improvement Curve is monotonically upward and crosses the cut-off at the second attempt – the joint signature of progression and attainment that the metric framework is designed to detect.

3.4.2 Findings: the Improvement Curve

The principal finding is a consistent positive trajectory across repeated attempts. The Improvement Rate distribution between the first and second attempts spanned 9% to 38%, with a median of 17% across the full pilot population. The distribution is right-skewed: the modal student gained moderately, while a substantial sub-population – most evident among students whose first attempt fell well below the cut-off – gained more steeply. This is the pattern Black et al (1998) identified as the signature effect of well-designed formative feedback, with the strongest gains concentrated among lower-achieving students.

Trajectories beyond the second attempt are consistent with the same mechanism. Across students who attempted three, four or five times on the same competency-aligned test, the curve continued to trend upward, with declining per-attempt gains as the curve approached the upper end of the scale (Figure 1, Figure 2, Figure 3). Some five-attempt trajectories were non-monotonic, with a partial dip on the fourth attempt followed by recovery on the fifth; this is consistent with students attempting items in a more challenging configuration after exhausting the easier remediation resources. Second and subsequent attempts systematically crossed the platform's competency cut-off of approximately 8/20, which means students were not improving in the abstract but were moving into demonstrated competency on the assessed skill. Two questions are therefore answered simultaneously by the dataset: *is the student improving?* (trajectory analysis) and *has the student reached competency?* (threshold analysis). Both questions receive a positive answer across the pilot population.

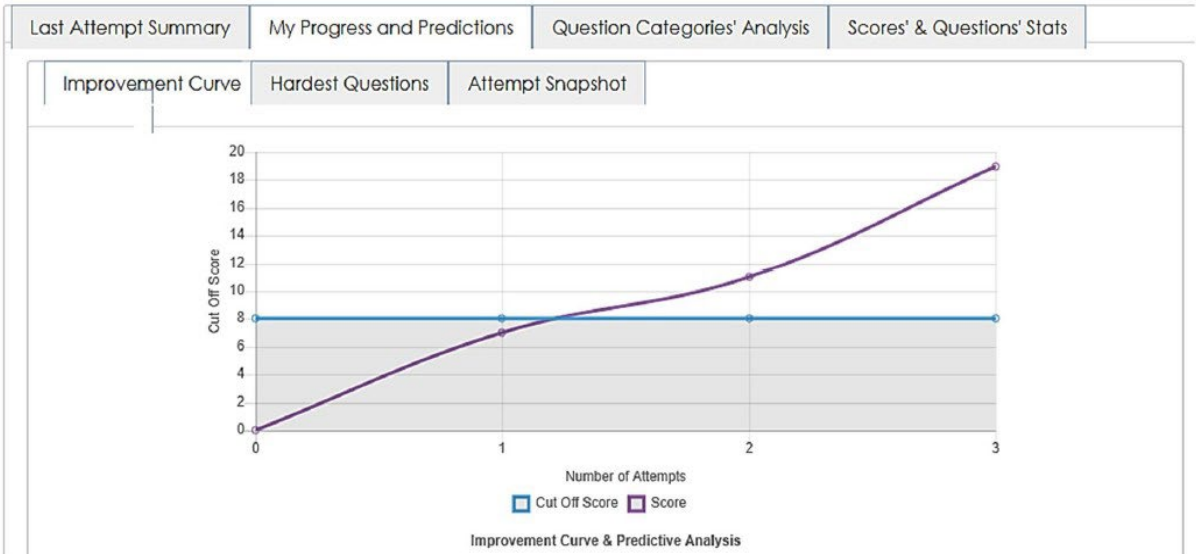


Figure 1: Improvement Curve for 3-attempt students

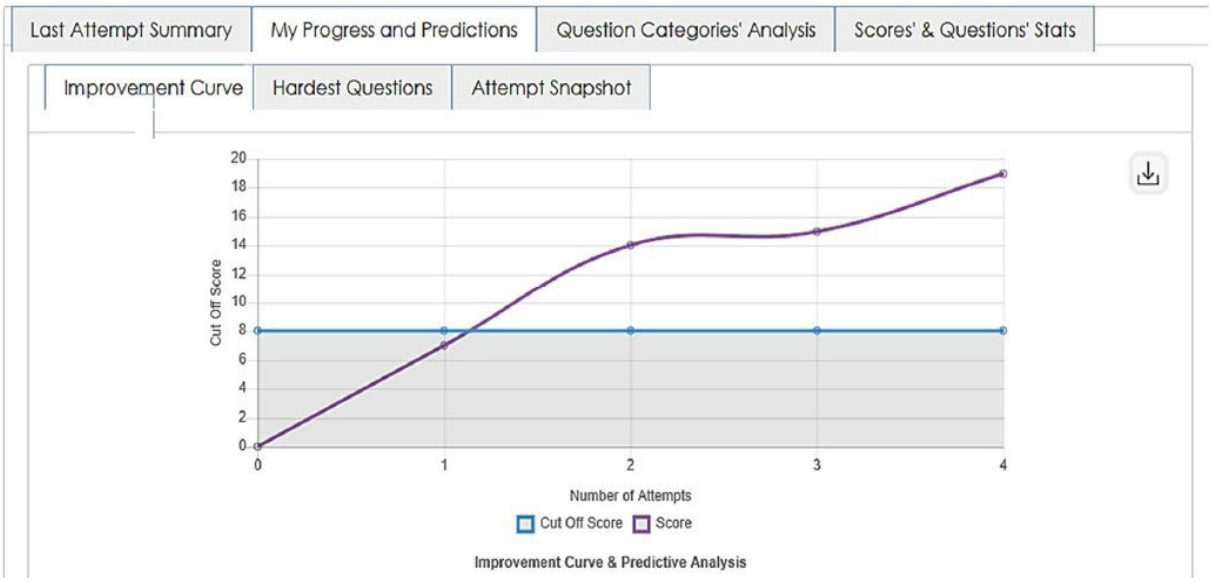


Figure 2: Improvement Curve for 4-attempt students

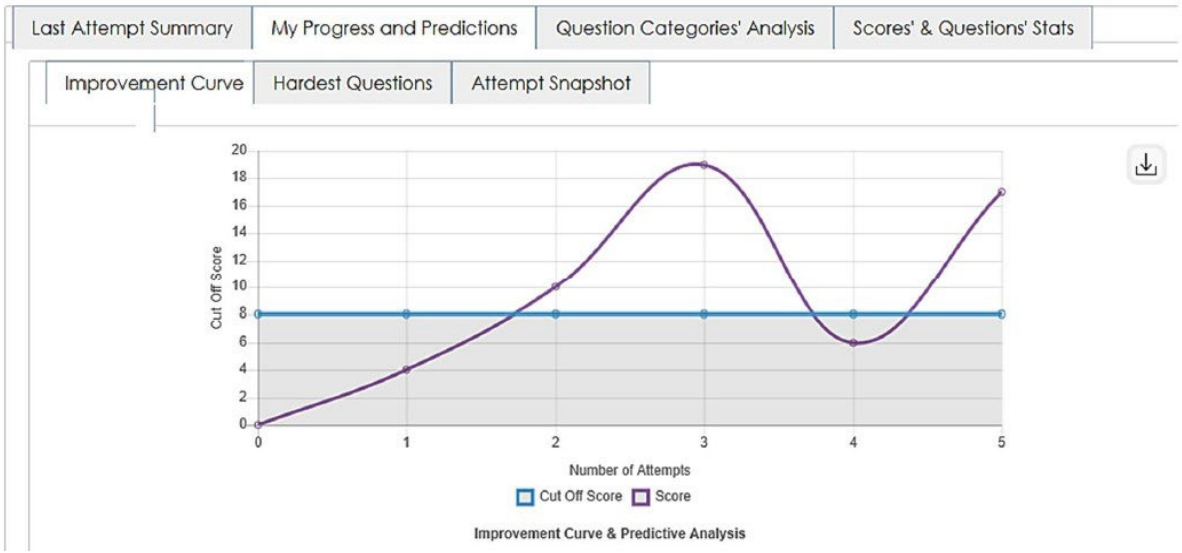


Figure 3: Improvement Curve for 5-attempt students

3.4.3 Survey Complement

The student and teacher surveys provide a qualitative complement to the assessment data. Across the responding cohort, students reported that the diagnostic feedback was the platform's most valued feature: the linkage between an identified gap and a specific resource was experienced as a key benefit. Teachers reported that authoring feedback blocks alongside test items was effortful in the preparation phase but generated reusable assets and surfaced misconceptions in their own course content that had previously been invisible to them.

4. Discussion

4.1 What the Curve Does and Does not Show

As stated above, the improvement-curve evidence is consistent with the formative-feedback theory established by Black et al (1998, 2009) and Hattie et al (2007): when feedback is task-specific, addresses an identified gap, attainment rises; it rises most steeply for students who entered below the competency line. The 17% median gain across 772 students is robust to institutional setting and discipline area, and the threshold-crossing pattern

strengthens the reading: students are not just accumulating attempt-related score increments, but they are passing a criterion-referenced bar.

What the given design does not allow is a controlled comparison against a non-personalised counterfactual. There was no parallel control test receiving conventional formative assessment without diagnostic feedback, and there were no random allocation of students between conditions. Test-wiseness and item-familiarity effects on later attempts are not partialled out, although the use of a shuffled item bank and the platform's *Always latest attempt* setting mitigate the most obvious familiarity artefacts. The claim the paper makes is therefore modest: when a student receives system-generated feedback tied to identified gaps and re-engages with the prescribed remediation material, gains are large, consistent and replicable across six institutions, ten competence areas and all tested students. The paper does not claim that personalised learning outperformed traditional delivery in a counterfactual sense. It claims that the feedback loop produces measurable learning under the tested conditions.

4.2 Transferable Object

The transferable object of the SMART-PL project is the four-stage design pattern. The benchmarking-to-guidelines-to-metric-to-pilot sequence is repeatable in other higher education systems facing analogous pressures: an Eastern Partnership country negotiating EHEA alignment, a candidate country seeking to embed student-centred learning into accreditation procedures, a national system absorbing displaced or returning learners. The platform itself is open-source Moodle with documented configuration and a sharing policy that permits mirror instances; this lowers the transfer cost relative to a proprietary adaptive-learning vendor.

4.3 Limitations

Three limitations deserve to be stated in the paper. First, the pilot is single-domain: all ten ESCO areas are within information and communications technology and generalisation to other disciplinary clusters cannot be assumed without further evidence. Disciplines with less codifiable knowledge structures may respond differently to the closed-loop diagnostic-feedback mechanism that the SMART platform implements. Second, survey self-selection is a known confound on the 42–71% response rates: respondents may differ systematically from non-respondents on engagement with the platform, which is precisely the variable being measured. Third, the wartime context that makes the resilience finding interesting also constrains transferability: the inclusivity infrastructure that Ukrainian partners brought to the project predates SMART-PL and reflects national policy priorities that are not equally salient in other Eastern Partnership systems (Tsybuliak et al, 2024).

5. Conclusion

SMART-PL set out to design, build and empirically evaluate a personalised learning model fit for Ukrainian higher education in its current circumstances. The four-stage design – benchmarking, guidelines, metric framework, platform-mediated pilot – produced an artefact that survived the regulatory feasibility test imposed by Ukrainian programme accreditation procedures and an empirical evaluation that engaged 772 students and 64 academic staff across six universities and ten ESCO-aligned ICT qualification fields. The Improvement Curve evidence – a median gain of 17% between first and second attempts, with sustained upward trajectories across third, fourth and fifth attempts – is fully consistent with the formative-feedback theory established Hattie at al (2007).

The transferable contribution is the design pattern successfully developed by the project team. Other transitional higher education systems facing EHEA-alignment can adopt the four-stage sequence without recourse to proprietary adaptive-learning infrastructure: the SMART platform is open-source Moodle, the sharing policy permits mirror instances and the institutional anchoring achieved through rectorial orders embedding the Co-working Centres into permanent university structures is reproducible administrative practice. The methodological partnerships with NAQA and IHE NAESU that route consortium outputs into Ukraine's national quality-assurance architecture are likewise replicable through equivalent national bodies in other Eastern Partnership.

A causal-design follow-up pilot, with a parallel non-personalised comparator and randomised allocation, is required to extend the within-subject improvement evidence into a controlled effect estimate. Extension beyond ICT subject areas is needed before the model can be claimed as broadly transferable. And the second-iteration testing cycle scheduled for the spring 2026 semester will allow the consortium to test whether the

Spring 2025 effect is robust to changes in cohort composition, including a likely increase in the proportion of returning veterans and internally displaced students relative to the first cohort.

Ethics Declaration

The pilot collected platform-generated assessment data (test attempts, item-level responses, scores, feedback delivery records) and survey responses from students and academic staff as part of regular course activity within the SMART-PL Erasmus+ CBHE project (Grant Agreement 101082928). Data collection followed each partner university's internal ethics and data-protection procedures and the consortium's sharing policy on access and authorisation. No personally identifiable information is reported in this paper or any project deliverables. The authors confirm that formal additional ethics clearance beyond routine pedagogical evaluation procedures was not required under the partner universities' research-ethics frameworks.

AI Declaration

The authors used an AI tools in two roles during the preparation of this manuscript: for language proof-reading of the authors' drafted text and for the formatting of bibliographic references into the Harvard style, including the resolution of DOI identifiers.

References

Alpert, L. (2017) 'The promise (and pitfalls) of personalized learning', *Phi Delta Kappan*, 98(3), pp. 8–13. doi:10.1177/0031721716677260.

Black, P. and Wiliam, D. (1998) 'Inside the Black Box: Raising standards through classroom assessment', *Phi Delta Kappan*, 80(2), pp. 139–148.

Black, P. and Wiliam, D. (2009) 'Developing the theory of formative assessment', *Educational Assessment, Evaluation and Accountability*, 21(1), pp. 5–31. doi:10.1007/s11092-008-9068-5.

Bloom, B.S. (1984) 'The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring', *Educational Researcher*, 13(6), pp. 4–16. doi:10.3102/0013189X013006004.

Boyarchuk, A. and Lysenko, D. (2025) *Adaptation of pilot disciplines for learning in a personalised approach and using the SMART online platform*. SMART-PL Deliverable D.5.2. Available at: https://smart-pl-project.khmnu.edu.ua/ (Accessed: 18 May 2026).

Hattie, J. (2009) *Visible Learning: A synthesis of 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research*, 77(1), pp. 81–112. doi:10.3102/003465430298487.

Kalashnikova, S. and Orzhel, O. (2022) 'Optimisation of the network of higher education institutions: Theoretical features and practical recommendations', *International Scientific Journal of Universities and Leadership*, 13, pp. 89–129. doi:10.31874/2520-6702-2022-13-89-129.

Kalashnikova, S., Vlasova, I., Palamarchuk, O., Debych, M. and Lytovchenko, O. (2023) *Benchmarking of the European and Ukrainian Practice of Applying a Personalised Approach to Learning*. SMART-PL Deliverable D.2.1. Kyiv: Institute of Higher Education, National Academy of Educational Sciences of Ukraine. Available at: https://ihed.org.ua/en/internationalization/project-smart-pl/ (Accessed: 18 May 2026).

Kremen, V.H. (2021) *National Report on the State and Prospects of Education Development in Ukraine: A monograph*. Kyiv: National Academy of Educational Sciences of Ukraine, Konvi Print. Available at: https://lib.iitta.gov.ua/730110/1/NAES-2021-en.pdf (Accessed: 18 May 2026).

Londar, L. and Pietsch, M. (2023) 'Providing distance education during the war: the experience of Ukraine', *Information Technologies and Learning Tools*, 98(6), pp. 31–51. doi:10.33407/itlt.v98i6.5408.

Lugovyi, V., Slyusarenko, O. and Talanova, Z. (2023) 'Transformation of higher education in Ukraine: Impact of the war and objectives for post-war recovery', *European Journal of Education*, 58(4), pp. 575–588. doi:10.1111/ejed.12584.

Lysenko, S. and Hnatchuk, E. (2024) *SMART-PL online platform*. SMART-PL Deliverable D.3.1. Available at: https://smart-pl-project.khmnu.edu.ua/ (Accessed: 18 May 2026).

Means, B. and Neisler, J. (2018) *Personalized Learning: What it really is and why it really matters*. Palo Alto, CA: Learning Policy Institute.

Morkun, N., Kalashnikova, S. and Loban, O. (2024) *Policies on Personalised Learning*. SMART-PL Deliverable D.2.3. Available at: https://smart-pl-project.khmnu.edu.ua/ (Accessed: 18 May 2026).

Nulty, D.D. (2008) 'The adequacy of response rates to online and paper surveys: what can be done?', *Assessment and Evaluation in Higher Education*, 33(3), pp. 301–314. doi:10.1080/02602930701293231.

Pane, J.F., Steiner, E.D., Baird, M.D. and Hamilton, L.S. (2017) *Informing Progress: Insights on personalized learning implementation and effects*. Santa Monica, CA: RAND Corporation. doi:10.7249/RR2042.

Rose, T. (2016) *The End of Average: How we succeed in a world that values sameness*. New York: HarperCollins.

Sadler, D.R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science*, 18(2), pp. 119–144. doi:10.1007/BF00117714.

Tsybuliak, N., Mytsyk, H., Suchikova, Y., Lopatina, H., Kovachov, S. and Popova, A. (2024) 'Inclusion in Ukrainian universities from an inside perspective', *Scientific Reports*, 14(1), 18041. doi:10.1038/s41598-024-69084-1.

VanLehn, K. (2011) 'The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems', *Educational Psychologist*, 46(4), pp. 197–221. doi:10.1080/00461520.2011.611369.

Integrating AI-Generated 3D/360° Content into VR-Based Secondary Education

Vasile-Ioan-Grigore Avram, Ovidiu Cosma and Nadia Barkoczi
Technical University of Cluj-Napoca, North University Center of Baia Mare, Romania

vasile.avram@campus.utcluj.ro

Abstract: This paper presents a pedagogical framework for integrating AI-generated educational content into Virtual Reality (VR)-based teaching within SmartLab-supported secondary education. The framework addresses the limited classroom adoption of AI and immersive technologies by combining AI-assisted content generation, VR integration, interdisciplinary SmartLab activities, and evaluation of learning outcomes into a structured teacher-centred workflow. The framework was implemented during the second semester of the 2025–2026 academic year in a Romanian technological high school, involving 702 secondary school students and 35 master's students who collaboratively designed and delivered 11 interdisciplinary SmartLab lessons. The evaluation combined classroom observations, Kahoot assessments, discussions, and an anonymous questionnaire completed by 550 students. The results indicate increased student engagement, motivation, interdisciplinary understanding, and positive perceptions of AI-assisted immersive learning. The study demonstrates that integrating AI-generated educational resources with VR technologies enhances classroom interaction while supporting the preparation of future teachers through interdisciplinary practice. The proposed framework provides a scalable model for the digital transformation of secondary education and can be adapted to different educational contexts.

Keywords: Artificial Intelligence, Virtual Reality, SmartLab, Secondary Education, Interdisciplinary Learning, Teacher Training.

1. Introduction

The rapid development of digital technologies has transformed contemporary education, with Artificial Intelligence (AI) and Virtual Reality (VR) emerging as key components of Education 4.0. These technologies support immersive, interactive, and student-centred learning while fostering digital competencies required in modern educational environments (Crompton and Burke, 2023; Tlili et al., 2023; European Commission, 2021). Virtual Reality enables students to explore complex concepts through immersive three-dimensional environments, improving engagement, motivation, conceptual understanding, and knowledge retention. At the same time, generative AI facilitates the creation of educational resources such as images, videos, 3D models, and 360° learning environments, reducing the effort required for developing immersive learning materials (Radianti et al., 2020; Hamilton et al., 2021; Lee and Liu, 2025).

Despite these advances, the integration of AI and VR into secondary education remains limited due to challenges related to infrastructure, teacher preparation, curriculum alignment, and the lack of structured pedagogical models. SmartLab environments provide suitable technological infrastructure, but their educational value depends on effective pedagogical integration rather than technology alone. This paper proposes a pedagogical framework for integrating AI-generated educational resources into VR-supported interdisciplinary teaching within SmartLab environments. The framework combines AI-assisted content generation, immersive VR learning, interdisciplinary instructional design, and teacher-guided implementation into a structured workflow suitable for secondary education.

The framework was implemented during the second semester of the 2025–2026 academic year in a Romanian technological high school involving 702 secondary school students and 35 master's students, who collaboratively designed and delivered 11 interdisciplinary SmartLab lessons. The framework was evaluated through classroom observations, Kahoot assessments, classroom discussions, and an anonymous questionnaire completed by 550 students. The main contributions of this study are: (1) the development of a scalable pedagogical framework integrating AI, VR, and SmartLab technologies; (2) its validation through large-scale classroom implementation; (3) the introduction of a collaborative model supporting the preparation of future teachers; and (4) empirical evidence regarding students' perceptions of AI-assisted immersive learning, contributing to research on digital transformation in secondary education (European Commission, 2021; Crompton and Burke, 2023; Tlili et al., 2023).

2. Literature Review / Related Work

2.1 Virtual Reality in Education

Virtual Reality (VR) has become an important educational technology due to its ability to create immersive and interactive learning experiences. Unlike traditional teaching methods, VR enables students to explore simulated environments that enhance engagement, motivation, conceptual understanding, and knowledge retention, particularly in subjects involving abstract or complex concepts (Radianti et al., 2020; Hamilton et al., 2021).

VR has been successfully applied across disciplines such as science, engineering, medicine, geography, and vocational education, supporting experiential learning through visualization, exploration, and active participation (Dede, 2009; Pirker and Dengel, 2021).

Despite these benefits, the adoption of VR in schools remains limited by infrastructure costs, teacher preparation, and the lack of structured pedagogical models for curriculum integration (Song et al., 2024; Liu et al., 2025). Moreover, relatively few studies report large-scale classroom implementations in secondary education or address interdisciplinary teaching and teacher training within authentic school environments (Lee and Liu, 2025; Zagami, 2025).

2.2 Artificial Intelligence and AI-Generated Educational Content

Artificial Intelligence (AI), particularly generative AI, has become an important educational technology by enabling the rapid creation of images, videos, simulations, 3D models, and other learning resources. These tools support personalized and interactive learning while reducing the time required for educational content development (Crompton and Burke, 2023; Tlili et al., 2023). Recent advances in generative AI have expanded its application to adaptive learning environments, intelligent tutoring systems, and immersive educational resources, supporting student-centred learning and improving the accessibility of digital education (Brown et al., 2020; OpenAI, 2023; Song et al., 2024). However, the adoption of AI in education remains challenged by concerns regarding content quality, ethical issues, teacher preparedness, and pedagogical integration. Moreover, most existing studies investigate AI or VR independently, with relatively few proposing structured pedagogical frameworks that combine AI-generated immersive content, Virtual Reality, interdisciplinary learning, and teacher preparation within authentic classroom settings (Tzirides et al., 2024; Liu et al., 2025).

2.3 Smart Learning Environments and SmartLabs

Therefore, there is a growing need for scalable pedagogical frameworks that combine Artificial Intelligence, Virtual Reality, SmartLab infrastructure, and interdisciplinary instructional design while remaining accessible to practicing teachers. (Gros, 2016). Furthermore, there is limited evidence regarding the implementation of such frameworks at a large educational scale or their role in preparing future teachers for technology-enhanced instruction. This study addresses these research gaps by proposing and validating a teacher-centred pedagogical framework implemented in a real SmartLab environment involving 702 secondary education students and 35 master's students from multiple academic disciplines. (European Commission, 2021).

Table 1: SmartLab Technologies and Educational Activities Implemented During the Study

Technology	Educational Use in This Study	Pedagogical Purpose
ClassVR Premium	Immersive exploration of AI-generated educational content	Experiential and interdisciplinary learning
Educational Robotics	Programming and automation activities	Computational thinking
Artificial Intelligence	Creation of images, 3D models, 360° environments and educational videos	Personalized educational content
Interactive Whiteboard	Classroom collaboration	Teacher-guided instruction
Arduino Platforms	STEM activities	Problem-solving
3D Printer & Scanner	Engineering and modelling	Practical digital competencies

3. Proposed Pedagogical Framework

The proposed pedagogical framework supports the integration of Artificial Intelligence (AI), Virtual Reality (VR), and SmartLab technologies into interdisciplinary secondary education. Unlike approaches focusing on individual technologies, it combines AI-assisted content generation, immersive learning, classroom implementation, and learning evaluation within a scalable teacher-centred workflow. The framework promotes interdisciplinary collaboration and experiential learning while maintaining alignment with curricular objectives (Brown et al., 2020; Song et al., 2024). A distinctive feature of the framework is the involvement of master's students in designing and delivering interdisciplinary SmartLab activities under pedagogical supervision. This collaborative approach supports both secondary school students and the professional development of future teachers. The framework consists of four stages: (1) AI-assisted educational content generation, (2) transformation into VR-compatible learning environments, (3) SmartLab-based classroom implementation, and (4) evaluation of learning outcomes and student engagement (Figure 1).

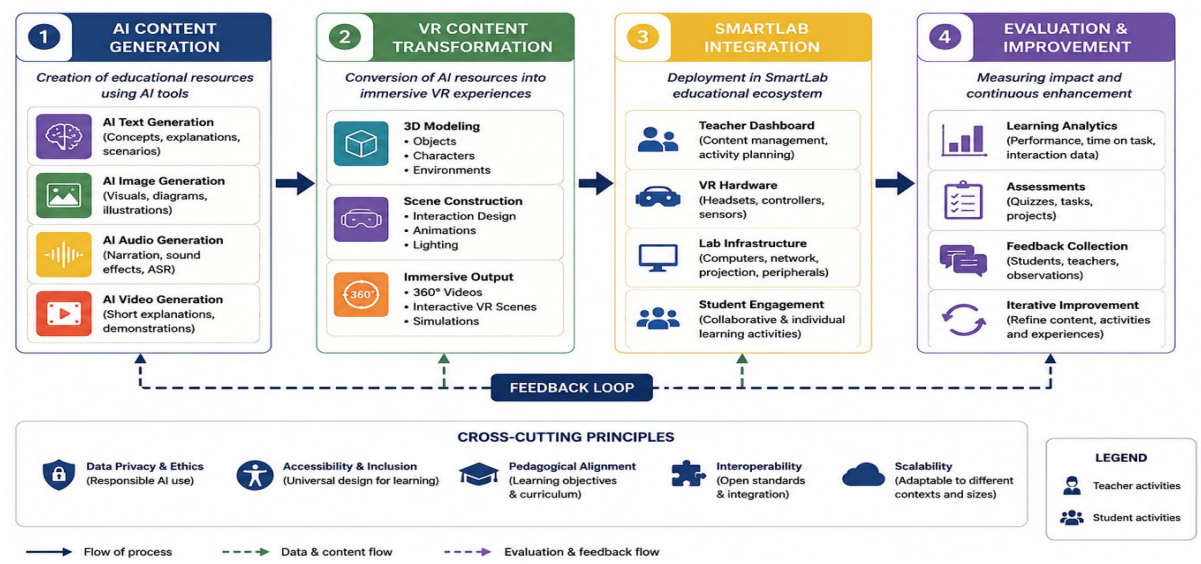


Figure 1: Framework Architecture for AI-Generated VR Learning Environments in SmartLab Ecosystems

3.1 AI-Based Educational Content Generation

The first stage of the framework focuses on generating educational resources using Artificial Intelligence. Generative AI enables the rapid creation of images, 3D models, simulations, videos, and other digital materials that can be adapted to different learning objectives and transformed into immersive VR experiences (OpenAI, 2023; Song et al., 2024). Within the proposed framework, AI-assisted resources support the development of interdisciplinary learning materials while reducing the time and effort required for content creation. Teachers remain responsible for selecting, adapting, and validating all AI-generated materials to ensure their pedagogical relevance and alignment with curriculum objectives. During the implementation, master's students selected AI tools according to the objectives of each lesson and produced images, 3D models, 360° panoramas, educational videos, and other digital resources, which were reviewed by the teaching staff before classroom use (Tlili et al., 2023).

3.2 Transformation into Immersive VR-Compatible Environments

The second stage transforms AI-generated educational resources into immersive VR-compatible environments, including interactive 3D models and 360° learning spaces that can be explored using VR technologies (Radianti et al., 2020; Lee and Liu, 2025). These immersive environments support experiential learning by enabling students to explore, observe, and interact with educational content, improving engagement and conceptual understanding compared with traditional instructional materials (Hamilton et al., 2021). The framework includes virtual laboratories, interactive simulations, and 360° educational scenarios compatible with SmartLab infrastructure and ClassVR devices (Pirker and Dengel, 2021). During implementation, students explored virtual environments, identified objects, observed complex processes, interacted with multimedia resources, and

completed inquiry-based activities using the ClassVR interactive joystick. The emphasis remained on achieving learning objectives rather than on technological novelty.

3.3 Integration into Classroom Activities Using SmartLab Infrastructure

The third stage integrates immersive VR activities into classroom teaching using SmartLab infrastructure. Lessons followed a common 50-minute structure consisting of an introduction, presentation of theoretical concepts, 15–30 minutes of VR exploration, collaborative discussion, and formative assessment. The SmartLab laboratory included 16 ClassVR Premium headsets, allowing students to participate in successive groups or collaborative pairs while maintaining continuous engagement. Teachers integrated VR activities through structured pedagogical scenarios aligned with curriculum objectives. Immersive technologies complemented traditional instruction by supporting guided exploration, collaboration, and problem-solving under continuous teacher supervision (Gros, 2016; European Commission, 2021). Beyond VR, the SmartLab environment incorporated interactive whiteboards, educational robotics, Arduino kits, a 3D printer, a 3D scanner, multimedia systems, and 30 educational laptops. These technologies supported interdisciplinary STEM activities and enriched authentic classroom learning experiences (Radianti et al., 2020).

3.4 Evaluation of Learning Outcomes and Student Engagement

The proposed framework was evaluated using a mixed-method approach combining systematic classroom observation, Kahoot assessments, classroom discussions, and an anonymous student questionnaire. This strategy provided quantitative and qualitative evidence regarding student engagement, motivation, collaboration, and learning outcomes. The implementation involved 702 secondary school students, of whom 550 completed the post-activity questionnaire. The collected data enabled the evaluation of student engagement, interdisciplinary learning, digital competence, and overall satisfaction with the proposed framework. The results are presented in the following section. The evaluation also identified challenges related to technical limitations, teacher preparation, infrastructure, and classroom management, highlighting the importance of scalable and pedagogically sustainable approaches for integrating AI and VR into secondary education.

Table 2: Overview of the Proposed Pedagogical Framework

Framework Stage	Main Activities	Expected Educational Outcomes
AI Content Generation	Images, videos, 3D models, 360° panoramas	Personalized educational resources
VR Transformation	Immersive learning scenarios	Active exploration and visualization
SmartLab Implementation	Interdisciplinary classroom activities	Collaboration and experiential learning
Learning Evaluation	Observation, Kahoot, discussions, questionnaire	Engagement and learning assessment

Overall, the proposed pedagogical framework provides a scalable model for integrating AI-generated educational resources, Virtual Reality, and SmartLab technologies into secondary education while supporting continuous pedagogical improvement.

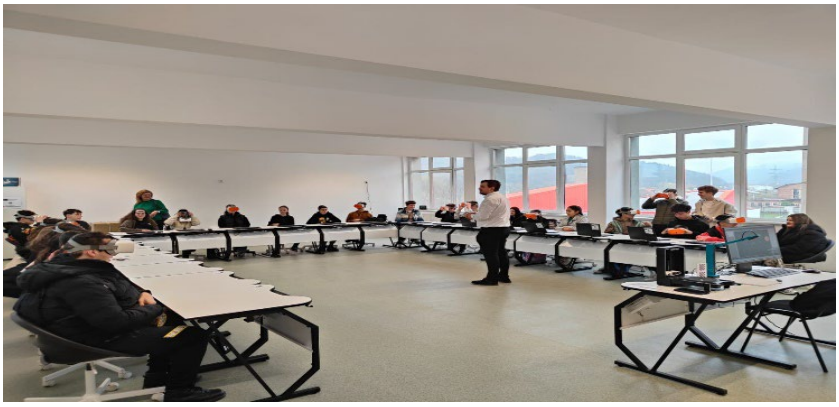


Figure 2: Overview of the SmartLab educational ecosystem integrating VR technologies and collaborative interdisciplinary activities.

4. Case Study

4.1 Educational Context

The proposed pedagogical framework was implemented during the second semester of the 2025–2026 academic year (March–June 2026) in a Romanian technological high school equipped with SmartLab infrastructure. The study was conducted during regular classroom activities to evaluate the practical integration of Artificial Intelligence (AI), Virtual Reality (VR), and interdisciplinary teaching in secondary education. The implementation involved 702 secondary school students from full-time, vocational, and evening education programmes, with an average class size of approximately 24 students. A total of 35 master's students from the Faculty of Science and Engineering participated as future teachers. Representing disciplines such as Mathematics, Computer Science, Biology, Chemistry, Environmental Protection, Economics, and Computer Engineering, they worked in interdisciplinary teams to design and deliver 11 SmartLab lessons under the supervision of the principal researcher. Before classroom implementation, the master's students completed training covering AI, VR, SmartLab technologies, lesson planning, and hands-on practice with the equipment. Each lesson lasted 50 minutes and followed a common pedagogical structure while allowing flexibility according to the educational topic.

Table 3: Participants involved in the educational implementation

Participant Group	Number of Participants
Full-Time Secondary Education Students (Grades 9–12)	305
Vocational Education Students (Grades 9–11)	202
Evening Education Students (Grades 9–13)	195
Total Secondary Education Students	**702**
Master's Students (Future Teachers)	**35**

4.2 SmartLab Infrastructure and Technological Environment

The educational implementation was conducted in a SmartLab environment designed to support immersive and interdisciplinary learning through Artificial Intelligence (AI), Virtual Reality (VR), educational robotics, and digital technologies. The laboratory was equipped with 16 ClassVR Premium headsets, enabling students to participate in immersive 3D and 360° learning experiences. Because the number of devices was smaller than the average class size, students worked in successive groups or collaborative pairs, ensuring active participation throughout each lesson. Additional equipment included 30 educational laptops, an interactive whiteboard, a 3D printer, a 3D scanner, Arduino kits, mBot2, Fable GO and Fable Explore robots, a digital weather station, and multimedia systems. AI tools were used to create images, 3D models, 360° panoramas, educational videos, and other learning resources integrated into the ClassVR platform. Students engaged in immersive learning through virtual exploration, object identification, multimedia interaction, educational games, and collaborative problem-solving. Minor technical issues, including occasional wireless synchronisation delays and headset sensor activation, were quickly resolved and had no significant impact on classroom activities. Overall, the SmartLab infrastructure provided an effective environment for integrating AI, VR, and interdisciplinary STEM education while promoting active learning, collaboration, and digital competence.

Table 4: SmartLab Infrastructure and Educational Technologies

Equipment	Educational Purpose
16 ClassVR Premium headsets	Immersive VR learning and 360° exploration
Interactive whiteboard	Teacher-guided instruction and collaborative learning
30 educational laptops	Digital content creation and classroom activities
3D printer	Prototyping and engineering education
3D scanner	Digital modelling and object digitisation
Arduino educational kits	Programming and STEM activities
mBot2 educational robots	Robotics and computational thinking

Equipment	Educational Purpose
Fable GO & Fable Explore	Educational robotics and collaborative programming
Digital weather station	Environmental monitoring and STEM investigations
High-performance audio system	Multimedia educational experiences



Figure 3: SmartLab educational environment during immersive VR-based interdisciplinary learning activities.

4.3 Learning Activities and Implementation Process

The educational implementation consisted of 11 interdisciplinary SmartLab lessons integrating Artificial Intelligence (AI), Virtual Reality (VR), and SmartLab technologies into authentic classroom instruction. The lessons were collaboratively designed by interdisciplinary teams of master's students under the supervision of the principal researcher, ensuring curriculum alignment and pedagogical coherence. AI tools were used to create educational images, 3D models, 360° panoramas, videos, and other digital resources integrated into the ClassVR platform. Each lesson followed a common 50-minute structure comprising: (1) introduction of the topic, (2) presentation of theoretical concepts, (3) 15–30 minutes of immersive VR exploration, (4) collaborative discussion, and (5) formative assessment through Kahoot and classroom reflection. Using 16 ClassVR headsets, students participated in successive groups or collaborative pairs, while those not using VR completed complementary learning tasks and discussions. Within the immersive environments, students explored virtual scenarios, identified objects, interacted with multimedia resources, and solved interdisciplinary challenges under continuous teacher guidance. One representative lesson focused on the human digestive system, integrating Biology, Chemistry, Mathematics, and ICT. Students explored immersive 3D visualisations, analysed digestive processes, calculated caloric requirements, and investigated digital applications supporting healthy lifestyles. This activity illustrated how interdisciplinary learning can connect concepts from multiple subjects within authentic educational contexts. Overall, the implementation demonstrated that combining AI, VR, and SmartLab technologies with interdisciplinary instructional design promotes active participation, collaboration, creativity, and meaningful learning in secondary education.

Table 5: Standard Structure of the Interdisciplinary SmartLab Lessons

Lesson Stage	Approximate Duration	Main Educational Activities
Introduction	5 minutes	Presentation of learning objectives and interdisciplinary topic
Theoretical preparation	10 minutes	Explanation of key concepts and classroom discussion
Immersive VR activities	15–30 minutes	Exploration of VR environments, object identification, interactive learning
Collaborative discussion	5–10 minutes	Reflection, knowledge consolidation, teamwork
Formative assessment	5 minutes	Kahoot quizzes and classroom feedback



Figure 4: Students participating in interdisciplinary SmartLab learning activities using ClassVR immersive technologies under teacher guidance.

4.4 Interdisciplinary Educational Scenarios

The interdisciplinary educational scenarios constituted a key component of the proposed framework by integrating concepts from multiple disciplines into coherent learning activities. Designed collaboratively by interdisciplinary teams of master's students under the supervision of the principal researcher, the lessons combined Artificial Intelligence, Virtual Reality, and SmartLab technologies while ensuring curriculum alignment and clear learning objectives. AI-assisted resources, including 3D models, 360° environments, educational videos, and multimedia materials, were integrated into the ClassVR platform to support immersive learning. Students engaged in virtual exploration, collaborative problem-solving, multimedia analysis, and classroom discussions, promoting inquiry-based and student-centred learning. One representative lesson focused on the Human Digestive System, integrating Biology, Chemistry, Mathematics, and ICT. Students explored immersive 3D visualisations, analysed digestive processes, calculated caloric requirements, and investigated digital applications supporting healthy lifestyles, demonstrating how interdisciplinary learning connects scientific concepts with real-world applications. Overall, the educational scenarios demonstrated that combining AI-generated resources with Virtual Reality within a structured pedagogical framework can effectively support collaborative, interdisciplinary, and experiential learning.

Table 6: Interdisciplinary Educational Scenarios Implemented During the Study

No.	Educational Scenario	Integrated Disciplines	Main Learning Objective
1	Human Digestive System	Biology, Chemistry, Mathematics, ICT	Understanding the digestive system, nutrition, caloric requirements, and digital health technologies.
2	Directed Graphs for Financial Fraud Detection	Computer Science, Economics	Applying graph theory to detect tax evasion and money laundering through real-world economic scenarios.
3	Backtracking Applications	Mathematics, Computer Science, Biology	Understanding recursive algorithms and their interdisciplinary applications through immersive learning.
4	Environmental Pollution Analysis	Environmental Protection, ICT	Analysing natural pollution factors and representing environmental data using digital technologies.
5	European Integration and Erasmus+ Budget Analysis	ICT, Economics	Analysing Erasmus+ project budgets using spreadsheet functions and immersive learning resources.
6	Creativity and Digital Entrepreneurship	Psychology, Entrepreneurship, ICT	Developing creativity, entrepreneurial thinking, collaboration, and responsible digital behaviour.
7	Electric Current and the Human Body	Biology, Electrical Engineering, Automation	Understanding electrical current, conductors, and the physiological effects of electrocution through immersive simulations.

No.	Educational Scenario	Integrated Disciplines	Main Learning Objective
8	Educational Robotics and Programming	Computer Science, Robotics	Developing computational thinking, programming skills, and problem-solving abilities using educational robots.
9	Sustainable Smart Cities	Geography, Environmental Science, ICT	Exploring sustainable urban development through immersive digital technologies and SmartLab resources.
10	Construction of the Future	Engineering, Environmental Protection, ICT	Investigating sustainable construction technologies and modern engineering solutions using immersive educational resources.
11	Nervous System and Human Creativity	Biology, Psychology, ICT	Exploring the relationship between brain functions, creativity, and digital technologies through Virtual Reality.

4.5 Evaluation Methods

The proposed pedagogical framework was evaluated using a mixed-method approach combining qualitative and quantitative assessment techniques. The evaluation included systematic classroom observation, formative Kahoot assessments, guided classroom discussions, and an anonymous Google Forms questionnaire, providing a comprehensive perspective on student engagement and learning outcomes. Classroom observations focused on students' participation, collaboration, interaction with immersive educational resources, and overall classroom behaviour. Kahoot assessments were used to evaluate conceptual understanding and provide immediate formative feedback. Following each activity, guided classroom discussions encouraged students to reflect on their VR experiences, interdisciplinary connections, and overall learning process. These discussions complemented the quantitative findings with qualitative insights. The main quantitative instrument was an anonymous questionnaire completed by 550 students (approximately 78% of participants). It included Likert-scale, multiple-choice, and open-ended questions addressing engagement, motivation, usability of VR technologies, interdisciplinary learning, digital competence, and overall satisfaction with the SmartLab experience. The combined evaluation methods provided a comprehensive assessment of the proposed framework. The statistical analysis of the questionnaire responses and the interpretation of the qualitative findings are presented in the following section.

Table 7: Evaluation Methods Used During the Educational Implementation

Evaluation Method	Purpose	Data Collected
Systematic classroom observation	Student behaviour and classroom interaction	Participation, collaboration, attention, engagement
Kahoot formative assessment	Learning outcome evaluation	Conceptual understanding and immediate feedback
Classroom discussion	Reflection and qualitative feedback	Students' perceptions and interdisciplinary understanding
Google Forms questionnaire	Quantitative and qualitative evaluation	Motivation, engagement, usability, satisfaction, digital competence

5. Results and Discussion

5.1 Questionnaire Results

An anonymous Google Forms questionnaire was administered immediately after the SmartLab activities to evaluate students' perceptions of Artificial Intelligence (AI), Virtual Reality (VR), and SmartLab technologies in secondary education. A total of 550 students (approximately 78% of all participants) completed the questionnaire, providing a representative dataset for analysing engagement, motivation, interdisciplinary learning, digital competence, and overall satisfaction with the proposed framework. The questionnaire included Likert-scale, multiple-choice, and open-ended questions. Responses were analysed using descriptive statistics, including frequencies and percentages, while qualitative answers complemented the quantitative findings. The following sections present the statistical results and their educational interpretation.

Table 8: Questionnaire Characteristics

Characteristic	Value
Total participants	702
Questionnaire respondents	550
Response rate	78.3%
Questionnaire platform	Google Forms
Question types	Likert-scale, multiple-choice, open-ended

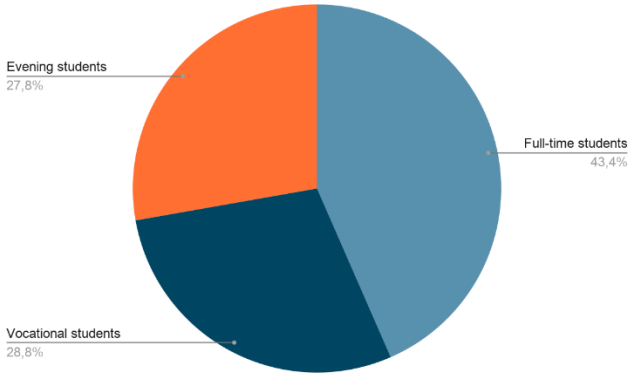


Figure 5: Distribution of questionnaire respondents according to educational pathway.

5.2 Student Engagement and Learning Experience

The questionnaire results indicate that students perceived the interdisciplinary SmartLab activities as highly engaging and educationally valuable, with mean scores ranging from 3.82 to 4.20 on a five-point Likert scale. The highest-rated statement concerned students' willingness to participate in similar activities in the future (M = 4.204, SD = 0.774), with 81.5% of respondents selecting *Agree* or *Strongly Agree*. Similarly, students highly appreciated the educational value of AI-assisted immersive learning (M = 4.173, SD = 0.752; 82.7%) and the interactive SmartLab environment (M = 4.178, SD = 0.745; 82.5%). Virtual Reality also positively influenced classroom participation. Students reported increased attention during VR activities (M = 4.109, SD = 0.768; 79.1%), considered VR-based learning more engaging than traditional instruction (M = 4.160, SD = 0.737; 82.2%), and indicated higher learning motivation (M = 4.029, SD = 0.807; 74.4%). Regarding learning outcomes, students reported improved interest in the studied topics (M = 4.116, SD = 0.778; 79.3%), better understanding of complex concepts (M = 3.942, SD = 0.879; 70.9%), improved information retention (M = 3.842), and stronger interdisciplinary learning (M = 3.818). AI-generated educational resources were also positively evaluated (M = 3.971, SD = 0.789; 73.5%), confirming their contribution to the immersive learning experience. Overall, the findings demonstrate that integrating AI, VR, and SmartLab technologies created an engaging learning environment that promoted motivation, active participation, and interdisciplinary understanding.

Table 9: Main Student Perceptions of the Proposed Pedagogical Framework

Evaluation Dimension	Mean	Agreement (%)
Future implementation of SmartLab activities	4.204	81.5
SmartLab interactivity	4.178	82.5
Educational value	4.173	82.7
VR compared with traditional learning	4.160	82.2
Interest in the learning topic	4.116	79.3
Attention during VR activities	4.109	79.1

Evaluation Dimension	Mean	Agreement (%)
Overall satisfaction	4.042	76.5
Motivation	4.029	74.4
AI-generated educational materials	3.971	73.5
Understanding of concepts	3.942	70.9

5.3 Educational Benefits of AI-Generated Immersive Learning

The questionnaire results, classroom observations, and formative assessments indicate that the proposed pedagogical framework generated educational benefits extending beyond increased student engagement. The integration of Artificial Intelligence (AI), Virtual Reality (VR), and SmartLab technologies supported interdisciplinary learning, conceptual understanding, and the development of digital competencies. One of the main benefits was the promotion of interdisciplinary learning. By combining concepts from Biology, Chemistry, Mathematics, Computer Science, Economics, Environmental Sciences, Robotics, and Engineering, students explored authentic real-world problems and established meaningful connections across multiple disciplines. AI-assisted technologies simplified the creation of immersive educational resources, including 3D models, 360° environments, videos, and multimedia materials, allowing teachers to focus on pedagogical design rather than content production. Virtual Reality further enhanced learning by supporting visualisation, exploration, and experiential learning, contributing to improved attention, conceptual understanding, and classroom participation. The implementation also promoted digital competencies, computational thinking, creativity, collaboration, and problem-solving through the use of VR, educational robotics, AI-generated resources, and SmartLab technologies. An additional benefit concerned teacher preparation. The involvement of 35 master's students in designing and delivering interdisciplinary lessons provided authentic teaching experience and strengthened their ability to integrate emerging technologies into educational practice. Overall, the proposed framework demonstrates that combining AI, VR, and SmartLab technologies provides a scalable model for interdisciplinary, student-centred, and technology-enhanced learning in secondary education.

Table 10: Educational Benefits of the Proposed Pedagogical Framework

Educational Dimension	Observed Educational Benefit
Student engagement	Increased motivation and classroom participation
Interdisciplinary learning	Integration of knowledge across multiple subjects
Conceptual understanding	Improved visualisation of complex concepts
Digital competence	Development of digital literacy and computational thinking
Collaboration	Enhanced teamwork and peer interaction
AI-assisted learning	Rapid creation of personalised immersive educational resources
Teacher training	Practical preparation of future teachers through authentic classroom implementation

5.4 Challenges and Limitations

Although the proposed framework produced positive educational outcomes, several technical, organisational, and methodological challenges were identified. These findings provide practical recommendations for future SmartLab implementations and are consistent with recent studies on AI and Virtual Reality integration in education (Song et al., 2024; Liu et al., 2025). One challenge concerned the management of the 16 ClassVR headsets, which required students to participate in successive groups or collaborative pairs due to the average class size of 24 students. Occasional wireless synchronisation delays and headset sensor activation issues were observed but were quickly resolved by the teaching staff without affecting the learning process. Another limitation was the time required to prepare interdisciplinary lessons, including AI-generated educational resources, lesson planning, testing, and classroom organisation. However, these resources can be reused and adapted for future activities, reducing preparation time. The study was conducted in a single Romanian technological high school, which may limit the generalisability of the findings. Furthermore, the evaluation relied

mainly on classroom observations, formative assessments, discussions, and students' perceptions. Future research should include longitudinal studies, comparative analyses across multiple schools, and objective measures of learning performance. Despite these limitations, the challenges were primarily organisational and technical rather than pedagogical, confirming the feasibility of integrating AI, VR, and SmartLab technologies into interdisciplinary secondary education.

Table 11: Challenges Identified During the Educational Implementation

Challenge	Impact	Implemented Solution
Limited number of VR headsets	Not all students could use VR simultaneously	Student rotation and collaborative pair work
Wireless synchronisation	Occasional connection delays	Immediate reconnection through the ClassVR management platform
VR proximity sensor	Some devices remained in energy-saving mode	Correct adjustment of the headset position
Lesson preparation time	Increased workload for teachers and master's students	Collaborative lesson design and reusable educational resources
Generalisability	Implementation conducted in one educational institution	Future multi-school validation studies

5.5 Discussion

The findings demonstrate that the proposed pedagogical framework provides an effective and scalable approach for integrating Artificial Intelligence (AI), Virtual Reality (VR), and SmartLab technologies into interdisciplinary secondary education. The combination of AI-generated educational resources, immersive learning, and collaborative instructional design promoted student engagement, interdisciplinary understanding, and digital competence (Song et al., 2024; Liu et al., 2025). The questionnaire results, classroom observations, and formative assessments indicate that AI-assisted VR environments effectively complement traditional teaching, supporting previous research while extending it through the validation of a structured pedagogical framework implemented in an authentic SmartLab environment (Radianti et al., 2020; Hamilton et al., 2021). A key contribution of this study is its interdisciplinary approach, demonstrating that immersive technologies can successfully integrate concepts from multiple scientific disciplines within authentic classroom activities (Zagami, 2025). In addition, AI was used as a creative tool for generating immersive educational resources, reducing content preparation time while enabling teachers to focus on pedagogical design (Tzirides et al., 2024). The involvement of 35 master's students in designing and delivering interdisciplinary lessons also highlights the framework's contribution to teacher preparation, showing that future teachers can successfully integrate AI and VR into educational practice when supported by appropriate pedagogical guidance (Lee and Liu, 2025; Tracy and Spantidi, 2025). Overall, the large-scale implementation involving 702 secondary school students and 550 questionnaire respondents demonstrates that the proposed framework is applicable in authentic secondary education settings and represents a promising approach for the digital transformation of education.

6. Conclusion and Future Work

This paper presented a scalable pedagogical framework for integrating Artificial Intelligence-generated educational resources, Virtual Reality, and SmartLab technologies into interdisciplinary secondary education. The framework combines AI-assisted content generation, immersive learning, interdisciplinary instructional design, and authentic classroom implementation within a teacher-centred educational model. The framework was validated through the implementation of 11 interdisciplinary SmartLab lessons involving 702 secondary school students and 35 master's students. The evaluation, based on classroom observations, Kahoot assessments, discussions, and questionnaires completed by 550 students, demonstrated high levels of engagement, motivation, interdisciplinary understanding, and positive perceptions of AI-assisted immersive learning. An additional contribution of this research is the integration of future teacher preparation through the collaborative design and delivery of interdisciplinary SmartLab activities. Although several organisational and technical challenges were identified, they did not prevent the successful implementation of the proposed framework and provide practical recommendations for future applications. Future work will focus on extending the framework to multiple educational institutions, conducting longitudinal studies, integrating adaptive AI techniques, and evaluating the long-term impact of immersive interdisciplinary learning on students' achievement, digital competencies, and teacher professional development.

Ethics Declaration

The study complied with institutional educational regulations and the General Data Protection Regulation (GDPR). Participation in the questionnaire was voluntary and anonymous, and no personally identifiable information was collected.

AI Declaration

Artificial Intelligence tools were used to support language editing, grammar refinement, manuscript organisation, and the generation of illustrative educational resources described in this paper. All research design, data collection, educational implementation, statistical interpretation, scientific analysis, and final conclusions were performed and validated exclusively by the authors, who assume full responsibility for the content of this manuscript.

References

D. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," Computers & Education, vol. 147, 2020.

J. Hamilton, D. McKechnie, E. Edgerton, and C. Wilson, "Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design," Journal of Computers in Education, vol. 8, pp. 1–32, 2021.

S. Mystakidis, "Metaverse," Encyclopedia, vol. 2, no. 1, pp. 486–497, 2022.

H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," International Journal of Educational Technology in Higher Education, vol. 20, no. 22, 2023.

A. Tlili, J. Burgos, R. Huang, and D. Kinshuk, "Artificial intelligence in education: A review of current research and future directions," Computers and Education: Artificial Intelligence, vol. 4, 2023.

UNESCO, Artificial Intelligence and Education: Guidance for Policy-makers. Paris: UNESCO Publishing, 2021.

European Commission, Digital Education Action Plan 2021–2027. Brussels: European Union, 2021.

B. Gros, "The design of smart educational environments," Smart Learning Environments, vol. 3, no. 15, 2016.

C. Dede, "Immersive interfaces for engagement and learning," Science, vol. 323, no. 5910, pp. 66–69, 2009.

P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," IEICE Transactions on Information Systems, vol. E77-D, no. 12, pp. 1321–1329, 1994.

Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

T. Brown et al., "Language models are few-shot learners," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.

OpenAI, GPT-4 Technical Report, 2023.

M. Resnick, Lifelong Kindergarten: Cultivating Creativity through Projects, Passion, Peers, and Play. Cambridge, MA: MIT Press, 2017.

J. Bers, Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom. London: Routledge, 2020.

J. Pirker and C. Dengel, "The potential of virtual reality for immersive learning in education," Proceedings of EDULEARN, 2021.

H. Wu, S. Lee, H. Chang, and J. Liang, "Current status, opportunities and challenges of augmented reality in education," Computers & Education, vol. 62, pp. 41–49, 2013.

M. Akçayır and G. Akçayır, "Advantages and challenges associated with augmented reality for education: A systematic review of the literature," Educational Research Review, vol. 20, pp. 1–11, 2017.

A. O. Tzirides, G. Zapata, N. P. Kastania, A. K. Saini, V. Castro, S. A. Ismael, Y.-L. You, T. A. dos Santos, D. Searsmith, C. O'Brien, B. Cope, and M. Kalantzis, "Combining human and artificial intelligence for enhanced AI literacy in higher education," *Computers and Education Open*, vol. 6, article 100184, 2024.

J. Zagami, "Virtual maker spaces: Enhancing student engagement through extended reality in the co-creation of learning spaces," *Computers & Education: X Reality*, vol. 7, article 100115, 2025.

L.-K. Lee and R. Liu, "Effects of immersive virtual reality on primary students' science performance in classroom settings: A generative AI pedagogical agents-enhanced 5E approach," Interactive Learning Environments, 2025.

K. Tracy and O. Spantidi, "Impact of GPT-Driven Teaching Assistants in VR Learning Environments," IEEE Transactions on Learning Technologies, vol. 18, pp. 192–205, 2025.

Z. Liu, S. Cheon, A. Stanbury, X. Jiao, W. Xing, and H. Kang, "Towards contextual-based AI: A scoping review of artificial intelligence in X reality for personalized learning," Computers and Education: Artificial Intelligence, vol. 10, article 100523, 2025.

Y. Song, K. Wu, and J. Ding, "Developing an immersive game-based learning platform with generative artificial intelligence and virtual reality technologies: 'LearningverseVR'," Computers & Education: X Reality, vol. 4, article 100069, 2024.

Cognitive Barriers in Solving Context-Based Geometry Tasks in an LMS Moodle Environment

Katerina Biernatova[1], Angelika Schmid[1] and Lilla Korenova[2]
[1]University of Ostrava, Faculty of Education, Czechia
[2]Comenius University Bratislava, Faculty of Education, Slovakia

katerina.biernatova@osu.cz
angelika.schmid@osu.cz
korenova@fedu.uniba.sk

Abstract: Context-based tasks provide an important means of connecting geometric knowledge with real-world situations. Successfully solving such tasks, however, requires more than knowledge of geometric concepts and relationships. Students need to understand the given situation, translate it into an appropriate geometric model, identify the relevant geometric relationships, and subsequently interpret the obtained solution within the original context. Each stage of this process may give rise to specific difficulties and cognitive barriers, which can be reflected in incorrect or incomplete solutions. This process can be viewed within the mathematical modelling cycle described by Blum and Leiß (2007). This study focuses on the analysis of open-ended responses produced by pre-service mathematics teachers when solving context-based geometry tasks in the Synthetic Geometry I course. One of the key topics addressed in the course is loci of points satisfying given conditions. As part of the course, students completed an e-test in the Moodle learning management system consisting of seven open-ended context-based tasks. Solving these tasks required students to translate real-world spatial situations into two-dimensional geometric models, identify the corresponding loci of points, and interpret their solutions in relation to the original problem context. Open-ended responses collected during two implementations of the course were analysed using qualitative content analysis with inductively developed categories. Particular attention was paid to incorrect and incomplete solutions in order to identify recurring error patterns and, on this basis, to characterise potential cognitive barriers arising at different stages of the problem-solving process. The analysis focuses particularly on difficulties related to understanding the problem context, constructing an appropriate geometric model, identifying the relevant locus of points, and interpreting the geometric solution within the original situation. The study highlights the potential of analysing incorrect responses as a means of gaining deeper insight into the geometric reasoning of pre-service mathematics teachers. Identifying cognitive barriers may also contribute to the more targeted design of instruction and e-assessment tasks that support not only the acquisition of geometric knowledge but also the development of students' ability to mathematically model and interpret real-world situations.

Keywords: Open-Ended E-assessment, Mathematical Modelling, Geometry Education, Cognitive Barriers, Pre-service Mathematics Teachers.

1. Introduction

Against this background, the present study analyses open-ended responses collected through a Moodle-based e-test in a Synthetic Geometry course for pre-service mathematics teachers. The test consisted of seven context-based tasks concerning loci of points satisfying given conditions. Previous research has addressed digital assessment in mathematics and the formative potential of Moodle-based assessment (Blanco and Ginovart, 2012; Drijvers, 2018; Weigand, Trgalova and Tabach, 2024). However, less attention has been paid to how open-ended responses collected through LMS-based e-assessment can be systematically analysed to distinguish qualitatively different patterns of difficulty in context-based geometry tasks. Rather than focusing solely on achievement, the present study examines incorrect and incomplete responses to identify recurring diagnostic patterns and potential cognitive barriers. The modelling cycle of Blum and Leiß (2007) is subsequently used as an interpretive lens for considering where difficulties may arise when students translate contextual situations into geometric representations. The study therefore contributes to research on e-assessment by demonstrating how open-ended response data collected in Moodle can provide diagnostic information beyond binary scoring, while also offering a geometry-specific framework for interpreting recurring patterns of difficulty.

2. Theoretical Background

Understanding students' difficulties in context-based geometry tasks requires several interconnected perspectives. Digital assessment provides an environment for systematically collecting students' responses, while mathematical modelling offers a framework for understanding the transition between real-world situations and mathematical representations. The analysis of incorrect and incomplete responses can further reveal recurring patterns of difficulty. The following sections therefore address e-assessment in mathematics

and LMS Moodle, mathematical modelling in context-based geometry tasks, and errors and potential cognitive barriers in mathematical problem solving.

2.1 E-Assessment in Mathematics and LMS Moodle

Digital assessment has become an established component of higher education, providing opportunities not only for efficient test administration but also for collecting information about students' learning and performance. Online assessment can support formative processes through feedback and the systematic collection of learner data that may inform subsequent instructional decisions (Heil and Ifenthaler, 2023). In mathematics education, however, digital assessment presents specific challenges because mathematical competence cannot always be adequately represented by selecting a predefined answer or entering only a final numerical result.

Mathematical problem solving may involve multiple representations, strategies, and intermediate steps. Drijvers (2018) therefore emphasises that digital mathematics assessment should provide sufficiently rich opportunities for students to engage in mathematical activity and notes the limitations of assessment focused primarily on final answers and automated scoring. This issue is particularly relevant for open-ended tasks, where an incorrect or incomplete response may contain information about students' reasoning that would remain hidden in a simple correct/incorrect score.

Learning management systems such as Moodle provide an accessible environment for integrating e-assessment into university mathematics courses. Moodle supports different question formats, feedback, repeated testing, and the systematic storage of individual responses. Blanco and Ginovart (2012) demonstrated the potential of Moodle quizzes for formative e-assessment in university mathematics and showed that assessment data can be used not only to evaluate students' performance but also to inform subsequent teaching and assessment.

In the present study, Moodle was used to administer an e-test containing open-ended context-based geometry tasks and to systematically collect students' responses. These responses provide the data for a qualitative analysis of incorrect and incomplete solutions. Thus, Moodle-based e-assessment is considered not only as a means of administering and scoring tasks but also as an environment that enables the collection of response data for examining recurring patterns of difficulty in students' geometric problem solving. Its diagnostic potential therefore lies in preserving open-ended response data for subsequent qualitative analysis rather than in automated scoring itself.

2.2 Context-Based Geometry Tasks and Mathematical Modelling

Context-based mathematical tasks require students to connect mathematical knowledge with situations outside purely mathematical contexts. Their solution involves identifying mathematically relevant aspects of a situation, representing them appropriately, applying mathematical concepts and methods, and interpreting the obtained result in relation to the original context. Mathematical modelling provides a theoretical framework for describing these transitions between reality and mathematics (Blum and Niss, 1991; Kaiser and Sriraman, 2006; Stillman, 2019).

One influential framework is the modelling cycle proposed by Blum and Leiß (2007), which distinguishes between the real and mathematical worlds and conceptualises modelling as a sequence of interconnected processes (Figure 1). Starting from a real situation, the solver constructs a situation model, which is subsequently simplified and structured into a real model. Through mathematisation, this is transformed into a mathematical model, within which mathematical reasoning produces mathematical results. These results are then interpreted as real results and validated against the original situation (Blum and Leiß, 2007).

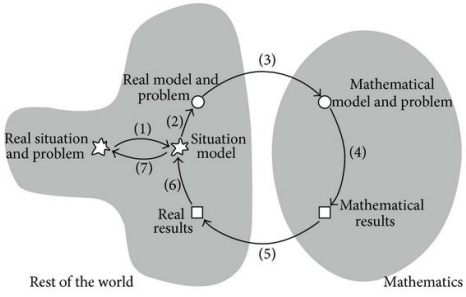


Figure 1: Mathematical modelling cycle according to Blum and Leiß (2007)

The modelling cycle should not be interpreted as a rigid linear sequence. Students may move between stages, return to previous phases, or follow individual modelling routes during problem solving (Borromeo Ferri, 2006; Greefrath and Vorhölter, 2016). Difficulties encountered in modelling tasks therefore cannot automatically be attributed to insufficient mathematical knowledge, as they may arise at different points in the modelling process.

This perspective is particularly relevant to context-based geometry tasks, where students need to translate real or spatial situations into appropriate geometric representations. In tasks involving loci of points satisfying given conditions, conditions expressed in contextual terms must be connected with corresponding geometric properties, and the resulting locus must subsequently be interpreted within the original situation. Research on geometric modelling indicates that mathematisation can be challenging when information derived from real objects or their representations needs to be transformed into mathematical models (Jablonski, 2024).

In the present study, the modelling cycle is used as a theoretical lens rather than a predetermined coding framework. Categories of incorrect and incomplete responses are developed inductively from the data and subsequently considered in relation to transitions within the modelling process. This approach enables recurring response patterns to be examined as possible indicators of difficulties or potential cognitive barriers without assuming that a particular written error directly reflects a specific cognitive process.

2.3 Errors and Cognitive Barriers in Mathematical Problem Solving

Errors can provide valuable information about students' mathematical understanding and reasoning rather than serving solely as indicators of failure. Borasi (1994) described mathematical errors as potential "springboards for inquiry", highlighting their value for examining students' thinking. Error analysis can therefore contribute to identifying recurring areas of difficulty that may inform both research and instructional decisions.

An important distinction must, however, be made between an observable error and its underlying cause. An incorrect response provides evidence that a difficulty occurred but does not necessarily reveal the cognitive process that produced it. Similar errors may originate from different sources, including conceptual misunderstanding, inappropriate strategy selection, difficulties in interpreting the task, or procedural problems. Consequently, recurring patterns across students' responses are more informative than treating individual errors as direct evidence of particular misconceptions or cognitive processes (Radatz, 1980; Borasi, 1994).

This distinction is particularly relevant in geometry, where problem solving involves conceptual knowledge, visualisation, representation, and mathematical reasoning. Research involving pre-service mathematics teachers has identified difficulties related to applying geometric concepts, visualising geometric objects, understanding problem situations, and selecting appropriate principles or definitions (Noto, Priatna and Dahlan, 2019). An incorrect geometric response should therefore not automatically be interpreted as evidence of insufficient geometric knowledge.

In the present study, the term potential cognitive barriers is used cautiously to describe difficulties that may be reflected in recurring patterns of incorrect or incomplete responses. Categories are developed inductively from students' written responses rather than assigned according to presumed cognitive processes. These patterns are subsequently considered in relation to the mathematical modelling framework described above. This approach allows the analysis to explore the diagnostic information contained in students' responses while avoiding direct inference about cognitive processes that cannot be established from final written responses alone. Together, these perspectives suggest a need to examine how open-ended e-assessment responses can provide diagnostic information about qualitatively different difficulties in context-based geometry tasks.

2.4 Research Questions

Building on the theoretical considerations outlined above, the study examines the diagnostic potential of open-ended e-assessment in context-based geometry tasks. Particular attention is paid to the recurring characteristics of students' responses, their possible interpretation through the mathematical modelling cycle, and the diagnostic information that can be obtained beyond binary correct/incorrect scoring. The study addresses the following research questions:

> RQ1: What recurring diagnostic patterns emerge from pre-service mathematics teachers' open responses to context-based geometry tasks in a Moodle-based e-assessment?

RQ2: How can these patterns be interpreted in relation to the stages of the mathematical modelling cycle proposed by Blum and Leiß (2007)?

RQ3: What diagnostic potential do open-ended responses in Moodle-based e-assessment offer beyond the distinction between correct and incorrect answers?

3. Methodology

3.1 Research Context and Participants

The study was conducted within the *Synthetic Geometry I* course for prospective mathematics teachers at the Faculty of Education, University of Ostrava. Loci defined by geometric properties constitute one of the key topics addressed in the course. As part of the course, students completed an e-test administered through the LMS Moodle. The e-test included seven open-ended context-based geometry tasks designed around real-world situations that could be represented and solved using concepts related to geometric loci.

The dataset comprised anonymised open-ended responses from 33 prospective mathematics teachers, collected during two implementations of the course. The first cohort consisted of 16 students and the second of 17 students, yielding a total of 231 responses across the seven tasks. Responses from the two course implementations were analysed as a single dataset, as the purpose of the study was to identify recurring diagnostic patterns rather than to compare the two cohorts. The analysis focused on the information contained in students' written responses rather than solely on whether the responses corresponded to the expected solutions.

3.2 Design of the E-Assessment Tasks

The e-test consisted of seven open-ended tasks focusing on loci defined by geometric properties. Each task presented a real-world context requiring students to translate a spatial situation into plane geometry, identify the locus corresponding to the specified conditions, and interpret the solution in relation to the original context. The tasks therefore reflected the general logic of mathematical modelling described by Blum and Leiß (2007). An overview of the task contexts and expected geometric solutions is provided in Table 1.

Table 1: Overview of the e-assessment tasks and expected geometric solutions

Task	Context	Expected Geometric Solution
T1	Path 10 m from a straight riverbank	Two lines parallel to the riverbank at a distance of 10 m
T2	Area reachable by a goat tied to a stake	Disk centred at K with radius 1 m
T3	Ship remaining 30 km from a lighthouse	Circle centred at the lighthouse with radius 30 km
T4	Cycle path midway between two parallel roads	Line parallel to and equidistant from both roads
T5	Hydrants equidistant from two intersecting streets	Angle bisector of the 60° angle formed by the two streets
T6	Zone equidistant from two sound towers A and B	Perpendicular bisector of segment AB
T7	Watering point equidistant from three barns A, B and C	Circumcentre of triangle ABC (intersection of the perpendicular bisectors of its sides)

The open-ended format required students to formulate their solutions independently rather than select from predefined alternatives. This enabled the responses to preserve evidence of incomplete representations, substitutions between geometric concepts, terminological or formalisation difficulties, procedural reasoning, and internal inconsistencies that could remain obscured by binary correct/incorrect scoring.

3.3 Data Analysis and Development of the Diagnostic Coding Framework

The 231 anonymised responses were analysed using qualitative content analysis with inductive category development. Both authors independently analysed the complete dataset. Observable characteristics of the responses were coded without attributing them to presumed cognitive processes. The independently generated codes were compared, and disagreements or ambiguous cases were discussed until consensus was reached. No

inter-rater agreement statistic was calculated, as the independent analyses were used to support qualitative comparison and consensus development rather than to assess coder reliability statistically. Analytical trustworthiness was supported through independent analysis of the complete dataset by two authors, systematic comparison of the resulting codes, and consensus discussion of disagreements and ambiguous cases. Recurring codes were subsequently grouped into seven higher-order diagnostic patterns. The complete framework is presented in Table 2.

Table 2: Diagnostic coding framework

Higher-Order Pattern	Open Code	Definition and Inclusion/Exclusion Criteria	Representative Example
Successful geometric representation	C1 Explicit correct geometric identification	The expected geometric object or locus is correctly and explicitly identified using appropriate geometric terminology. Includes responses that specify the essential conditions required by the task. Excludes incomplete representations and responses in which the relationship is described correctly but the standard geometric term is not used.	"Circle centred at the lighthouse with radius 30 km."
Successful geometric representation	C2 Correct descriptive identification without standard term	The expected geometric relationship or locus is correctly described, but the conventional geometric term is not explicitly used. Includes responses demonstrating the relevant relationship through an accurate verbal description. Excludes terminologically inaccurate expressions and incomplete representations.	Correctly describes the set of points equidistant from A and B without naming the perpendicular bisector.
Incomplete geometric representation	C3 Incomplete / under-specified locus or object	A relevant component of the expected geometric representation is identified, but the complete locus or all required conditions are not represented. Includes omission of a necessary component, distance, radius, or other defining condition. Excludes responses based on a different geometric concept.	"Rovnoběžka s břehem řeky ve vzdálenosti 10 m." — only one of the two required parallel lines is identified.
Conceptual geometric substitution	C4 Substitution by a different geometric object/locus	The expected geometric object or locus is replaced by another geometric object or locus. Excludes responses that identify the correct object but omit only a defining condition.	A different locus is identified instead of the expected one.
Conceptual geometric substitution	C5 Boundary–region distinction not maintained	The response does not distinguish correctly between the boundary of a geometric set and the entire region satisfying the contextual condition.	*Kružnice* [circle] instead of *kruh* [disk] for the area reachable by the goat.
Conceptual geometric substitution	C6 Substitution among notable triangle points	The required notable point of a triangle is replaced by another notable point. Includes substitutions among geometrically related triangle centres.	*Těžiště* [centroid] instead of the circumcentre.
Terminological/formalisation issue	C7 Terminological imprecision / non-standard formalisation	The relevant geometric relationship is recognisable, but it is expressed using inaccurate, non-standard, or insufficiently formal geometric terminology. Excludes responses in which a genuinely different geometric concept is substituted.	An appropriate geometric relationship expressed using non-standard terminology.
Observable procedural reasoning	C8 Construction-oriented expression of the solution	The response contains an observable construction or procedural route used to obtain the geometric solution rather than merely identifying the resulting locus.	"uděláme kolmici k úsečce AB přesně ve středu" [we construct a perpendicular to segment AB exactly at its midpoint].
Mixed/inconsistent or ancillary error	C9 Internally inconsistent / mixed conceptual response	The response contains mutually incompatible geometric statements or combines an appropriate idea with a contradictory geometric identification.	A response correctly referring to the circumcentre and subsequently

Higher-Order Pattern	Open Code	Definition and Inclusion/Exclusion Criteria	Representative Example
			identifying the same point as the centroid.
Mixed/inconsistent or ancillary error	C10 Ancillary numerical / unit error	The principal geometric representation is appropriate but is accompanied by a secondary numerical, measurement, or unit-related error. Excludes errors that change the underlying geometric concept itself.	Correct locus accompanied by an incorrect numerical value or unit.
Non-interpretable response	C11 No response / insufficiently interpretable response	The response is absent, excessively vague, or contains insufficient information to support a reliable diagnostic interpretation. No specific conceptual difficulty is inferred from such a response.	Blank response or a statement too limited to determine the intended geometric representation.

3.4 Post-Hoc Interpretation Through the Mathematical Modelling Cycle

After the diagnostic coding framework had been developed inductively, the identified response patterns were considered in relation to the mathematical modelling cycle proposed by Blum and Leiß (2007). The modelling cycle therefore served as a post-hoc interpretive lens rather than an a priori coding framework**.**

The analysis considered how the identified patterns could plausibly relate to processes involved in translating a real-world situation into a mathematical representation, working within the mathematical model, and relating mathematical results to the original context. These associations were interpreted cautiously because the dataset consisted of final written responses rather than process-oriented data. Individual responses were therefore not treated as direct evidence of the cognitive stage at which a difficulty occurred.

Finally, the identified patterns were considered from an e-assessment perspective, focusing on the diagnostic information provided by open-ended responses beyond binary correct/incorrect scoring.

4. Results

4.1 Recurring Diagnostic Patterns in Students' Open Responses

The analysis of 231 open responses across the seven context-based geometry tasks identified 11 open codes, which were organised into seven higher-order diagnostic patterns. As individual responses could display more than one analytically relevant characteristic, the diagnostic patterns were not treated as mutually exclusive categories. The overall distribution of the higher-order diagnostic patterns is presented in Table 3.

Table 3: Distribution of Higher-Order Diagnostic Patterns Identified in Students' Open Responses

Diagnostic Pattern	n	% of Responses	Diagnostic Characteristic
Successful geometric representation	126	54.5	Expected geometric relationship identified explicitly or descriptively
Incomplete geometric representation	36	15.6	Relevant representation present but incomplete or under-specified
Conceptual geometric substitution	34	14.7	Another geometric object, locus, region/boundary concept, or notable point substituted
Terminological/formalisation issue	8	3.5	Geometric relationship recognisable but expressed imprecisely
Observable procedural reasoning	2	0.9	Construction or procedural route visible in the response
Mixed/inconsistent or ancillary error	5	2.2	Contradictory concepts or an additional numerical/unit error
Non-interpretable response	44	19.0	Insufficient evidence for reliable diagnostic interpretation

The most frequently identified pattern was successful geometric representation, observed in 126 responses (54.5%). This pattern comprised both explicit identification of the expected geometric concept (117 responses; 50.6%) and correct descriptive identification without the use of the standard geometric term (9 responses;

3.9%). The latter distinction was diagnostically relevant because it indicated that an appropriate geometric relationship could be recognised even when formal terminology was not used.

The second major pattern was incomplete geometric representation, identified in 36 responses (15.6%). In these cases, the response contained a relevant component of the expected geometric model but did not fully represent the locus or all conditions specified in the task. A characteristic example occurred in the riverbank task, in which a student responded: *"Rovnoběžka s břehem řeky ve vzdálenosti 10 m."* [A line parallel to the riverbank at a distance of 10 metres.]. The response identifies the relevant parallel direction and distance but specifies only one of the two parallel lines forming the complete locus. Thus, the response demonstrates partial geometric understanding that would be obscured if it were recorded only as incorrect.

A comparable proportion of responses displayed conceptual geometric substitution (34 responses; 14.7%). This higher-order pattern included substitution of the expected locus by another geometric object (12 responses; 5.2%), failure to maintain the distinction between a boundary and a region (14 responses; 6.1%), and substitution among notable points of a triangle (10 responses; 4.3%). For example, in a task asking for the entire region accessible to a goat tied to a fixed point, some students identified a *kružnice* [circle] rather than a *kruh* [disk]. In another task, *těžiště* [centroid] was given instead of the centre of the circumscribed circle. Such responses did not merely represent an absence of knowledge; they revealed which alternative geometric construct had been activated.

Terminological or formalisation issues were identified in eight responses (3.5%). These responses indicated an appropriate or recognisable geometric relationship but expressed it using imprecise or non-standard terminology. This pattern was therefore distinguished from conceptual substitution: the difficulty concerned the formal expression of the geometric idea rather than necessarily the underlying identification of the relevant relationship.

Less frequent but diagnostically informative patterns were also identified. Observable procedural reasoning occurred in two responses (0.9%), in which students described a construction procedure rather than merely naming the resulting locus. Mixed or internally inconsistent responses and ancillary errors occurred in five responses (2.2%). These included responses combining mutually incompatible geometric concepts or an otherwise appropriate geometric solution with an additional numerical or unit-related error.

Finally, 44 responses (19.0%) were classified as non-interpretable because they were absent or did not contain sufficient information to support a reliable diagnostic interpretation. These responses were retained as a separate pattern rather than being assigned to a presumed conceptual difficulty for which the written data provided no evidence.

Table 4: Distribution of Higher-Order Diagnostic Patterns across the Seven E-assessment Tasks

Task	Successful representation	Incomplete representation	Conceptual substitution	Terminological/ formalisation issue	Procedural reasoning	Mixed/inconsistent or ancillary error	Non-interpretable
T1	2	24	6	0	0	0	1
T2	20	0	13	0	0	3	0
T3	27	10	4	0	0	1	2
T4	22	0	0	8	0	0	10
T5	16	0	0	0	0	0	17
T6	30	1	1	0	2	0	1
T7	9	1	10	0	0	1	13

Note: Diagnostic patterns were not mutually exclusive; therefore, the frequencies within individual tasks may exceed the number of responses (n = 33 per task).

The distribution of diagnostic patterns varied across the seven tasks (Table 4). Incomplete geometric representation was particularly prominent in Task 1 (24/33), whereas conceptual geometric substitution was frequent in Task 2 (13/33) and Task 7 (10/33). Task 6 showed the highest frequency of successful geometric representations (30/33), while Task 7 also generated a comparatively high number of non-interpretable responses (13/33). Terminological or formalisation issues were concentrated in Task 4 (8/33), and observable procedural reasoning occurred only in Task 6 (2/33). These task-specific differences indicate that the diagnostic patterns were associated with the particular geometric distinctions elicited by individual contextual situations rather than representing a uniform type of difficulty across the e-test.

4.2 Interpretation of the Diagnostic Patterns Through Mathematical Modelling

The diagnostic patterns were subsequently interpreted in relation to the mathematical modelling cycle of Blum and Leiß (2007), using the cycle as a post-hoc interpretive framework rather than as a basis for the original coding.

Incomplete geometric representations may be consistent with incomplete mathematisation. For example, identifying one parallel line at the specified distance from a riverbank demonstrates recognition of the relevant relationship while failing to represent the complete locus.

Conceptual geometric substitutions may reflect difficulty in constructing the mathematical model or working mathematically with the concepts involved. For example, distinguishing between a circle and a disk requires determining whether the contextual condition specifies a boundary or an entire region. Similarly, substituting the centroid for the circumcentre indicates difficulty in differentiating between related geometric constructs.

Responses displaying procedural reasoning contained evidence of working mathematically within the geometric model, whereas terminological or formalisation difficulties could not be reliably associated with a specific modelling process.

Overall, the modelling cycle provides a framework for considering processes potentially implicated by the observed patterns. Final written responses, however, do not allow the cognitive processes underlying individual difficulties to be reliably reconstructed.

4.3 Diagnostic Information Beyond Correct/Incorrect Scoring

The analysis demonstrated that open-ended responses provided diagnostic information beyond binary correct/incorrect scoring. Responses receiving the same incorrect outcome differed in their mathematical structure, including incomplete representations, conceptual substitutions, procedural reasoning, terminological difficulties, and internal inconsistencies.

In the Moodle-based e-test, open-ended responses preserved students' individual formulations, allowing these qualitatively different patterns to be distinguished. Their diagnostic value therefore lies not in automatically identifying students' cognitive difficulties, but in providing richer evidence for subsequent qualitative analysis and feedback.

5. Discussion

5.1 From Correctness to Diagnostic Interpretation in Open-Ended E-Assessment

Responses receiving the same incorrect score may reflect qualitatively different forms of mathematical understanding and therefore require different instructional interpretations. An incomplete geometric representation, conceptual substitution, and terminological imprecision represent different diagnostic situations despite potentially receiving the same incorrect score.

The findings therefore support the use of open-ended e-assessment as a source of diagnostic evidence that may inform more differentiated feedback and subsequent instructional decisions.

5.2 Context-Based Geometry Tasks and Mathematical Modelling

The identified patterns suggest that difficulties in context-based geometry tasks cannot be reduced to insufficient knowledge of geometric concepts. Students must translate contextual situations into mathematical representations, creating several points at which difficulties may emerge. This supports previous findings that mathematisation can be particularly demanding when information from real or represented situations must be transformed into an appropriate mathematical model (Jablonski, 2024).

The distinction between incomplete representations and conceptual substitutions further shows that similar incorrect outcomes may reflect qualitatively different difficulties. These patterns should, however, be interpreted as potential indicators of cognitive barriers rather than direct evidence of specific cognitive processes.

5.3 Diagnostic Affordances of Open-Response E-Assessment

Open-ended responses can provide a richer basis for feedback by distinguishing between qualitatively different difficulties that would remain indistinguishable under binary scoring.

Such distinctions may support more targeted feedback. For example, an incomplete representation may require attention to omitted conditions, whereas a conceptual substitution may call for clarification of relationships between geometric concepts. Terminological imprecision may instead require support in mathematical communication. This aligns with the formative potential of digital assessment, in which response data can inform subsequent instructional decisions (Blanco and Ginovart, 2012; Heil and Ifenthaler, 2023).

However, this diagnostic value does not arise automatically from the digital environment. Moodle facilitates the collection and preservation of open-ended responses, but their interpretation requires an appropriate analytical framework and professional judgement. Moodle-based e-assessment can therefore support differentiated feedback when responses are treated not only as objects of scoring but also as diagnostic evidence.

5.4 Implications for Prospective Mathematics Teacher Education

The findings are particularly relevant to pre-service mathematics teacher education. Prospective teachers need not only to solve geometric problems correctly but also to recognise different forms of difficulty in students' solutions. Analysing open-ended responses can highlight distinctions between partial understanding, conceptual substitution, and imprecise mathematical communication that may be obscured by correctness-based assessment.

Such analysis may support the development of prospective teachers' diagnostic competence by encouraging them to consider what mathematical understanding is visible in a response and what instructional support may be appropriate. Open-ended e-assessment can thus serve not only as an assessment tool but also as a resource for interpreting students' mathematical thinking.

5.5 Limitations and Directions for Further Research

The study is limited by its context-specific sample of 33 pre-service mathematics teachers from a single university course and is not intended to provide statistically generalisable conclusions. Moreover, final written responses do not allow the cognitive processes underlying particular errors to be reconstructed; the identified patterns should therefore be interpreted as potential indicators of cognitive barriers rather than direct evidence of specific cognitive processes.

Future research could combine open-ended e-assessment with process-oriented methods, such as think-aloud protocols, interviews, or observations, to investigate how particular response patterns emerge. Applying the diagnostic framework to larger samples, different geometric topics, or other mathematical domains could further examine its transferability and refine the identified categories.

6. Conclusion

This study examined the diagnostic potential of open-ended Moodle-based e-assessment through responses to context-based geometry tasks completed by pre-service mathematics teachers. The findings showed that responses receiving the same broad correct/incorrect judgement may reflect qualitatively different forms of mathematical understanding, including incomplete geometric representations, conceptual substitutions, terminological or formalisation difficulties, and internally inconsistent responses.

The identified diagnostic patterns were subsequently interpreted through the mathematical modelling cycle of Blum and Leiß (2007). This interpretation highlighted possible difficulties in translating contextual situations into appropriate geometric representations while also demonstrating the methodological limits of drawing conclusions about underlying cognitive processes from final written responses. The modelling cycle was therefore used as a post-hoc interpretive lens rather than as a framework for reconstructing students' cognitive processes.

Overall, the study demonstrates that the value of open-ended Moodle-based e-assessment lies in preserving diagnostically rich response data that would be obscured by binary scoring alone. Such responses can provide a basis for more differentiated analysis and feedback, while the diagnostic framework developed in this study

offers a systematic approach to distinguishing qualitatively different characteristics of students' geometric responses.

Acknowledgement

This research was supported by the Student Grant Competition of the University of Ostrava within the project "Integration of Electronic Testing into Constructivist Geometry Education in the Preparation of Pre-Service Mathematics Teachers" (SGS02/PdF/2026).

Ethics Declaration

The data analysed in this study were collected as part of regular teaching and assessment activities in the Synthetic Geometry I course. The dataset consisted exclusively of students' written responses to the Moodle-based e-test. Prior to the research analysis, all responses were anonymised and any personally identifiable information was removed. The research analysis was conducted independently of the students' course assessment and had no impact on their grades or academic outcomes.

AI Declaration

Generative AI tools were used to support language editing, improve the clarity and readability of the manuscript, and assist with the organisation and refinement of the text. The authors retained full responsibility for the research design, data analysis, interpretation of the results, and the final content of the manuscript.

References

Blanco, M. and Ginovart, M. (2012) 'On how Moodle quizzes can contribute to the formative e-assessment of first-year engineering students in mathematics courses', RUSC. Universities and Knowledge Society Journal, 9(1), pp. 354–370. doi: 10.7238/rusc.v9i1.1277.

Blum, W. and Leiß, D. (2007) 'How do students and teachers deal with modelling problems?', in Haines, C., Galbraith, P., Blum, W. and Khan, S. (eds.) Mathematical Modelling (ICTMA 12): Education, Engineering and Economics. Chichester: Horwood Publishing, pp. 222–231. doi: 10.1533/9780857099419.5.221.

Blum, W. and Niss, M. (1991) 'Applied mathematical problem solving, modelling, applications, and links to other subjects – State, trends and issues in mathematics instruction', Educational Studies in Mathematics, 22(1), pp. 37–68. doi: 10.1007/BF00302716.

Borasi, R. (1994) 'Capitalizing on errors as "springboards for inquiry": A teaching experiment', Journal for Research in Mathematics Education, 25(2), pp. 166–208. doi: 10.2307/749507.

Borromeo Ferri, R. (2006) 'Theoretical and empirical differentiations of phases in the modelling process', ZDM – Mathematics Education, 38(2), pp. 86–95. doi: 10.1007/BF02655883.

Drijvers, P. (2018) 'Digital assessment of mathematics: Opportunities, issues and criteria', Mesure et Évaluation en Éducation, 41(1), pp. 41–66. doi: 10.7202/1055896ar.

Greefrath, G. and Vorhölter, K. (2016) *Teaching and Learning Mathematical Modelling: Approaches and Developments from German Speaking Countries*. Cham: Springer. doi: 10.1007/978-3-319-45004-9.

Heil, J. and Ifenthaler, D. (2023) 'Online assessment in higher education: A systematic review', Online Learning, 27(1), pp. 187–218. doi: 10.24059/olj.v27i1.3398.

Jablonski, S. (2024) 'Challenges in geometric modelling – A comparison of students' mathematization with real objects, photos, and 3D models', EURASIA Journal of Mathematics, Science and Technology Education, 20(3), em2414. doi: 10.29333/ejmste/14321.

Kaiser, G. and Sriraman, B. (2006) 'A global survey of international perspectives on modelling in mathematics education', ZDM – Mathematics Education, 38(3), pp. 302–310. doi: 10.1007/BF02652813.

Noto, M.S., Priatna, N. and Dahlan, J.A. (2019) 'Mathematical proof: The learning obstacles of pre-service mathematics teachers on transformation geometry', Journal on Mathematics Education, 10(1), pp. 117–126. doi: 10.22342/jme.10.1.5379.117-126.

Radatz, H. (1980) 'Students' errors in the mathematical learning process: A survey', For the Learning of Mathematics, 1(1), pp. 16–20.

Stillman, G.A. (2019) 'State of the art on modelling in mathematics education – Lines of inquiry', in Stillman, G.A. and Brown, J.P. (eds.) Lines of Inquiry in Mathematical Modelling Research in Education. Cham: Springer, pp. 1–20. doi: 10.1007/978-3-030-14931-4_1.

Weigand, H.-G., Trgalova, J. and Tabach, M. (2024) 'Mathematics teaching, learning, and assessment in the digital age', ZDM – Mathematics Education, 56(4), pp. 525–541. doi: 10.1007/s11858-024-01612-9.

Structured over Chat-Based Feedback: Comparing User Preferences on Different Presentation Feedback Reports

Noemi Bontempo[1], Nina Mouhammad[2] and Daniele Di Mitri[2,3]
[1]Goethe University, Frankfurt am Main, Germany
[2]DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany
[3]German University of Digital Science, Potsdam, Germany

bontemponoemig@gmail.com
n.mouhammad@dipf.de
d.dimitri@dipf.de

Abstract: Digital presentation training systems provide automated feedback during presentation practice, but less attention has been paid to how feedback should be presented after a training session to support review and reflection. This paper investigates how learners perceive different feedback representations for understanding presentation performance and reflecting on it. A Rehearsal Report was designed for Presentable, a digital presentation training system that provides automated feedback on presentation delivery. Three concepts were developed: a timeline-based design that connects feedback to specific moments in the presentation, a chat-based design that presents feedback conversationally, and a storytelling-based design that organizes feedback into thematic categories. The concepts were evaluated in an online within-subject study with 84 participants. Participants reviewed all three designs and evaluated them regarding clarity, usefulness, reflection support, preference, and perceived support for understanding presentation performance. The results show that the chat-based design received lower ratings than both structured alternatives. Participants preferred designs with clear visual organization and contextual information. The timeline-based design was rated most helpful overall, while the storytelling-based design was rated highest for ease of understanding. These findings suggest that structured visual representations may be perceived as more supportive for post-training reflection than conversational feedback formats.

Keywords: Presentation Training, Feedback Visualization, Reflection, Educational Technology, Learning Analytics

1. Introduction

Presentations are an important form of communication in educational and professional contexts. Students present their work in seminars, group projects, or final examinations, while professionals use presentations to communicate ideas, results, or products. Although presentations are common, giving an effective presentation remains challenging for many learners. Successful presenting requires not only knowledge of the topic, but also delivery-related skills such as clear speech, confident posture, appropriate gestures, and audience orientation.

Digital presentation training systems have been developed to support learners in practising these skills. Such systems can provide automated feedback on aspects of presentation delivery, for example speaking pace, posture, gestures, or gaze direction. Previous research shows that automated feedback can increase learners' awareness of their presentation behaviour and support public speaking practice, especially when feedback is provided during training (Chollet, Marsella and Scherer, 2022; Jariyah, Patak and Novia, 2025; Sichterman et al., 2025).

However, immediate feedback alone does not fully support reflection after a presentation has ended. During a presentation, learners must focus on speaking, remembering their content, managing body language, and reacting to the situation. Processing feedback deeply at the same time can therefore be difficult. Post-training feedback offers a different learning opportunity: learners can review their performance as a whole, identify recurring patterns, and reflect on strengths and weaknesses. This relates to self-regulated learning, where feedback supports learners in evaluating their performance and planning future improvements (Nicol and Macfarlane-Dick, 2006; Theobald and Bellhäuser, 2022).

A central challenge is how post-training feedback should be presented. Dashboards, scores, graphs, or isolated feedback messages can provide useful information, but they do not automatically help learners understand what the feedback means for their own behaviour. Research on learning analytics and feedback visualization therefore emphasizes the importance of clear, learner-centred representations that help users interpret performance data (Ahn et al., 2019; McKenna et al., 2019).

This paper investigates how learners perceive different representations of post-training presentation feedback in terms of clarity, usefulness, preference, and perceived support for reflection. For this purpose, three Rehearsal Report designs were developed for Presentable: a timeline-based design, a chat-based design, and a

storytelling-based design. The concepts were evaluated in an online within-subject study with 84 participants. By comparing these approaches, the paper contributes design insights for post-training feedback in digital presentation training systems.

2. Related Work

Digital presentation training systems support learners by providing automated feedback on presentation delivery. Such systems can analyse verbal features, such as speaking pace or filler words, and non-verbal features, such as posture, gestures, gaze direction, or body movement. Previous research suggests that automated presentation feedback can increase learners' awareness of their behaviour and support public speaking practice, especially when human feedback is limited (Chollet, Marsella and Scherer, 2022; Jariyah, Patak and Novia, 2025; Sichterman et al., 2025). Existing presentation training systems often provide feedback during or shortly after presentation practice and differ in the technical features they use to capture, analyse, and present feedback (Hummel et al., 2025). However, comparatively less attention has been given to how learners can review and reflect on their performance after the training session has ended.

Immediate feedback can help presenters notice specific behaviours and adjust them during practice, but processing feedback while presenting may be demanding. Learners must divide their attention between speaking, remembering their content, managing body language, and reacting to the presentation situation. This relates to cognitive load theory, which emphasizes that learning tasks should avoid overloading learners' limited working memory capacity (Sweller, 1988; Sweller, van Merriënboer and Paas, 1998). Post-training feedback addresses a different phase of learning. After the presentation, learners can review their performance more calmly, identify recurring patterns, and plan future improvements. This reflective process is closely related to self-regulated learning, where feedback helps learners evaluate their current performance and close the gap between current and desired outcomes (Nicol and Macfarlane-Dick, 2006; Theobald and Bellhäuser, 2022). Research on formative assessment similarly emphasizes that feedback should not only inform learners about their performance, but also help them interpret the feedback and remain engaged in the learning process (Gikandi, Morrow and Davis, 2011).

However, delayed or post-training feedback also has limitations. Since it is presented after the performance has ended, learners can no longer adjust their behaviour in the moment. Research on feedback timing suggests that immediate and delayed feedback can support learning in different ways, depending on the task, feedback content, and learning goals (Taxipulati et al., 2021). In the context of post-training feedback, the representation therefore needs to support memory, interpretation, and planning rather than immediate correction. If feedback remains too abstract or disconnected from the original presentation situation, learners may struggle to relate it to their own behaviour.

The usefulness of feedback depends not only on its content, but also on how it is presented. Dashboards, charts, scores, or summary views can make performance data accessible, but they do not automatically help learners interpret what the data means for their own behaviour. Learning analytics research therefore emphasizes the need for learner-centred feedback visualizations that support interpretation rather than only displaying data (Ahn et al., 2019; McKenna et al., 2019). Jivet et al. (2017) similarly argue that awareness is not enough: learning analytics dashboards need to support sense-making and meaningful action. Zha and Harrison (2019) also show that feedback visualization design influences how users understand and use visual feedback.

For presentation training, temporal representations are particularly relevant because presentation behaviour unfolds over time. A timeline can help learners reconstruct when specific feedback events occurred and how their behaviour developed during the presentation (Shneiderman, 2003). However, temporal information alone may not be sufficient. Research on informative tutoring feedback suggests that feedback is more useful when it provides explanations and context rather than only indicating that something happened (Narciss and Huth, 2004). Narrative and conversational formats offer alternative ways of presenting feedback. Data storytelling can combine structure, explanation, and visual information to make feedback easier to interpret (Maheshi et al., 2024), while chat-based feedback may feel familiar and supportive because it resembles a conversation with a coach. At the same time, presenting feedback step by step may reduce overview and make it harder to compare feedback points. The present study addresses this design tension by comparing three Rehearsal Report designs: a timeline-based representation, a chat-based representation, and a storytelling-based representation.

3. Rehearsal Report Designs

The Rehearsal Report designs were developed as post-training feedback concepts for Presentable, a digital presentation training system that supports learners during presentation rehearsal. The training process can be understood as consisting of three phases. First, learners prepare their presentation through a short planning or message-composition phase. Second, they rehearse their presentation while receiving automated feedback on observable aspects of delivery, such as posture, gestures, body movement, and orientation towards the camera or audience. Third, after the rehearsal has ended, the Rehearsal Report provides a structured reflection phase in which learners can review and interpret the feedback from the session.

For the purpose of this study, three mockups of the rehearsal dashboard were created. All three mockups were based on the same fictional rehearsal scenario and the same underlying feedback content. In this scenario, a learner gives a short practice presentation, initially shows expressive gestures and distracting movements, briefly loses connection with the audience in the middle, and becomes calmer and more controlled towards the end. Using the same scenario ensured that participants evaluated different versions of the rehearsal dashboard rather than different feedback content.

Figures 1a and 1b show excerpts of the three Rehearsal Report designs compared in the study. Figure 1a presents an excerpt of the timeline-based design, while Figure 1b presents excerpts of the storytelling-based and chat-based designs.

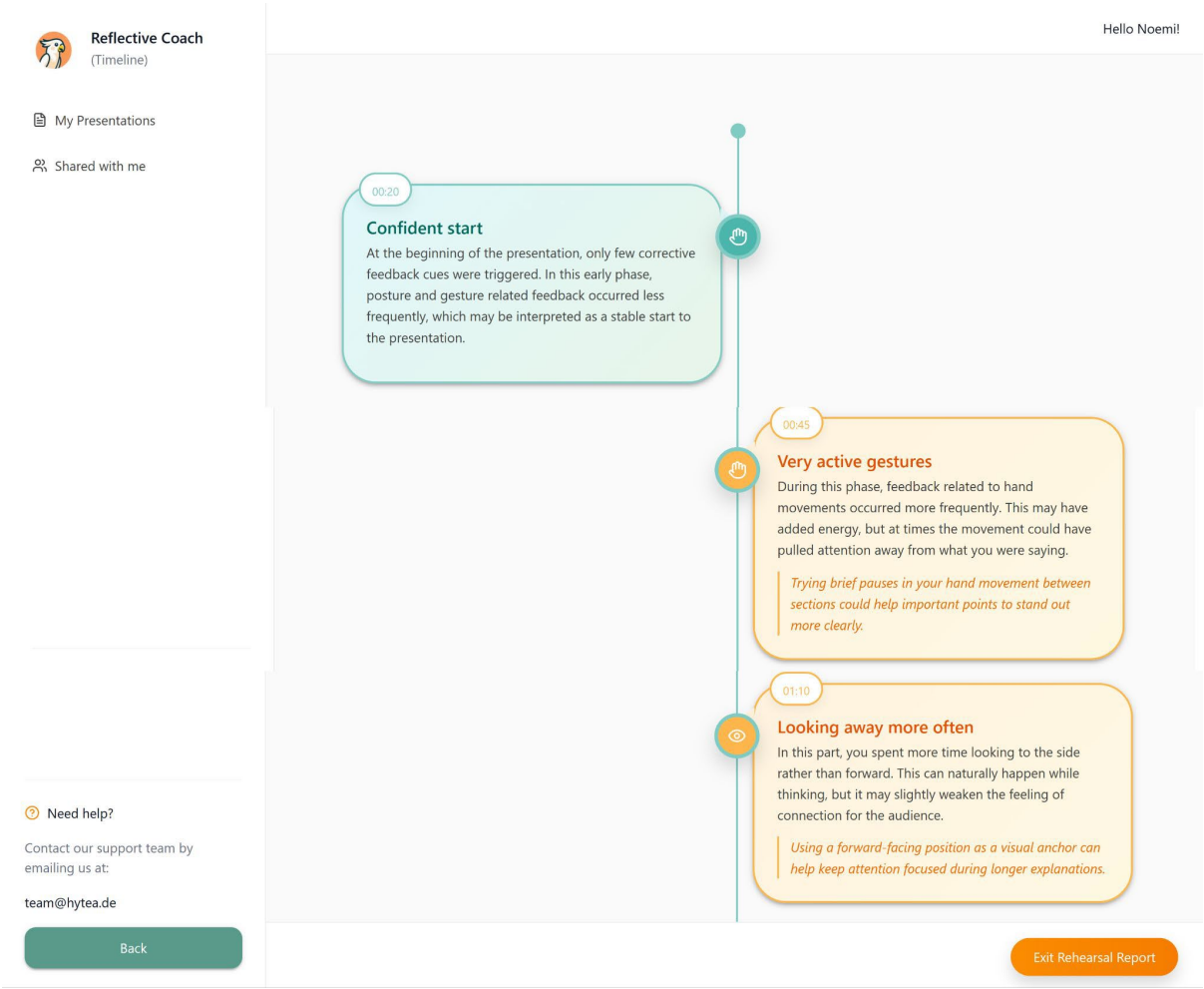


Figure 1a: Excerpt of the timeline-based Rehearsal Report design.

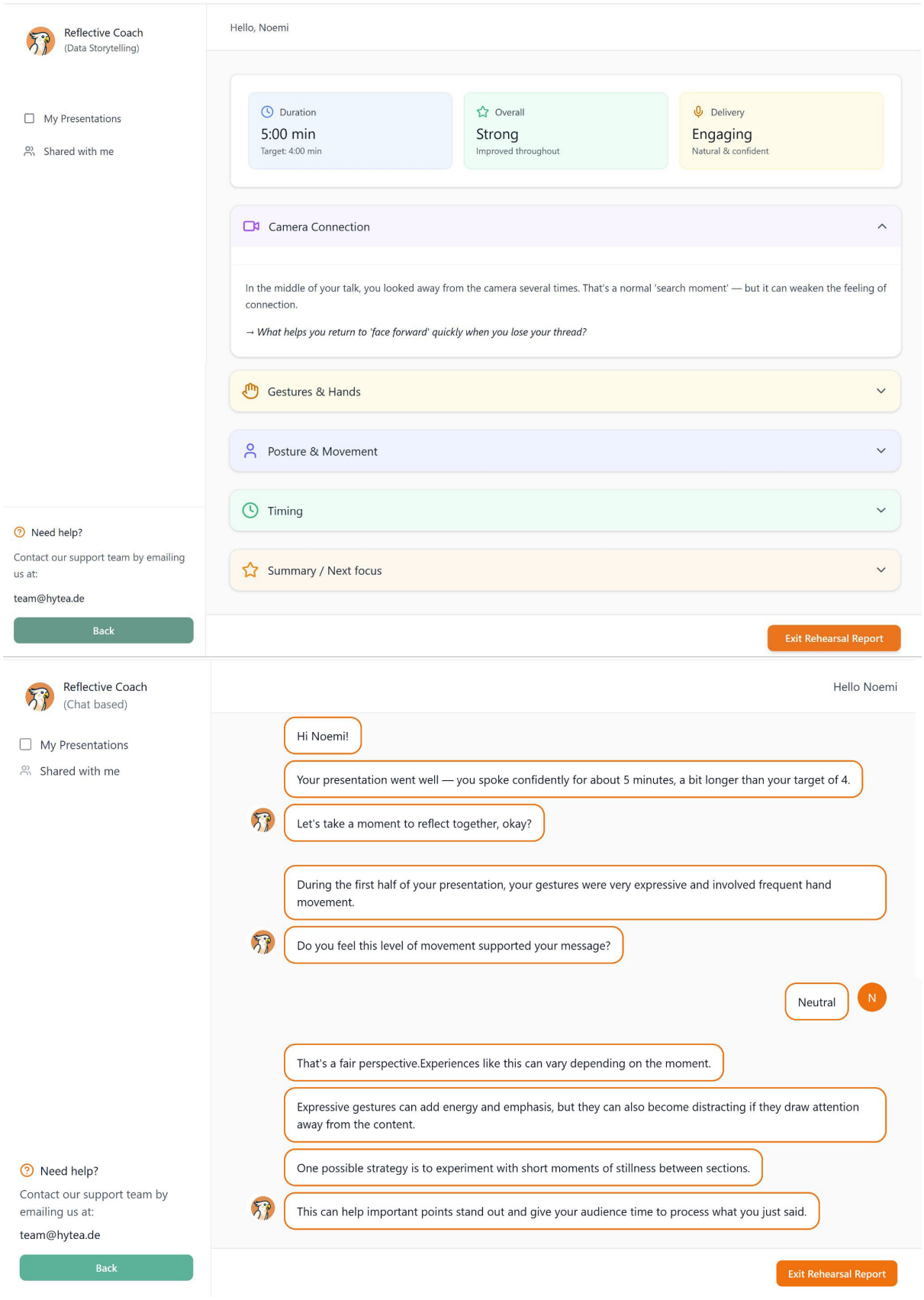


Figure 1b: Excerpts of the storytelling-based and chat-based Rehearsal Report designs.

3.1 Timeline-Based Design

The timeline-based design presents feedback as a sequence of events along the duration of the presentation. Feedback items are connected to timestamps and placed on a vertical timeline, allowing learners to see when specific behaviours occurred and how their presentation developed over time. This design was chosen because the feedback generated during rehearsal is event-based and connected to specific moments. A timeline can therefore preserve this temporal structure and help learners reconstruct the presentation process after the session. This follows visualization principles that emphasize overview, temporal organization, and access to details (Shneiderman, 2003).

Positive feedback and improvement feedback are visually separated, and individual feedback items can be opened to show a short explanation and advice. The design combines an overview of the whole presentation with detailed feedback on specific moments. It aims to help learners identify patterns, such as whether certain behaviours occurred mainly at the beginning, middle, or end of the presentation. This is consistent with feedback visualization research, which highlights the importance of clear representations that help learners interpret performance information rather than only displaying data (Ahn et al., 2019; McKenna et al., 2019).

3.2 Chat-Based Design

The chat-based design presents feedback as a conversation with a virtual coach. Instead of showing all feedback at once, the report guides the learner through short messages and reflective prompts. Learners can respond using predefined options, such as agree, neutral, or disagree. This design was included because conversational interfaces are familiar to many users and may make feedback feel more personal, supportive, and less evaluative.

At the same time, the sequential structure provides less immediate overview because feedback points are not visible all at once. The design therefore prioritizes guidance and approachability over detailed comparison and temporal overview. This contrast is relevant because learning analytics dashboards should not only create awareness, but also support sense-making and meaningful action (Jivet et al., 2017).

3.3 Storytelling-Based Design

The storytelling-based design organizes feedback into thematic categories, such as gestures, posture, camera connection, timing, and overall summary. Each category contains short explanatory text and a reflective question. Instead of focusing on detailed timestamps, the design presents feedback as a structured and readable overview.

This design was chosen to explore whether feedback becomes easier to understand when it is organized as an explanatory narrative rather than as a temporal sequence or conversation. The aim is to reduce the effort required to interpret the feedback by grouping related observations and explaining them in everyday language. This is in line with research on data storytelling, which suggests that combining structure, explanation, and visual information can make feedback analytics easier to interpret (Maheshi et al., 2024). The storytelling-based design therefore focuses on clarity, categorization, and interpretation, but provides less precise temporal information than the timeline-based design.

4. Method

4.1 Study Design and Procedure

The three Rehearsal Report designs were evaluated in an online within-subject study. This design was chosen because the aim was to compare how the same participants perceived three different feedback representations. Each participant reviewed all three designs: the timeline-based design, the chat-based design, and the storytelling-based design.

At the beginning of the survey, participants received a short introduction to the study context. They were asked to imagine that they had completed a practice presentation using a digital presentation training tool and were now receiving a feedback report to support reflection and improvement. The three design concepts were then shown in randomized order to reduce order effects. For each design, participants viewed a screenshot and watched a short demonstration video showing how the report could be used, for example by opening feedback sections or navigating through the interface.

After each design presentation, participants answered the same set of evaluation questions, which are described in Section 4.3. After all three designs had been evaluated individually, participants were shown a thumbnail overview of the designs to support recall. The final part of the survey focused on direct comparison, including overall preference, ranking, motivation to improve, and perceived support for understanding presentation performance.

4.2 Participants

A total of 84 participants completed the online survey. Participants were recruited mainly through online survey exchange platforms and, to a smaller extent, through personal contacts and university-related channels. The only formal requirement for participation was that participants had to be at least 18 years old.

The sample included participants with different educational and academic backgrounds. Regarding the highest completed level of education, 49% reported a Bachelor's degree, 20% a Master's degree, 23% a high school qualification, 5% vocational training, and 4% a doctoral degree. Academic backgrounds included humanities or social sciences (48%), other STEM fields (14%), computer science (7%), education (4%), and other fields (27%).

Most participants had at least some presentation experience. Regarding presentation frequency, 26% reported giving presentations frequently or monthly, 57% occasionally, 8% rarely, and 8% very frequently. Previous use of digital presentation training tools was less common, with 33% reporting previous use and 67% reporting no previous use.

4.3 Measures and Analysis

The survey combined quantitative and qualitative measures. For each design, participants answered seven Likert-scale items on a seven-point scale from strongly disagree to strongly agree. The items were developed specifically for this study and were not adopted from a standardized questionnaire. They focused on perceived usefulness for improving presentation skills, ease of understanding, perceived support for reflection, clarity of structure and layout, support for understanding when specific behaviours occurred, intention to use the report after practice, and motivation to improve the next presentation.

In the comparison part of the survey, participants selected which design they found most helpful, which design made it easiest to understand their performance, and which design would most motivate them to improve. They also ranked the three designs from most helpful to least helpful for reflecting on presentation performance and selected which design elements contributed most to their motivation to reflect. Open-ended questions allowed participants to describe what they liked about each design, what they found confusing or less helpful, and why they preferred a particular design.

The quantitative data were analysed descriptively using frequencies, percentages, means, and standard deviations. The open-ended responses were analysed thematically to identify recurring positive and critical comments for each design.

5. Results

5.1 Individual Design Ratings

As shown in Figure 2, the timeline-based and storytelling-based designs received higher ratings than the chat-based design across the seven evaluation items. The timeline-based design was rated positively, with mean values ranging from 5.20 to 5.86 on the seven-point scale. Its highest rating concerned helping participants understand when specific behaviours occurred during the presentation (M = 5.86). The storytelling-based design received similarly positive ratings, with mean values ranging from 5.24 to 5.87, and was rated highest for ease of understanding (M = 5.87). In contrast, the chat-based design received lower ratings, with mean values ranging from 3.93 to 4.93. Its lowest rating concerned participants' intention to use this type of report after presentation practice (M = 3.93).

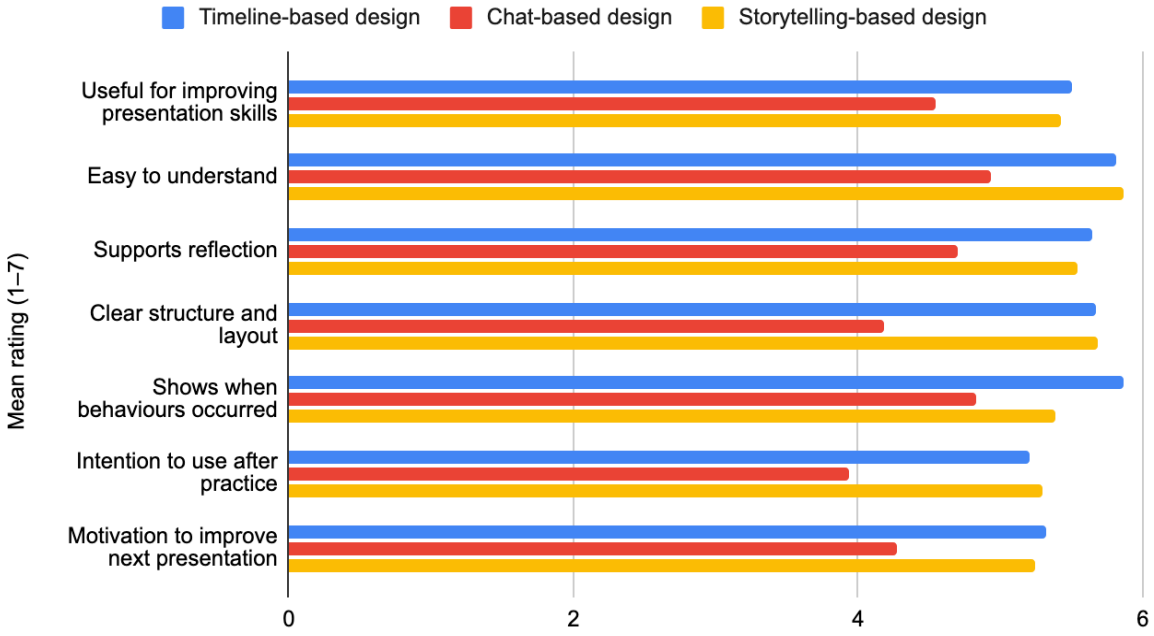


Figure 2: Mean ratings of the three Rehearsal Report designs.

5.2 Design Preference and Comparison

As shown in Figure 3, the timeline-based design was selected as most helpful overall by 48% of participants, followed by the storytelling-based design with 40% and the chat-based design with 12%. For understanding presentation performance, the storytelling-based design was selected most often (49%), followed by the timeline-based design (40%) and the chat-based design (11%). For motivation to improve the next presentation, the timeline-based design was selected by 42% of participants, the storytelling-based design by 40%, and the chat-based design by 18%. The ranking results confirmed this tendency: the timeline-based design was most often ranked first, while the chat-based design was most frequently ranked last.

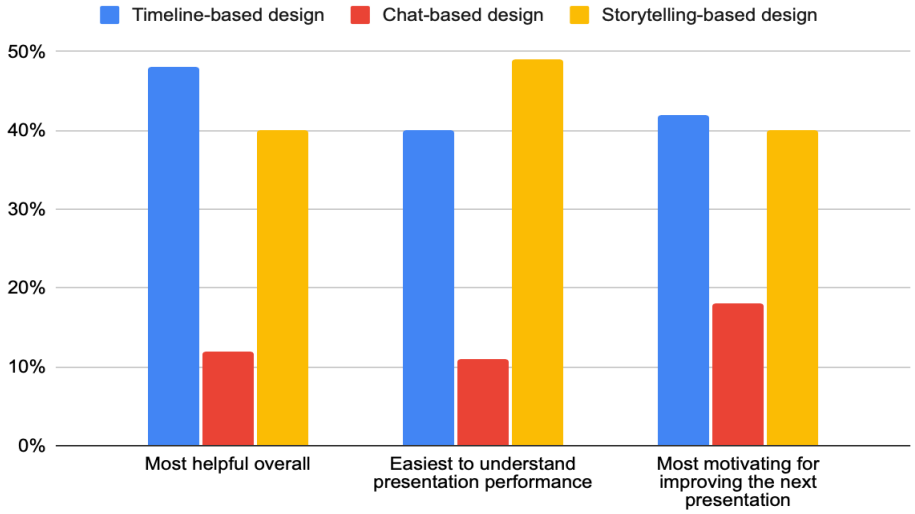


Figure 3: Direct comparison of the three Rehearsal Report designs.

5.3 Motivating Design Elements and Qualitative Feedback

Figure 4 shows which design elements participants selected as contributing most to their motivation to reflect on their presentation performance. Participants most frequently selected clear visual grouping of categories (73%), followed by highlighting key moments (61%), a clear connection between feedback and time (55%), chronological structure (50%), summary statements (48%), a visual overview of the entire presentation (46%),

and visual indicators such as icons and colours (42%). In contrast, narrative explanations were selected by 29%, and the conversational format by only 5%.

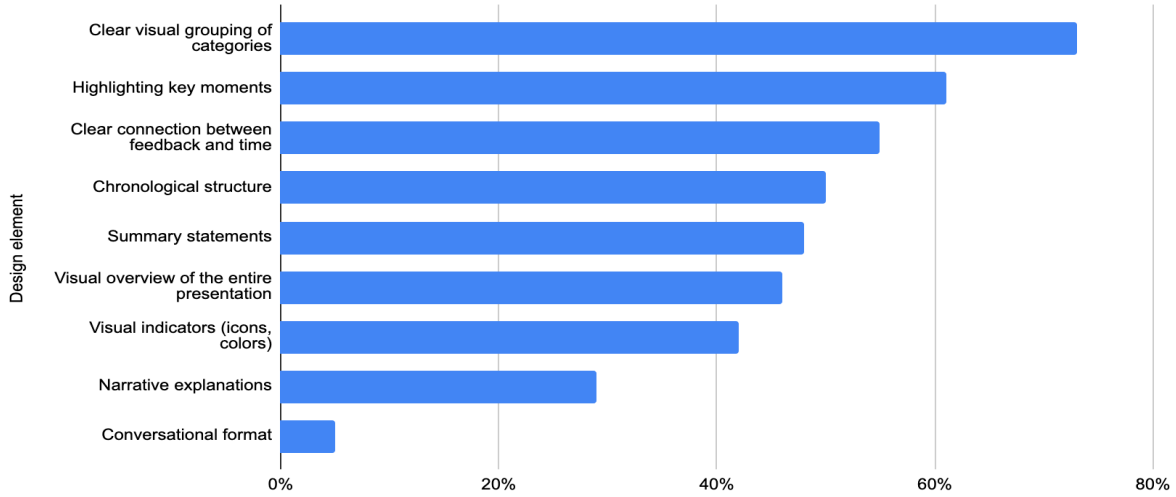


Figure 4: Design elements contributing to motivation for reflection.

The qualitative responses supported these findings. For the timeline-based design, participants often mentioned the clear structure, timestamps, chronological organization, and separation of positive feedback and improvement feedback. For the storytelling-based design, participants highlighted the clear grouping into categories, the overview, and the understandable language, but some missed precise time information. For the chat-based design, participants appreciated the personal and interactive character, but criticized the limited response options, lack of overview, and difficulty of revisiting specific feedback points.

6. Discussion

The results show that participants evaluated the three representations of post-training feedback differently, although all designs were based on the same feedback scenario. The timeline-based and storytelling-based designs were evaluated more positively than the chat-based design. This suggests that structured representations may be particularly important for how learners perceive post-training feedback after presentation practice.

The timeline-based design was especially strong in helping participants connect feedback to specific moments in the presentation. Since the feedback used in the scenario was event-based and linked to specific moments, participants may have benefited from seeing these events arranged in temporal order. This aligns with visualization principles that emphasize the value of providing both an overview and access to details (Shneiderman, 2003).

The storytelling-based design was selected most often as the design that made it easiest to understand presentation performance. This indicates that temporal detail is most likely not the only factor shaping participants' preferences. Participants also valued clear categories, structured summaries, and explanations in everyday language, which is consistent with feedback visualization research emphasizing interpretation rather than only data display (Ahn et al., 2019; McKenna et al., 2019).

The chat-based design received more mixed reactions. Some participants appreciated its personal and interactive character, but the design was rated lower overall and was most frequently ranked as the least helpful format. A central issue was the lack of overview: because information was presented sequentially, participants found it harder to review feedback quickly or compare feedback points. In addition, the predefined answer options did not match participants' expectations of a chat interface.

Across the three designs, participants consistently valued clear visual grouping, highlighting key moments, connection between feedback and time, and chronological structure. These findings suggest that participants preferred feedback formats that helped them organize and interpret performance data. For post-training feedback, clarity and contextualization appeared to be more important than conversational interaction alone. Future designs could therefore combine the structured overview of the storytelling design with the temporal detail of the timeline design.

7. Limitations and Future Work

This study has several limitations. First, the evaluation was conducted as an online survey in which participants reviewed screenshots and demonstration videos. They did not use the reports after giving their own presentation. Therefore, the results reflect perceived usefulness, clarity, motivation, and preference rather than actual use in a real training situation. Second, the feedback was based on a fictional rehearsal scenario and mock data rather than participants' own presentation recordings. Learners may respond differently when feedback is personally relevant and connected to their own performance.

The study focused on comparing different representations of the same feedback content and did not investigate how automated feedback messages themselves should be generated or formulated. The participant sample was also limited, as many participants were recruited through online survey platforms and university-related contexts. Other user groups, such as school students, professional presenters, or learners with little presentation experience, may have different expectations regarding post-training feedback. In addition, the study focused on perceived reflection support and motivation, but did not measure whether the designs improve presentation skills over time.

Future work should integrate the Rehearsal Report into Presentable and connect report items to the recorded presentation video. This would allow learners to click on feedback points, review the corresponding moment, and relate the feedback more directly to their own behaviour. Such an implementation would make it possible to evaluate the reports in real presentation training sessions rather than through screenshots and demonstration videos. Future studies should also investigate the effect of Rehearsal Reports across repeated practice and explore hybrid designs that combine the structured overview of the storytelling design with the temporal detail of the timeline design.

8. Conclusion

This paper investigated how learners perceive different representations of post-training feedback for presentation practice. Three Rehearsal Report designs were developed and compared: a timeline-based design, a chat-based design, and a storytelling-based design. The results of the online study show that participants preferred structured feedback formats over the conversational chat-based approach.

The timeline-based design was especially valued for connecting feedback to specific moments in the presentation, while the storytelling-based design was perceived as particularly easy to understand. In contrast, the chat-based design was appreciated by some participants for its personal and interactive character, but was also criticized for its limited overview and restrictive interaction options.

Overall, the findings suggest that post-training feedback should combine clear structure, visual grouping, and contextual information. For presentation training systems, Rehearsal Reports may therefore complement immediate feedback by providing a structured basis for reviewing performance after the session.

Acknowledgements

The authors would like to thank the participants of the study for their time and valuable feedback.

Ethics Declaration

The study was conducted as an anonymous online survey with adult participants who voluntarily agreed to participate. No personally identifying data was collected.

AI Declaration

AI-based writing support was used for language refinement, restructuring, and condensation of existing thesis material. The conceptual content, study design, results, and interpretation are based on the first author's bachelor thesis and were reviewed and edited by the authors.

References

Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in Context: Reaching beyond Usability in Learning Analytics Dashboard Design. *Journal of Learning Analytics*, *6*(2), 70-85.

Chollet, M., Marsella, S., & Scherer, S. (2022). Training public speaking with virtual social interactions: Effectiveness of real-time feedback and delayed feedback. *Journal on Multimodal User Interfaces*, *16*(1), 17-29.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & education*, *57*(4), 2333-2351.

Hummel, S., Schneider, J., Mouhammad, N., Klemke, R. and Di Mitri, D. (2025) "Enhancing presentation skills: key technical features of automated feedback systems – a systematic feature analysis", International Journal of Technology Enhanced Learning, Vol. 17, No. 6, pp. 1–25.

Jariyah A, A., Patak, A. A., & Novia, L. (2025). The Impact of Microsoft PowerPoint Online 'Rehearse with Coach' Feature in Improving BEC Students' Presentation Skills. *JoEELE: Journal of Excellence in English Language Education*, *4*(2), 288-296.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *European conference on technology enhanced learning* (pp. 82-96). Springer, Cham.

Maheshi, B., Milesi, M. E., Palihena, H., Zheng, A., Martinez-Maldonado, R., & Tsai, Y. S. (2024). Data Storytelling for Feedback Analytics. In *Data Storytelling and Learning Analytics Workshop 2024* (pp. 118-123). CEUR-WS.

McKenna, K., Pouska, B., Moraes, M. C., & Folkestad, J. E. (2019). Visual-form learning analytics: a tool for critical reflection and feedback. *Contemporary educational technology*, *10*(3), 214-228.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. *Instructional design for multimedia learning*, *181-195*(3).

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199-218.

Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364-371). Morgan Kaufmann.

Sichterman, B., Banihashem, S. K., Verstappen, M., Schipper, M., Waisvisz, P., Noroozi, O., & van Ginkel, S. (2025). Can AI feedback stand alone for fostering students' presentation performance? A comparison of AI-only versus teacher-supported AI feedback. *Interactive Learning Environments*, *34*(3), 1280-1297.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. Educational psychology review, 10(3), 251-296.

Taxipulati, S., Li, Y., Chen, Y., Tian, A. and Huang, Q. (2021) "The Influence of Feedback Content and Feedback Time on Multimedia Learning Achievement of College Students and Its Mechanism", Frontiers in Psychology, Vol. 12, 706821.

Theobald, M., & Bellhäuser, H. (2022). How am I going and where to next? Elaborated online feedback improves university students' self-regulated learning and performance. *The Internet and Higher Education*, *55*, 100872.

Zha, M., & Harrison, L. T. (2019). Designing and Evaluating Feedback Schemes for Improving Data Visualization Performance. *Worcester Polytechnic Institute, Tech. Rep*.

From Action to Judgement: Experiential Learning Processes in a Business Simulation Environment

Yvonne Costin, Michael P. O'Brien and William Hogan
University of Limerick, Castletroy, Limerick, Ireland

yvonne.costin@ul.ie
michaelp.obrien@ul.ie
william.hogan@ul.ie

Abstract: This study examines how postgraduate students develop decision-making capability through experiential learning processes within a business simulation environment. While simulation-based learning is widely used in business education, there is limited empirical insight into how learners internalise decision-making through experience, feedback, and consequence. This research focuses on student-reported learning processes rather than performance outcomes, providing insight into how judgement develops through repeated decision cycles. The study draws on survey data collected from master's students enrolled in an 18-credit business analytics simulation module incorporating SimVenture Evolution. The survey captured quantitative measures relating to perceived realism, confidence in decision-making, preparedness for complex business challenges, risk-taking orientation, and perceived development of judgement, complemented by open-ended responses. Findings indicate that students experienced the simulation as a realistic and demanding decision environment requiring trade-offs under uncertainty. Quantitative results show increased confidence and preparedness, while qualitative analysis highlights a learning process characterised by action, feedback, and visible consequences. Students consistently reported that learning emerged from observing the outcomes of their own decisions, with both successful and unsuccessful outcomes prompting reflection and adaptation. The iterative structure of the simulation enabled learners to refine their strategies across successive decision cycles. The study contributes empirical insight into how experiential learning processes operate in simulation environments and demonstrates how consequence, iteration, and reflection support the development of judgement in complex, uncertain decision environments.

1. Introduction

Developing decision-making capability is a central objective of contemporary business and management education. Graduates are increasingly expected not only to possess disciplinary knowledge, but also to demonstrate the ability to make informed judgements under conditions of uncertainty, complexity, and incomplete information. Traditional pedagogical approaches, which emphasise content transmission and the acquisition of procedural knowledge, have been criticised for their limited effectiveness in developing such capabilities.

In response, there has been growing emphasis on experiential learning approaches that position learners as active participants in decision-making contexts. Experiential learning is widely understood as a process through which knowledge is created through the transformation of experience (Kolb, 1984). Within business education, simulation-based learning has emerged as a prominent form of experiential pedagogy, offering learners opportunities to engage in complex, dynamic decision environments (Salas, Wildman and Piccolo, 2009; Sitzmann, 2011). In particular, digital simulation environments represent a form of technology-mediated learning that enables learners to engage with complex, dynamic decision contexts in ways that are difficult to replicate through traditional instructional approaches.

Despite the widespread adoption of simulations, much of the existing literature focuses on engagement, satisfaction, or learning outcomes, rather than the underlying processes through which learning occurs. Recent studies suggest that while simulations can enhance learning outcomes, there remains limited empirical insight into how learners interpret feedback, respond to consequences, and develop judgement over time (Eckhaus et al., 2017; Villarroel et al., 2020).

This study addresses this gap by examining how postgraduate students experience learning within a business simulation environment. Rather than focusing on performance metrics, the research investigates student-reported learning processes, with particular attention to the role of action, feedback, and consequence in shaping decision-making capability. This study therefore contributes by providing empirical insight into how experiential learning processes operate within simulation environments, with particular attention to the development of judgement through consequence and iteration.

2. Literature Review

The literature review adopts a narrative approach to synthesise research relating to experiential learning, simulation-based learning, and judgement development in higher education. The review draws on foundational theoretical contributions and contemporary empirical studies relevant to the study's focus.

2.1 Experiential Learning and Capability Development

Experiential learning theory conceptualises learning as a cyclical process involving concrete experience, reflective observation, abstract conceptualisation, and active experimentation (Kolb, 1984). This cycle emphasises that learning emerges through the transformation of experience, rather than through passive absorption of information. Subsequent work has extended this perspective by highlighting the importance of learning environments in supporting experiential processes. Kolb and Kolb (2005) argue that effective learning requires environments that enable learners to engage actively with tasks, reflect on their experiences, and apply insights in new contexts.

Recent research continues to demonstrate the relevance of experiential learning in higher education. Experiential approaches have been shown to support the development of critical thinking, problem-solving, and decision-making capabilities, particularly in complex and uncertain environments (Jones-Roberts and Bechtold, 2024). However, systematic reviews suggest that experiential learning is often only partially implemented, with limited attention to the full cycle of action, feedback, and reflection (Villarroel et al., 2020).

2.2 Simulation-Based Learning in Business Education

Simulation-based learning provides a structured environment in which learners can engage with complex decision-making scenarios (O'Brien and Costin, 2023; O'Brien et al., 2025). These environments enable students to experiment with decisions, observe outcomes, and explore the consequences of strategic choices in a risk-free setting (Salas, Wildman and Piccolo, 2009). Meta-analytic evidence suggests that simulation-based training can be highly effective in developing applied skills and improving learning outcomes (Sitzmann, 2011). In business education, simulations are particularly valued for their ability to integrate multiple functional areas and expose learners to dynamic, interdependent systems (Heyworth-Thomas, 2023). More recent work highlights the importance of feedback and reflection in enhancing the effectiveness of simulation-based learning. Structured reflection processes, such as debriefing, have been shown to support deeper learning by helping students interpret and internalise their experiences (Kumar et al., 2025). Despite these insights, much of the literature remains focused on outcomes rather than processes. There is limited understanding of how learners interpret simulation experiences, particularly in relation to feedback, uncertainty, and iterative decision-making.

2.3 Learning through Consequence and Iteration

A defining feature of simulation-based learning is the ability to experience the consequences of decisions (O'Brien and Costin, 2023; O'Brien et al., 2025). Unlike traditional instructional approaches, simulations provide immediate and visible feedback, allowing learners to observe the impact of their actions.

Research suggests that learning through consequence is central to the effectiveness of simulations. Students develop understanding by observing outcomes, particularly when those outcomes challenge expectations and highlight discrepancies between intended and actual performance (Hattie and Timperley, 2007). This aligns with broader constructivist perspectives, which emphasise active engagement and reflection in knowledge construction. Iteration also plays a critical role in this process. Repeated cycles of decision-making enable learners to refine their understanding, recognise patterns, and develop more sophisticated strategies. However, empirical research examining how iteration contributes to learning remains limited.

2.4 Research Gap

While existing literature highlights the importance of experiential learning and simulation-based education, there remains a lack of empirical insight into how learning processes unfold within simulation environments. This study addresses this gap by examining how learners experience and interpret decision-making, feedback, and outcomes.

3. Teaching Context

The study was conducted within a postgraduate business analytics module delivered at an Irish university. The module is an 18-credit component of a master's programme and is designed to provide students with an applied understanding of business decision-making in a simulated environment. The module incorporates the SimVenture Evolution simulation, in which students operate a virtual company and make decisions across multiple functional areas, including marketing, operations, finance, and strategy. Students work in teams and participate in multiple decision rounds, each representing a discrete period of business activity.

The simulation is designed to create a dynamic and uncertain decision environment. Decisions made in one round influence outcomes in subsequent rounds, requiring learners to continuously reassess their strategies. Feedback is provided through performance metrics and system-generated reports, allowing students to observe the consequences of their decisions.

4. Methodology

This study adopted a mixed-methods research design combining quantitative survey data with qualitative thematic analysis. This approach was considered appropriate because the study sought both to identify patterns in students' perceptions of learning and to explore how learners interpreted their experiences, capturing both the distribution of learner perceptions and the processes through which those perceptions were formed (Creswell and Plano Clark, 2018). The study was informed by a pragmatic research perspective, which prioritises the use of methods most appropriate to addressing the research question (Creswell and Creswell, 2018). Quantitative survey items provided insight into students' perceptions of realism, confidence, preparedness, and judgement development, while qualitative responses enabled exploration of the mechanisms through which learning occurred.

4.1 Data Collection

Data were collected from postgraduate students participating in the simulation-based module incorporating SimVenture Evolution. The survey was administered at the conclusion of the simulation activity, ensuring that responses reflected students' full experience of the decision-making process. The instrument included Likert-scale items designed to measure key dimensions of the learning experience, including perceived realism, engagement, confidence in decision-making, preparedness to address complex business challenges, risk-taking orientation, and perceived development of judgement. These items provided an overview of how students evaluated the simulation as a learning environment. In addition, open-ended questions were included to capture qualitative insights into how students experienced decision-making within the simulation. These responses focused on perceptions of uncertainty, the role of feedback and outcomes, and how students adapted their strategies over time.

A total of 33 responses were collected.

4.2 Data Analysis

Quantitative data were analysed using descriptive statistics to identify overall patterns in student responses across the survey items. Frequencies and levels of agreement were examined to provide an indication of how students perceived the simulation experience and its contribution to decision-making capability. Given the exploratory nature of the study and the relatively small cohort size, the quantitative analysis was used primarily to identify broad trends and support interpretation of the qualitative findings.

Qualitative responses were analysed using an inductive thematic analysis approach (Braun and Clarke, 2006, 2022). Thematic analysis was selected because it provides a flexible and systematic method for identifying patterns of meaning within qualitative datasets. Responses were reviewed repeatedly to achieve familiarity with the data before inductive coding was undertaken. Related codes were subsequently grouped into broader themes relating to consequence, feedback, reflection, and iteration. These themes emerged through iterative comparison of participant responses and were subsequently refined to ensure they accurately reflected the dataset. To enhance trustworthiness, themes were reviewed against the original dataset to ensure consistency between participant responses and thematic interpretations.

The combination of quantitative and qualitative data enabled a more comprehensive understanding of the learning experience, allowing patterns in student perceptions to be interpreted alongside detailed accounts of how those perceptions were formed.

5. Findings

5.1 Perceived Realism and Decision Complexity

Students consistently described the simulation as providing a realistic and demanding decision-making environment. Quantitative responses indicated consistently high levels of agreement regarding the realism of the simulation, with over 80% of students selecting 'agree' or 'strongly agree' across related items.

Qualitative responses provide further insight into how this realism was experienced. Students frequently referred to the interconnected nature of decisions, noting that choices made in one functional area (e.g. pricing or marketing) had implications for other aspects of the business, such as cash flow, staffing, or production capacity. This interdependence contributed to a perception of complexity that extended beyond isolated decision tasks.

A recurring theme in the data was the challenge of making decisions without complete information. Students highlighted the difficulty of predicting outcomes and described a sense of uncertainty that required them to move beyond formulaic approaches. Rather than applying predefined solutions, learners reported needing to interpret available data, make assumptions, and accept the possibility of unintended consequences. This suggests that the simulation successfully created a context in which decision-making was experienced as uncertain, contingent, and context-dependent. Importantly, realism was not only associated with the structure of the simulation but also with the consequences of decisions. Students emphasised that outcomes felt meaningful, particularly when poor decisions resulted in visible negative impacts on performance. This reinforced the perception that decisions carried weight, contributing to a more engaged and reflective learning experience.

5.2 Development of Confidence and Decision-Making Capability

Quantitative findings indicate that participation in the simulation was associated with increased confidence in decision-making. A clear majority of respondents (over 80%) reported increased confidence, with responses strongly concentrated in the 'agree' and 'strongly agree' categories. Students also reported greater willingness to engage with risk, although responses in this area exhibited slightly greater variation, suggesting that comfort with uncertainty developed unevenly across the cohort.

Qualitative data suggest that this increase in confidence was not simply a result of familiarity with the simulation interface, but rather emerged from repeated engagement with decision-making tasks. Students described an initial period of uncertainty and hesitation, followed by a gradual increase in confidence as they gained experience with the consequences of their decisions.

Several responses highlighted that confidence developed through exposure to both successful and unsuccessful outcomes. Early mistakes were often described as challenging, but also as important learning moments that contributed to improved decision-making in later rounds. As students progressed through the simulation, they reported becoming more comfortable making decisions under uncertainty and more willing to take calculated risks. This suggests that confidence was not derived from success alone, but from the process of engaging with decision-making over time. The iterative structure of the simulation allowed students to test their assumptions, learn from outcomes, and refine their approach, contributing to a more robust sense of perceived capability.

5.3 Learning through Consequences of Decisions

A central finding of the study is the importance of learning through the consequences of decisions. Students consistently emphasised that observing the outcomes of their actions was the primary mechanism through which learning occurred.

Quantitative responses indicate that students perceived a strong link between their decisions and subsequent outcomes, with the majority indicating agreement that outcomes reflected their strategic choices. This perception of causality appears to be critical, as it enabled learners to connect their actions with results in a meaningful way.

Qualitative responses reinforce this finding. Students frequently described learning as emerging from seeing "what happened" after making decisions, particularly when outcomes deviated from expectations. Negative outcomes were often highlighted as especially impactful, prompting learners to reconsider their assumptions and approach.

Importantly, students did not attribute learning primarily to instruction, feedback from instructors, or external guidance. Instead, they emphasised the role of experiencing consequences directly. This suggests that the simulation functioned as an environment in which learners could generate their own feedback through action, rather than relying solely on external evaluation. This finding highlights the importance of consequence visibility in simulation design. When learners are able to clearly see the outcomes of their decisions, they are more likely to engage in reflection and develop a deeper understanding of cause-and-effect relationships.

5.4 Iterative Decision-Making and Strategy Refinement

The iterative structure of the simulation emerged as a key factor supporting learning. Students engaged in multiple rounds of decision-making, allowing them to test strategies, observe outcomes, and adjust their approach over time. Most respondents (approximately three-quarters of the cohort) indicated that repeated decision cycles supported their learning.

Qualitative responses provide a more detailed picture of how iteration contributed to learning. Students described a process in which initial decisions were often exploratory or uncertain, followed by increasing levels of intentionality as they gained experience. Early rounds were characterised by trial-and-error, while later rounds involved more deliberate and informed decision-making.

Learners also reported becoming more aware of patterns and interdependencies over time. Through repeated exposure to decision–outcome relationships, students developed a more nuanced understanding of how different variables interacted. This enabled them to anticipate potential consequences and make more strategic decisions in subsequent rounds. The temporal dimension of learning was particularly evident in student reflections. Many participants emphasised that understanding developed gradually, rather than immediately. The opportunity to revisit decisions, observe longer-term outcomes, and refine strategies was seen as essential to the learning process.

5.5 Reflection, Feedback, and Adaptation

Reflection emerged as a critical component of the learning process, closely linked to feedback and outcome visibility. Students described engaging in both individual and group reflection following each decision round, using outcomes to evaluate their performance and inform future decisions.

Quantitative responses indicate that students valued the feedback provided by the simulation, particularly in terms of its immediacy and clarity, with the majority reporting that feedback helped them understand the impact of their decisions. Qualitative data suggest that reflection was often triggered by unexpected or undesirable outcomes. When results did not align with expectations, students reported being prompted to reconsider their assumptions and analyse their decision-making process. This reflective activity frequently led to adjustments in strategy, demonstrating a clear link between feedback, reflection, and adaptation.

Students also highlighted the role of team discussions in facilitating reflection. Collaborative analysis of outcomes allowed learners to share perspectives, challenge assumptions, and develop a more comprehensive understanding of the situation. This suggests that reflection was both an individual and social process, contributing to deeper learning. Importantly, reflection was not limited to correcting mistakes. Positive outcomes also prompted reflection, with students seeking to understand why certain strategies were successful. This indicates that reflection operated as a continuous process, supporting both reinforcement and revision of decision-making approaches.

5.6 Synthesis of Learning Process

Taken together, the findings suggest a coherent experiential learning process characterised by the interaction of action, feedback, reflection, and adaptation. Students engaged in cycles of decision-making in which each stage informed the next, contributing to the development of judgement over time. This process can be understood as a dynamic sequence: learners make decisions, observe outcomes, reflect on those outcomes, and adjust their approach in subsequent iterations. The visibility of consequences and the opportunity for repeated

engagement appear to be critical factors enabling this process. Rather than learning through instruction alone, students developed understanding through direct engagement with decision-making and its consequences. The simulation provided a structured environment in which learners could experience responsibility for decisions, encounter the effects of those decisions, and refine their approach through iterative cycles.

This synthesis highlights the importance of designing learning environments that support not only action, but also meaningful feedback, reflection, and iteration. It is the interaction of these elements, rather than any single component, that appears to underpin the development of decision-making capability.

6. Discussion

6.1 Experiential Learning as a Process, Not a Pedagogical Label

A key contribution of this study is the identification of experiential learning as a *processual phenomenon* rather than a pedagogical label. Although simulation-based learning is frequently described as "experiential", the findings suggest that experiential learning does not arise automatically from the use of simulation technology. Instead, it emerges through a specific sequence of learner activities: making decisions, observing outcomes, reflecting on those outcomes, and adapting subsequent behaviour.

This aligns with the cyclical model of experiential learning proposed by Kolb (1984), but the present study extends this perspective by evidencing how learners themselves experience and articulate this cycle. Importantly, the data indicate that students do not consciously move through formal stages of experiential learning. Rather, the process is enacted through engagement with the simulation, with reflection and adaptation triggered by the visibility and impact of decision outcomes. This suggests that experiential learning in simulation contexts is less about structured pedagogical stages and more about the *conditions* that enable learners to engage in iterative cycles of action and reflection. In this sense, the findings reinforce the argument that effective experiential learning depends on the design of the learning environment, rather than on the nominal use of experiential methods.

6.2 The Central Role of Consequence in Learning

One of the most significant findings of this study is the central role of consequence in shaping learning. Students consistently reported that learning occurred through observing the outcomes of their decisions, particularly when those outcomes differed from expectations. This highlights consequence as a key mechanism through which learners make sense of their actions and develop understanding.

Prior research has emphasised the importance of structured reflection and debriefing in simulation-based learning (Kumar et al., 2025), while feedback more broadly is recognised as one of the most powerful influences on learning (Hattie and Timperley, 2007). The findings of this study suggest that what matters is not simply the presence of feedback, but its perceived authenticity and causal linkage to learner actions. Students interpreted outcomes as meaningful because they were understood as direct consequences of their own decisions rather than externally imposed evaluations.

Simulation environments therefore provide a distinctive learning context in which learners generate feedback through action. This creates opportunities to develop a deeper understanding of cause-and-effect relationships and encourages greater ownership of learning. The findings suggest that the effectiveness of simulation-based learning depends not only on providing feedback, but on ensuring that outcomes are visible, meaningful, and clearly connected to learner decisions.

6.3 Iteration and the Temporal Dimension of Learning

The study highlights iteration as a core mechanism of experiential learning. Students did not develop decision-making capability through isolated experiences, but through repeated cycles of action, feedback, and reflection. Through successive decision rounds, learners became more confident, recognised patterns in outcomes, and refined their strategies over time.

The findings suggest that experiential learning is inherently temporal. Understanding developed gradually as students tested assumptions, observed consequences, and adapted their approach. The sequencing of decision rounds and the persistence of consequences across those rounds were therefore not simply features of the simulation design but important pedagogical mechanisms that supported learning. This suggests that the value

of simulation-based learning lies not only in providing realistic decision environments, but in enabling sustained cycles of experimentation, consequence, and adaptation.

6.4 Learning from Failure and the Productive Role of Error

An important insight emerging from the findings is the constructive role of failure in learning. Students frequently described negative outcomes as particularly valuable learning experiences, as they prompted reflection and reconsideration of assumptions. This supports the view that error and failure are integral to experiential learning, rather than outcomes to be avoided.

In traditional educational settings, failure is often minimised or penalised, which can discourage risk-taking and experimentation. In contrast, the simulation environment allowed students to experience failure without real-world consequences, creating a safe space for exploration and learning. This aligns with research suggesting that simulations can support learning by allowing learners to engage with risk and uncertainty in a controlled environment (Sitzmann, 2011).

However, the findings also suggest that failure alone is not sufficient to support learning. It is the combination of failure with visible consequences and opportunities for iteration that enables learners to reflect and adapt. Without these elements, failure may not translate into meaningful learning. This highlights the importance of designing learning environments that not only allow for failure, but also support learners in interpreting and responding to it. In this study, the immediacy of feedback and the opportunity to apply revised strategies in subsequent rounds appear to have been critical in transforming failure into learning.

6.5 From Decision-Making to Judgement Development

While much of the literature on business education focuses on decision-making as a skill, the findings of this study suggest that simulation-based learning supports the development of *judgement* rather than simply decision-making techniques. Students did not merely learn how to make decisions; they developed a more nuanced understanding of when and how to make decisions in complex and uncertain contexts. This distinction is important. Decision-making can be understood as the application of tools or frameworks, whereas judgement involves the ability to interpret situations, evaluate trade-offs, and make contextually appropriate decisions. The findings indicate that judgement develops through experience, particularly through engagement with uncertainty and consequence.

Students reported becoming more aware of interdependencies, recognising patterns in outcomes, and adapting their strategies accordingly. These behaviours are indicative of developing judgement, rather than simply applying predefined rules. This suggests that simulation-based learning can support higher-order learning outcomes that are difficult to achieve through traditional instructional approaches.

6.6 Contribution to Research and Practice

This study makes several contributions to the literature on experiential learning and simulation-based education. First, it provides empirical evidence of how experiential learning processes operate in a business simulation context, addressing a gap identified in recent reviews (Villarroel et al., 2020). Rather than simply applying experiential learning theory, the study offers insight into how specific mechanisms – particularly consequence visibility, iterative decision cycles, and reflection – interact in practice to support the development of judgement.

Second, the findings extend existing work on simulation-based learning by shifting the focus from outcomes such as engagement or performance to the underlying processes through which learning occurs. The results demonstrate that learning is not an automatic outcome of simulation use, but emerges through structured interaction with decision-making, feedback, and consequence over time.

From a practical perspective, the study highlights that the effectiveness of simulation-based learning depends on the design of the learning environment. Specifically, simulations should be structured to:

- Make consequences visible and meaningful
- Support iterative decision-making over time
- Encourage reflection through feedback and discussion
- Allow for failure as part of the learning process

By foregrounding these elements, educators can move beyond using simulations as engagement tools and instead leverage them as environments for developing judgement and decision-making capability.

7. Implications for Practice

The findings suggest that the effectiveness of simulation-based learning depends less on the use of simulation technology itself and more on how the learning environment is designed to support experiential learning processes. In particular, consequence, iteration, and reflection emerged as key mechanisms underpinning learning. Simulations should make the consequences of decisions visible and meaningfully linked to learner actions, allowing students to observe cause-and-effect relationships and reflect on outcomes. Learning is further strengthened through repeated decision cycles that enable learners to test assumptions, refine strategies, and build confidence over time.

The findings also highlight the importance of structured opportunities for reflection, both individually and collaboratively. Educators play an important role in facilitating interpretation of outcomes rather than providing solutions. Finally, simulation environments should allow learners to experience productive failure and adapt their approach in subsequent rounds.

Overall, the study suggests that simulation-based learning can support the development of judgement when learning environments are deliberately designed to foreground consequence, iteration, and reflection.

8. Limitations and Future Research

This study provides insight into experiential learning processes within a business simulation environment; however, several limitations should be acknowledged. First, the study is based on self-reported data, which reflects students' perceptions of their learning rather than direct measures of decision-making performance. While these perceptions are valuable in understanding how learners interpret their experiences, they may be subject to bias. In particular, self-reported data does not provide direct evidence of behavioural change in decision-making capability, which represents an important avenue for future research.

Second, the study is situated within a single module and institutional context. Although the findings are consistent with broader experiential learning theory, their generalisability to other disciplines, levels of study, or simulation designs may be limited. Differences in cohort characteristics, facilitation approaches, and simulation structures could influence how learning processes unfold.

Third, the analysis focuses on learning within the timeframe of the simulation. While students reported increased confidence and perceived capability, the study does not examine whether these effects persist over time or transfer to real-world decision-making contexts.

These limitations suggest several directions for future research. Longitudinal studies could examine how decision-making capability develops beyond the simulation and whether learning is retained over time. Further research could incorporate behavioural or performance-based measures to complement self-reported data and provide a more robust assessment of learning outcomes. In addition, comparative studies across different simulation designs or pedagogical approaches would help identify which features most effectively support experiential learning processes. Overall, while the findings should be interpreted within the context of these limitations, they provide a strong foundation for further investigation into how experiential learning environments can be designed to support the development of judgement and decision-making capability.

9. Conclusion

This study set out to examine how decision-making capability develops within a business simulation environment, with a particular focus on experiential learning processes. The findings demonstrate that learning in this context is not primarily driven by instruction or content delivery, but by structured engagement with decision-making, consequence, and iteration. Students developed capability through repeated cycles of action, feedback, reflection, and adaptation, rather than through the application of predefined models or techniques.

The study contributes to the literature by providing empirical evidence of how experiential learning processes unfold in practice. In doing so, the study contributes not only to simulation-based learning, but to a broader understanding of how judgement develops through structured experiential processes in higher education. The findings highlight that it is the visibility of consequences, the opportunity for repeated decision cycles, and the space for reflection that underpin meaningful learning.

Importantly, the development observed is better understood as the formation of judgement rather than the acquisition of decision-making skills alone. Learners demonstrated increasing capacity to interpret complex situations, evaluate trade-offs, and adapt their strategies in response to changing conditions. This suggests that simulation-based environments can support higher-order learning outcomes that are difficult to achieve through more traditional instructional approaches.

The implications of this study extend beyond business simulations. More broadly, it highlights the importance of designing learning environments that enable learners to experience the consequences of their actions, engage in iterative decision-making, and reflect on outcomes over time. By foregrounding these elements, educators can create conditions that support the development of capability rather than the reproduction of knowledge.

Future research should build on this work by examining how these processes evolve over longer timeframes and across different simulation designs. In particular, there is scope to explore how variations in feedback, structure, and facilitation influence the development of judgement in experiential learning contexts.

Ethics Declaration

This study involved human participants, and formal ethical approval was obtained from the university's ethics committee prior to data collection. All participants were informed of the purpose of the study and provided informed consent in accordance with institutional ethical guidelines. Participation was voluntary, and all responses were collected anonymously to ensure confidentiality.

AI Declaration

Artificial intelligence tools were used during the preparation of this manuscript to support proofreading, language refinement, and clarification of ideas. All intellectual contributions, critical analysis, and final decisions regarding the content were made by the authors.

References

Braun, V. and Clarke, V. (2006) "Using thematic analysis in psychology". *Qualitative Research in Psychology*, 3(2), pp. 77-101.

Braun, V. and Clarke, V. (2022) "*Thematic Analysis: A Practical Guide"*. London: Sage.

Creswell, J.W. and Creswell, J.D. (2018) "*Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*". Sage, Los Angeles.

Creswell, J.W., and Plano Clark, V.L. (2018) "Designing and conducting mixed methods research" (3rd ed.). Thousand Oaks, CA: SAGE.

Eckhaus, E., Klein, G. and Kantor, J. (2017) "Experiential learning in management education", *Business, Management and Education*, 15(1), pp. 42-56.

Hattie, J. and Timperley, H. (2007) "The power of feedback", *Review of Educational Research*, 77(1), pp. 81-112.

Heyworth-Thomas, E.M. (2023) "Creating experiential learning opportunities in enterprise education: an example of a facilitator-led business simulation game in a taught setting", *Journal of Work-Applied Management*, 15(1), pp. 173-187.

Jones-Roberts, C. and Bechtold, C. (2024) "Experiential Learning in Higher Education", Journal of Technology-Integrated Lessons and Teaching, 3(2), pp. 53-63.

Kolb, D.A. (1984) *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.

Kolb, A.Y. and Kolb, D.A. (2005) "Learning styles and learning spaces: enhancing experiential learning in higher education", *Academy of Management Learning & Education*, 4(2), pp. 193-212.

Kumar, P., Harrison, N., McAleer, K. *et al.* (2025). "Exploring the role of self-led debriefings within simulation-based education: time to challenge the status quo?", *Advances in Simulation,* 10(9).

O'Brien, M.P. and Costin, Y. (2023) "Using gamification to develop students as strategic thinkers", *Proceedings of the 22nd European Conference on e-Learning*, 22(1), pp. 226-233.

O'Brien, M.P., Costin, Y., and Hogan, W. (2025) "Beyond the Game: Identity, Ethics, and the Transformative Power of Role Play in Business Simulations", *Proceedings of the 19th European Conference on Games-Based Learning*, 19(2), pp. 665-673.

Salas, E., Wildman, J., and Piccolo, R. (2009) "Using simulation-based training to enhance management education. Management: Educating managers from an evidence-based perspective", *Academy of Management Learning & Education*, 8, pp. 559-573.

Sitzmann, T. (2011) "A meta-analytic examination of the instructional effectiveness of computer-based simulation games", *Personnel Psychology*, 64(2), pp. 489-528.

Villarroel, V., Benavente, M., Chuecas, M.J., and Bruna, D. (2020) "Experiential learning in higher education. A student-centered teaching method that improves perceived learning", *Journal of University Teaching & Learning Practice*, 17(5).

Technology-Enhanced Peer Assessment via FeedbackFruits: A Multi-Framework Evaluation of Digital Technology Integration in Pedagogical Design

Xiangping Du and Bobby Pothabathula
University of Hertfordshire Business School, Hatfield, United Kingdom

x.du1@herts.ac.uk
g.pothabathula@herts.ac.uk

Abstract: The rapid digitalisation of higher education has intensified interest in educational technologies that support engagement, collaboration and meaningful learning. However, research consistently demonstrates that technology alone does not improve educational outcomes; rather, impact depends on the quality of pedagogical integration. Emerging educational technologies such as FeedbackFruits are increasingly used to support innovative, student-centred pedagogy in higher education, particularly within digitally mediated and culturally diverse learning environments. This study investigates how the integration of FeedbackFruits within a large postgraduate programme at a UK university influences student engagement, peer interaction, and feedback practices among a predominantly international cohort. This study adopts an integrated theoretical lens combining the Community of Inquiry (CoI), Constructive Alignment (CA), and Technological Pedagogical Content Knowledge (TPACK) frameworks to evaluate the pedagogical integration of FeedbackFruits for technology-enhanced peer assessment. The CoI framework is used to examine how meaningful learning emerges through the interplay of cognitive, social, and teaching presence in online and blended environments. Within this framing, FeedbackFruits is analysed as a mediating tool that structures and sustains these presences by enabling dialogic peer interaction and iterative feedback processes. CA informs the pedagogical design by emphasising coherence between intended learning outcomes, learning activities, and assessment strategies. The platform's affordances, including rubric-guided peer review and sequenced collaborative tasks, are examined in relation to how they support aligned and transparent assessment practices. The TPACK framework further situates technology use within the intersection of pedagogy and disciplinary content, highlighting that effective integration depends on pedagogical intent rather than technological novelty. Drawing on a case study involving over 700 postgraduate students who were introduced to FeedbackFruits, the research analyses student participation data, reflective feedback, and learning analytics to explore how structured peer review and collaborative activities shape the learning experience. Findings indicate strong evidence of engagement, reflective learning, accountability, feedback literacy and collaborative participation. Students spent substantial time reviewing peer contributions, generated extensive feedback comments and engaged in structured self-assessment processes through FeedbackFruits. The study argues that transformative digital learning emerges when technology is selected through TPACK principles, embedded through CA and enacted through CoI processes. Overall, the paper suggests that the value of educational technology lies in its meaningful integration with pedagogical frameworks, offering insights for designing inclusive and engaging learning environments in higher education.

Keywords: Transformative Digital Pedagogy, Educational Technology, FeedbackFruits, Constructive Alignment (CA), Community of Inquiry (CoI), Technological Pedagogical Content Knowledge (TPACK)

1. Introduction

The increasing digitalisation of higher education has transformed how teaching, learning and assessment are designed and experienced. Universities worldwide have adopted a growing range of educational technologies to enhance student engagement, support collaborative learning and facilitate flexible modes of delivery. While the COVID-19 pandemic accelerated the adoption of digital learning environments, the subsequent challenge for educators has been to move beyond emergency remote teaching towards sustainable and pedagogically meaningful digital transformation (Rapanta et al., 2021). Within this context, educational technologies are increasingly evaluated not solely on their technical functionality but on their ability to support student-centred learning, critical engagement and inclusive educational practices (Stephenson and Torn, 2023).

Among contemporary educational technologies, FeedbackFruits has gained prominence as a platform designed to facilitate collaborative learning, peer feedback and active student participation. Through features such as structured peer review, collaborative annotation, interactive discussions and rubric-guided assessment, the platform aligns with contemporary pedagogical approaches that emphasise dialogue, reflection and learner agency (FeedbackFruits, 2025). However, early research by Kirkwood and Price (2014) suggests that educational technologies do not generate meaningful learning outcomes simply through their adoption. Rather, their effectiveness depends upon how they are integrated within coherent pedagogical designs. Similarly, Fullan and Langworthy (2014) suggests digital pedagogy extends beyond the use of technological tools and focus instead on how technology reshapes interactions among learners, educators and knowledge. Rapanta et al. (2020)

further affirm the importance of balancing technology, pedagogy and the new normal, and suggest effective digital pedagogy should consider active learning, collaboration, student agency and authentic assessment. This echoes with Stephenson and Torn (2023) who advocate higher education institutions to review, recall and reposition post-Covid and to adapt delivery and design in the digital world. Rather than viewing students as passive recipients of information, digital pedagogy encourages learners to participate actively in knowledge construction through dialogue, reflection and collaborative inquiry. Educational technologies therefore function as mediating tools that support social and cognitive learning processes.

The shift towards student-centred approaches is particularly significant in higher education, where diverse student populations increasingly participate in blended and online learning environments. International postgraduate cohorts often bring varied educational experiences, cultural perspectives and expectations regarding learning and assessment. Consequently, digital learning environments must support not only academic achievement but also social integration, intercultural communication and equitable participation (Mittelmeier and Rienties, 2024)

Despite a growing body of research recognising FeedbackFruits as an effective platform for active learning and collaborative peer assessment, with demonstrated benefits for student satisfaction, reflective practice, feedback quality, and the overall peer assessment experience (Campbell et al., 2020; Vaz-Serra, 2023; Myklebust, 2025), limited attention has been paid to the pedagogical design underpinning its implementation or the theoretical foundations that explain its effectiveness, particularly within diverse international student cohorts. To address this gap, this paper draws on three influential theoretical perspectives that collectively provide a robust foundation for understanding the pedagogical value of FeedbackFruits in higher education: the Community of Inquiry (CoI) framework, Constructive Alignment (CA), and the Technological Pedagogical Content Knowledge (TPACK) framework. It argues that these complementary perspectives are intrinsically interconnected and, when considered together, offer a holistic explanation of how technology-enhanced learning environments can foster meaningful engagement, collaborative knowledge construction, reflective learning, and transformative educational experiences.

2. Literature Review

2.1 Community of Inquiry (CoI): Understanding the Learning Processes

The Community of Inquiry (CoI) framework, developed by Garrison, et al. (2000), remains one of the most influential theoretical models for understanding learning in online and blended environments. Rooted in social constructivist perspectives, the framework conceptualises meaningful educational experiences as emerging through the interaction of three interconnected dimensions: cognitive presence, social presence and teaching presence. Cognitive presence refers to the extent to which learners can construct and confirm meaning through sustained reflection and discourse. It emphasises higher-order thinking, problem-solving and critical inquiry (Akyol and Garrison, 2011). Social presence concerns learners' ability to project themselves socially and emotionally within a learning community. It involves the development of trust, communication and interpersonal relationships that facilitate collaboration (Garrison et al., 2010). Teaching presence encompasses instructional design, facilitation and direct instruction. It reflects the role of educators in creating and sustaining productive learning environments (Garrison, et al., 2010). Shea and Bidjerano (2009) argue that within CoI framework, teaching presence significantly influences both cognitive and social presence by providing guidance, scaffolding and opportunities for interaction.

Recent studies have extended the CoI framework to contemporary digital learning environments characterised by learning analytics, collaborative platforms and artificial intelligence-supported tools (Fiock, 2020; Li, et al., 2025). These developments suggest digital and AI tools help revolutionise higher education as they boost student engagement in learning through constructive and social collaborative learning. Meanwhile, technologies should not be viewed merely as delivery mechanisms but as mediating artefacts that facilitate the interactions underpinning meaningful learning. FeedbackFruits, as an emerging digital technological tool exemplifies this approach by structuring dialogue, guiding peer assessment and supporting iterative feedback processes that strengthen all three dimensions of the CoI framework. Myklebust's (2025) research affirms that FeedbackFruits provide a digital learning environments supporting participation, discussion and reflection which knowledge is constructed with cognitive presence. Structured peer interaction facilitated by digital platforms can help establish connections between learners and foster a sense of belonging. However, despite the widespread adoption of CoI and its relevance in analysing student engagement and collaborative learning in

digitally enhanced higher education, Cleveland-Innes and Campbell (2012) has argued that the model underrepresents emotional engagement and learner identity, particularly within multicultural learning contexts. Fiock (2020) echoes that in culturally diverse setting, the ability to socially and emotionally engage is particularly important as students navigate differences in educational backgrounds, communication styles and cultural expectations.

2.2 Constructive Alignment (CA): Designing Meaningful Learning and Assessment

While CoI explains how learning occurs, Constructive Alignment (CA) provides a framework for understanding how learning environments should be designed. Introduced by Biggs (1996), CA provides a foundational framework for instructional design and assessment in higher education. The theory combines constructivist learning principles with systematic curriculum alignment, arguing that effective learning occurs when intended learning outcomes, teaching activities and assessment tasks are coherently integrated. The core of CA is the premise that learners actively construct knowledge through engagement with learning activities. Therefore, educational design should ensure that teaching methods and assessment strategies directly support the achievement of intended learning outcomes (Biggs, et al., 2022). Misalignment between these elements can lead to superficial learning, confusion and lack of motivation. The relevance of CA has increased significantly in digital education contexts where educators often face challenges integrating technology with pedagogical objectives. CA also has implications for inclusivity and learner diversity, especially with assessment. It supports transparent assessment criteria and clear expectations which helps reduce ambiguity for international students who may be unfamiliar with local academic conventions.

Contemporary higher education increasingly emphasises self-assessment, peer assessment and feedback literacy as mechanisms for promoting active learning and self-regulation (Carless and Winstone, 2023). Peer assessment has been regarded as an effective learning tool as Topping (2017) where students are learning by judging and discussing the work of other learners. The research of Iglesias Pérez, et al. (2022) on self and peer assessment reveals they are more effective than lecturer feedback, especially when anonymity retained through online learning platforms. It supports critical, responsible and reflective learners as students benefit as much from providing feedback as from receiving it, as the evaluative process encourages deeper engagement with assessment criteria and learning standards.

Digital tools like FeedbackFruits supports these approaches through structured self and peer review processes, rubric-guided evaluation and opportunities for iterative improvement (Campbell, et al. 2020). Meanwhile, they provide explicit rubrics, and guided feedback processes can therefore contribute to more equitable learning experiences across diverse cohorts. By embedding peer review within a structured framework, FeedbackFruits aligns assessment activities with intended learning outcomes and promotes transparency in evaluation practices (Myklebust, 2025).

2.3 Technological Pedagogical Content Knowledge (TPACK): Technology Integration

While the CoI framework focuses on learning processes and CA addresses curriculum design, the TPACK framework provides a lens for understanding effective technology integration in educational practice. Developed by Mishra and Koehler (2006), TPACK model proposes that effective teaching with technology requires the integration of three domains: content knowledge (CK), pedagogical knowledge (PK) and technological knowledge (TK). Chai, et al. (2013) suggest meaningful educational technology use occurs at the intersections of these domains; while Koehler, et al. (2013) argue that successful technology integration depends more on pedagogical intentionality than technological sophistication and technology should be selected and implemented in ways that support disciplinary learning objectives and pedagogical goals. Thus, rather than assuming that educational technologies automatically improve learning, the framework emphasises the importance of pedagogical decision-making. Meanwhile, it also highlights the importance of contextual factors influencing technology adoption. Institutional culture, disciplinary norms, student characteristics and available resources all shape how educational technologies are implemented (Koehler, et al., 2013).

This perspective is particularly relevant when examining platforms such as FeedbackFruits employed in higher education, where digital tools must accommodate complex disciplinary knowledge and increasingly diverse student populations. The educational value of the platform does not derive solely from its digital features but from how those features support collaborative learning, assessment and disciplinary knowledge construction (Vaz-Serra et al, 2023). In business studies amongst an internationally diverse groups, these contextual considerations become particularly significant as students bring varying levels of digital literacy and diverse

educational experiences (Li, et al., 2025). Therefore, it is vital to understand how FeedbackFruits decodes the relationship between pedagogy, technology and content knowledge in business studies under TPACK model.

2.4 Integrating CoI, Constructive Alignment and TPACK

Although CoI, CA and TPACK each provide valuable insights, none individually offers a complete explanation of transformative digital pedagogy. CoI focuses on learning experiences, CA emphasises curriculum coherence and TPACK addresses pedagogical decision-making regarding technology use. Importantly, all three frameworks share a common assumption: technology is not inherently transformative. Instead, educational technologies become valuable when integrated within pedagogically informed learning environments that support engagement, collaboration and knowledge construction. This perspective aligns with contemporary scholarship advocating a shift from technology-centred to pedagogy-centred approaches to digital transformation in higher education (Kirkwood and Price, 2014).

Together, the three frameworks offer complementary perspectives on educational technology integration. They suggest that the platform's effectiveness lies not in its technological capabilities alone but in its capacity to support collaborative learning, aligned assessment and pedagogically purposeful engagement. Collectively, they offer a robust theoretical foundation for examining the educational value of FeedbackFruits in higher education. As higher education continues to evolve within increasingly digital and culturally diverse contexts, these frameworks provide valuable guidance for designing inclusive, engaging and effective learning experiences.

3. Methodology

3.1 Research Design

This case study research (Yin, 2018) investigated the pedagogical impact of a technological tool of FeedbackFruits within a large postgraduate marketing module at a UK University in the 2025-26 academic year. The module enrolled 701 students and was among the largest postgraduate modules within the University. The cohort was highly international, representing a diverse range of cultural, linguistic, and educational backgrounds. This diversity provided an opportunity to explore technology-supported peer learning within multicultural environments. Informed by three conceptual frameworks: Community of Inquiry (CoI), Constructive Alignment (CA), and Technological Pedagogical Content Knowledge (TPACK), this study examined how technology-enhanced peer evaluation supports student involvement, feedback literacy, and collaborative learning within a culturally diverse higher education environment.

The study was situated within an authentic assessment in which students worked collaboratively to develop a strategic digital marketing campaign for a live client organisation, including evidence-based marketing solutions informed by consumer insights, market research, and campaign planning. The assessment comprised a group poster exhibition, a self-assessment and a peer-evaluation activity using FeedbackFruits within the University's virtual learning environment (VLE). FeedbackFruits was implemented as a structured self- and peer-assessment mechanism to address challenges commonly associated with group assessment, such as unequal contributions, free-riding, and perceptions of unfairness, and accordingly designed to promote accountability, transparency, critical reflection, and feedback literacy (Campbell, et al., 2020). Accordingly, students were instructed to evaluate their peers against five criteria: Contribution, Timeliness, Quality of Work, Team Collaboration, and Reflection which were preset within FeedbackFruits.

This study adopted a quantitative approach using institutional learning analytics generated through the FeedbackFruits platform to evaluate student engagement, peer assessment behaviour, reflective practice feedback literacy (Carless and Winstone, 2023; Iglesias Pérez, 2022). Students were introduced to FeedbackFruits through face-to-face teaching sessions and written guidance provided within the VLE. The purpose of the peer evaluation activity, assessment criteria, expectations for constructive feedback, and the contribution of peer evaluation to the overall module assessment were clearly explained. Students were encouraged to complete all components of the activity, although completion of every task was not mandatory to pass the module. Throughout the assessment, FeedbackFruits automatically generated learning analytics, including participation rates, review completion, self-assessment progress, peer ratings, review duration, and qualitative feedback comments. This platform generated analytics formed the primary dataset for this study.

The dataset included:

- Student Completion Records
- Peer review participation
- Self-assessment completion rates
- Review duration metrics
- Feedback comments generated
- Peer evaluation ratings
- Group-level engagement indicators

The data were then reported and analysed using descriptive statistics, including frequencies, percentages, means, and distributional characteristics. The analysis concentrated on four key dimensions:

1. Student engagement
2. Peer review performance
3. Feedback literacy and Reflective learning
4. Collaborative learning within international cohort

3.2 Findings and Analysis

3.2.1 Student Engagement with FeedbackFruits

Of the 701 postgraduate students enrolled on the module, 397 students successfully completed all required peer evaluation activities, while a further 192 students partially completed the activity, resulting in an overall engagement rate of 84%. This indicates that a substantial proportion of students accessed the platform and participated in the peer review process. Further findings showed that feedback completion reached 62%, while self-assessment completion reached 77%, demonstrating substantial engagement with reflective and evaluative learning activities. Students generated an average of 12 feedback comments per group and spent an average of 164 minutes engaging with review activities (see Table 1 below).

Table 1: Summary of FeedbackFruits Peer Review Engagement Metrics

Engagement Metric	Engagement Data
Students Completing All Tasks	397
Students Partially Completing Tasks	192
Full Engagement Rate	56.6%
Overall Engagement	84.0%
Feedback Completion Progress	62%
Self-Assessment Completion Progress	77%
Average Comments per Group	12
Average Review Duration	164 minutes

The high engagement rates suggest that students were willing to invest substantial time and effort in the peer-evaluation process. The average review duration and number of comments generated indicate meaningful engagement with assessment and feedback literacy.

3.2.2 Peer Review Performance

Students evaluated one another against five criteria: Contribution, Timeliness, Quality of Work, Team Collaboration, and Reflection. Although the example shown in Figure 1 demonstrates maximum ratings across all categories, the average peer-evaluation score across the 397 fully engaged students was 4.728 out of 5.

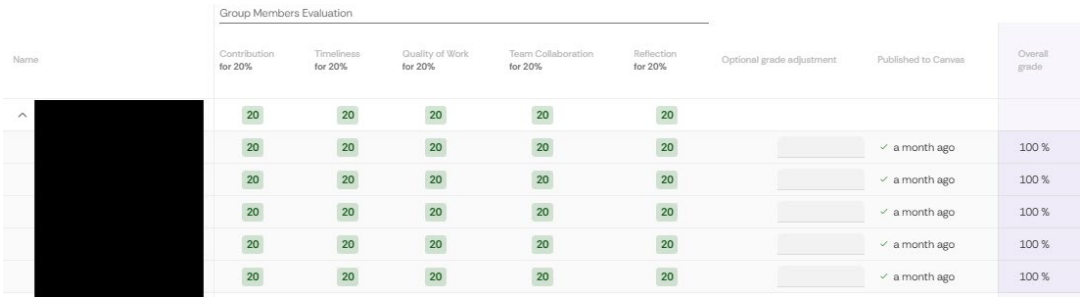


Figure 1*:* FeedbackFruits peer review rubric and grading dashboard

Table 2: Peer evaluation ratings

Peer Review Criteria	Contribution	Timeliness	Quality of Work	Team Collaboration	Reflection
Average grade	4.737	4.614	4.697	4.755	4.740

Table 2 presents the average ratings for each assessment criterion. Ratings were consistently high across all dimensions, with Team Collaboration receiving the highest average score (4.755/5), followed by Reflection (4.740/5) and Contribution (4.737/5). These findings suggest that students generally perceived positive levels of participation, cooperation, and accountability within their groups.

3.2.3 Feedback Literacy and Reflective Learning

Figure 2 below presents anonymised examples of students' given and received peer reviews. Students provided both ratings and qualitative comments against the assessment criteria, while tutors were able to monitor participation and review progress through the dashboard. Receiving peer feedback enabled students to reflect on their performance, identify strengths and areas for improvement, and engage with feedback as an active learning process. Students therefore occupied dual roles as both feedback providers and feedback recipients.

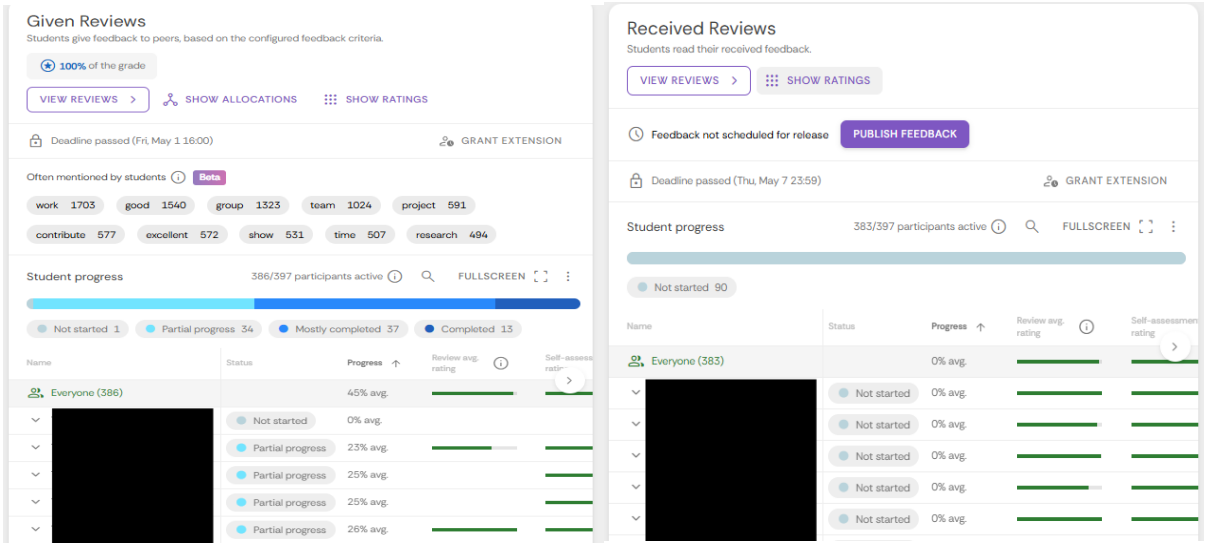


Figure 2: FeedbackFruites 'Given Reviews' and 'Received Reviews' dashboard

The combination of peer review, self-assessment, and qualitative feedback appears to have supported the development of evaluative judgement and reflective practice. The average of 12 comments per group and substantial review duration further suggest that students actively engaged in feedback processes rather than merely completing assessment requirements.

3.2.4 Collaborative Learning Within International Cohort

The findings suggest that FeedbackFruits enhanced collaborative learning and interaction among students. This resembles the research of Li et al. (2025) investigating a different digital tool within which international students

feel safe and secure to collaborate and learn. The strong ratings for Team Collaboration and Reflection indicate that students actively recognised the contributions of others and engaged in collaborative reflection throughout the project. Given the highly international nature of the cohort, the peer evaluation process provided opportunities for intercultural dialogue, communication, and teamwork. Students were required to negotiate expectations, evaluate contributions, and reflect upon team performance, supporting the development of collaborative working practices.

Meanwhile, the findings provide substantial evidence of feedback literacy development. Students were positioned as both feedback providers and recipients, enabling them to actively engage in generating, interpreting, and responding to feedback. The peer evaluation process created multiple opportunities for students to engage in evaluative judgement and develop a deeper understanding of assessment expectations. Through reviewing peers and reflecting on their own performance, students strengthened their ability to make informed judgements about academic work, a skill recognised as essential for lifelong learning.

In addition, the results show that students spent an average of 164 minutes on review activities and generated an average of 12 comments per group. The substantial time investment suggests that students engaged in reflective evaluation rather than simply completing administrative assessment requirements. These findings indicate that FeedbackFruits supported greater engagement with assessment criteria and encouraged students to become active participants in the evaluation process. In all, the findings demonstrate that through the deliberate alignment of technology, pedagogy, and assessment design, FeedbackFruits served as an effective peer assessment platform which encouraged students to participate actively, reflect on their performance, and engage constructively with peers.

4. Discussion

Viewed through the Community of Inquiry, Constructive Alignment, and TPACK frameworks, the study supports growing evidence that educational technologies achieve their greatest impact when embedded within coherent curricular strategies rather than implemented as standalone innovations (Kirkwood and Price, 2014).Summarised from the findings, Table 3 echoes the effectiveness of FeedbackFruits as a structured peer evaluation platforms to offer scalable mechanisms for promoting engagement, feedback literacy, belonging, and student success, and how they are aligned with the theoretical frameworks.

Table 3: Findings, evidence and link to theoretical frameworks

Finding	Evidence	Framework
High engagement	84% participation	Social Presence
Reflective learning	Reflection score 4.74	Cognitive Presence
Structured assessment	Rubrics & guidance	Teaching Presence
Technology-enabled assessment	FeedbackFruits integration	TPACK
Alignment with outcomes	Peer review criteria	Constructive Alignment

The analysis above supports the argument that transformative digital pedagogy emerges through the purposeful interaction of pedagogical design, technology integration, and learner engagement. From a Community of Inquiry (CoI) perspective, the findings provide evidence of cognitive, social, and teaching presence within FeedbackFruits. (Garrison, et al. 2001) Cognitive presence was strengthened through reflective evaluation, feedback generation, and critical engagement with assessment criteria where students were required to analyse peer contributions, compare performance against standards, and provide constructive feedback, encouraging deeper reflection and evaluative judgement. Social presence emerged through collaboration, dialogue, and recognition of peer contributions, particularly within a diverse international cohort. Teaching presence was evident in the structured design of the activity, including the use of clear rubrics, assessment guidance, and scaffolded peer review processes that supported student participation and engagement.

The study also reinforces the principles of Constructive Alignment (CA) (Biggs, et al., 2022) when peer assessment was intentionally linked to module learning outcomes relating to teamwork, communication and reflection. Students were therefore not engaging in isolated technological tasks but participating in activities that directly contributed to the achievement of both disciplinary knowledge and skills around digital marketing. The close alignment between learning outcomes, assessment requirements, and peer evaluation activities helped ensure that engagement with FeedbackFruits remained meaningful and relevant.

Similarly, the results support the Technological Pedagogical Content Knowledge (TPACK) framework (Mishra and Koehler, 2006), which argues that effective technology integration occurs when technological tools are aligned with pedagogical goals and disciplinary content. The findings suggest that FeedbackFruits supported learning not because of the technology itself, but because its functionality was purposefully embedded within a coherent pedagogical and assessment strategy. This reinforces the view that educational technologies achieve their greatest impact when used to enhance learning design rather than simply digitise existing practices.

The study also contributes to the growing literature on feedback literacy. Students were positioned as both feedback providers and recipients, requiring them to actively generate, interpret, and respond to feedback. Through reviewing peers and reflecting on their own performance, students developed evaluative judgement and a deeper understanding of quality standards and assessment expectations. These findings align with previous research suggesting that feedback is most effective when learners actively participate in feedback processes rather than passively receiving comments from educators (Carless and Winstone, 2023).

The findings also demonstrate the value of structured peer evaluation within large cohorts (Topping, 2017). FeedbackFruits provided a scalable mechanism for promoting accountability, transparency, and meaningful interaction among more than 700 students. However, the findings also identify opportunities for future enhancement. While overall engagement reached 84%, only 56.6% of students completed all peer evaluation activities. This suggests that although most students accessed and engaged with the platform, not all students completed every stage of the review process. Furthermore, feedback completion (62%) was lower than self-assessment completion (77%), indicating that students may have been more comfortable reflecting on their own performance than providing constructive feedback to peers. Future iterations of the module could explore additional strategies to strengthen participation, such as providing further guidance on feedback practices, integrating formative peer-review opportunities earlier in the semester, and offering exemplars of effective feedback comments.

Finally, similar to Li, et al. (2025) identified that the international nature of the cohort provides important insights into constructive and collaborative learning through digital technology tools. Students from diverse cultural, linguistic, and educational backgrounds participated successfully in collaborative activities supported by transparent criteria and structured evaluation processes. The findings therefore suggest that technology-enhanced peer assessment can contribute to intercultural engagement, community building, and a stronger sense of belonging within diverse higher education environments.

Overall, the study demonstrates that FeedbackFruits can support engagement, feedback literacy, accountability, collaboration, and reflective practice when embedded within a pedagogically aligned assessment strategy. As higher education institutions continue to expand digitally enhanced provision, structured peer evaluation platforms like FeedbackFruits can offer scalable approaches for fostering student success and creating more inclusive and participatory learning environments.

5. Conclusion

This study examined the implementation of FeedbackFruits within a large postgraduate marketing module at a UK higher education institution to investigate its contribution to student engagement, feedback literacy, accountability, reflective learning, and collaborative participation. Drawing on the frameworks of Community of Inquiry (CoI), Constructive Alignment (CA), and Technological Pedagogical Content Knowledge (TPACK), the findings demonstrate that FeedbackFruits can serve as more than a digital peer-assessment platform when embedded within a coherent pedagogical design. Rather than technology alone driving learning outcomes, its educational value emerges through the purposeful alignment of technology, pedagogy, curriculum, assessment, and meaningful student interaction.

The findings indicate that FeedbackFruits supported student engagement, reflection, collaboration, and feedback literacy through a structured peer-evaluation process aligned with intended learning outcomes. Importantly, the study suggests that technology-enhanced peer assessment provides a scalable and pedagogically robust approach for fostering accountability, engagement, and a sense of belonging within large and culturally diverse higher education classrooms. These findings reinforce the importance of designing technology-enhanced learning experiences that prioritise pedagogical intent over technological functionality.

The study also makes a theoretical contribution by demonstrating the complementary value of integrating CoI, CA, and TPACK as a multi-framework lens for evaluating technology-enhanced peer assessment. While each framework offers distinct insights into collaborative learning, curriculum alignment, and technology integration,

their combined application provides a more comprehensive understanding of how digital technologies can support meaningful, inclusive, and transformative learning experiences. This integrated perspective extends existing research, which has largely focused on student satisfaction and platform usability, by explaining the pedagogical mechanisms through which educational technologies create value.

However, several limitations should be acknowledged. The research was conducted within a single postgraduate marketing module at one UK institution and relied primarily on learning analytics and peer-evaluation data. Although the cohort was highly international, students' prior levels of digital literacy were not measured, making it difficult to determine how digital competence may have influenced engagement with FeedbackFruits or peer-assessment behaviours. Future research should therefore incorporate qualitative methods to explore students' lived experiences, examine implementation across different disciplines and institutional contexts, investigate the influence of digital literacy on technology-enhanced assessment, and evaluate the longer-term development of feedback literacy and collaborative learning.

Ultimately, this study argues that the transformative potential of educational technology resides not in the technology itself, but in the quality of its pedagogical integration. By positioning FeedbackFruits within an integrated CoI-CA-TPACK framework, the study provides both a theoretical foundation and practical guidance for educators seeking to design technology-enhanced peer assessment that is pedagogically coherent, scalable, and capable of supporting inclusive and engaging learning in increasingly diverse higher education environments.

Ethics Declaration

Formal ethic approval was not required as this study involved the analysis of anonymised data automatically generated by University-licensed digital technology platform, i.e. FeedbackFruits suited within the University's Virtual Learning Environment and collected as part of the module performance evaluation process.

AI Declaration

ChatGPT was used to support the exploration of theoretical frameworks and proofreading during the preparation of this manuscript. All substantive academic content, analysis, interpretation of findings, and final editorial decisions were undertaken by the authors.

References

Akyol, Z. and Garrison, D.R. (2011) 'Understanding cognitive presence in an online and blended community of inquiry', Distance Education, 32(2), pp. 233–250.

Biggs, J. (1996) 'Enhancing teaching through constructive alignment', Higher Education, 32(3), pp. 347–364.

Biggs, J., Tang, C. and Kennedy, G. (2022) *Teaching for Quality Learning at University*. 5th edn. Maidenhead: Open University Press.

Campbell, C., Borer, L., & McCarthy, S. (2020). Enhancing feedback with FeedbackFruits: Scaling for transformative implementation. *ASCILITE Publications*, 322-327.

Carless, D. and Winstone, N.E. (2023) *Teacher Feedback Literacy and Student Feedback Literacy in Higher Education*. London: Routledge.

Chai, C.S., Koh, J.H.L. and Tsai, C.C. (2013) 'A review of technological pedagogical content knowledge', Educational Technology & Society, 16(2), pp. 31–51.

Cleveland-Innes, M. and Campbell, P. (2012) "Emotional presence, learning, and the online learning environment", *International Review of Research in Open and Distance Learning*, 13(4), pp. 269–292

FeedbackFruits (2025) "FeebackFruits" [online], https://feedbackfruits.com/

Fullan, M. and Langworthy, M. (2014) *A rich seam: How new pedagogies find deep learning,* Pearson, London

Fiock, H.S. (2020) 'Designing a community of inquiry in online courses', International Review of Research in Open and Distance Learning, 21(1), pp. 135–153.

Garrison, D.R., Anderson, T. and Archer, W. (2000) "Critical inquiry in a text-based environment", *The Internet and Higher Education*, 2(2–3), pp. 87–105.

Garrison, D. R., Anderson, T. and Archer, W. (2010) "The first decade of the community of inquiry framework: a retrospective". The Internet and Higher Education, 13 (1-2), pp. 5-9

Iglesias Pérez, M. C., Vidal-Puga, J., & Pino Juste, M. R. (2022). The role of self and peer assessment in higher education. *Studies in Higher Education*, *47*(3), 683-692.

Kirkwood, A. and Price, L. (2014) "Technology-enhanced learning and teaching in higher education", *Learning, Media and Technology*, 39(1), pp. 6–36.

Koehler, M.J., Mishra, P. and Cain, W. (2013) "What is technological pedagogical content knowledge (TPACK)?", *Journal of Education*, 193(3), pp. 13–19.

Li, C.; Du, X. and Xiao, P. (2025) "Revolutionising Education: Leveraging AI to Boost Student Engagement through Constructivist and Social Collaborative Learning-A Study of Perusall" *UK Academy for Information Systems Conference Proceeding 2025* Spring 7-10-2025 Association for Information Systems

Mittelmeier, J. and Rienties, B.(2024). "The Community of Inquiry in international and intercultural settings", In Cleveland-Innes, M. F., Stenbom, S., Garrison, D.R. *The Design of Digital Learning Environments*, pp. 152-171, Routledge.

Myklebust, A. (2025). 'Improving peer assessment through methodological changes: a study of student feedback and participation', In *ICERI2025 Proceedings* (pp. 4082-4091).

Mishra, P. and Koehler, M.J. (2006) "Technological pedagogical content knowledge", *Teachers College Record*, 108(6), pp. 1017–1054.

Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L. and Koole, M. (2021) "Balancing technology, pedagogy and the new normal: Post-pandemic challenges for higher education". *Postdigital Science and Education*, 3(3), pp.715-742.

Shea, P. and Bidjerano, T. (2009) "Community of inquiry as a theoretical framework", *Computers & Education*, 52(3), pp. 543–553.

Stephenson M and Torn A (2023), "Review, rapid recall and reposition: how one HEI adapted delivery and design in the digital world in response to COVID-19". *Higher Education, Skills and Work-based Learning*, Vol. 13 No. 4 pp. 834–845

Tai, J., Ajjawi, R., Boud, D., Dawson, P. and Panadero, E. (2018) "Developing evaluative judgement: Enabling students to make decisions about the quality of work", *Higher Education*, 76(3), pp. 467–481.

Topping, K.J. (2017) "Peer assessment: Learning by judging and discussing the work of other learners", *Interdisciplinary Education and Psychology*, 1(1), pp. 1–17.

Vaz-Serra, P., Fleckney, P., Thompson, J. and Hui, F.K.P (2023) Improving Student Engagement in Learning and Assessment Activities Using Peer-To-Peer Feedback. In *Australasian Universities Building Education Association Conference* (pp. 347-360). Singapore: Springer Nature Singapore.

Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods*. 6th edn. Thousand Oaks, CA: Sage.

Beyond Audio: An Empirical Assessment of AI-Generated Podcasts and Videocasts in Business Education

Gökhan Eğilmez and Karolina Schneider
Plaster College of Business and Entrepreneurship, Lindenwood University, USA

Gegilmez@Lindenwood.Edu
KSchneider@lindenwood.edu

Abstract: Generative artificial intelligence (AI) is opening new possibilities for personalized, engaging instructional content in higher education. Extending our earlier work on AI-generated podcasts, this study examines the addition of AI-generated videocasts (video overviews) alongside podcasts in an undergraduate business analytics course, asking whether a multimodal (audio + video) approach improves students' learning experience beyond audio alone in a subject area that is conceptually dense and data-intensive. The study takes a quasi-experimental approach, comparing two consecutive sections of the same course: one with access only to AI-generated audio podcasts (Fall 2025, n = 24), and one where the same podcasts were supplemented with companion AI-generated videocasts (Spring 2026, n = 20). Student-perceived academic performance, engagement, and satisfaction were measured through a 15-item survey covering prior podcast experience, AI familiarity, usage patterns, perceived impact on quiz preparation and comprehension, effectiveness relative to traditional study methods, and open-ended feedback. Results showed a consistent, though modestly powered, advantage for the multimodal condition: students in the audio + video section rated the materials more favorably than the audio-only section on nearly every measured dimension, with a statistically significant increase in willingness to recommend the format for other courses (84.2% vs. 47.6% favorable, p = .026). Qualitative responses reinforced this pattern, with audio + video students more often citing improved comprehension and valuing multiple ways to engage with material, while a smaller subset preferred less synthetic-sounding narration. Across both sections, most students rated podcasts as at least as effective as traditional study methods, and roughly half had never used podcasts as a learning tool before the course. These findings offer preliminary but encouraging evidence that integrating AI generated video and audio content strengthens students' learning experience and engagement in business analytics education, while underscoring the need for larger, multi-term samples and objective performance measures to confirm the effect. The study contributes to the emerging literature on AI-enhanced instructional design by identifying student-reported benefits and design considerations relevant to integrating generative AI tools into business education curricula.

Keywords: Business Analytics; AI-Generated Podcasts; AI-Generated Videocasts; Student Academic Performance

1. Introduction

The next generation of business professionals will operate in an environment where data-informed decision-making is not a specialized skill but a baseline expectation. Roles across marketing, operations, finance, and strategy increasingly require employees to develop and interpret dashboards, evaluate statistical output, and translate quantitative findings into actionable business recommendations. Business analytics education is therefore no longer a niche elective but a foundational component of undergraduate business curricula, tasked with preparing students to function competently in data-saturated workplaces regardless of their eventual specialization. How well this education succeeds has direct consequences for the analytical readiness of the future workforce. Yet succeeding at this task requires more than exposing students to analytics tools; it requires building a durable, working knowledge of concepts that are often abstract by nature e.g. probability distributions, statistical inference, optimization logic, and model assumptions. These topics often require a deeper understanding to build up analytics skills to be ready as a workforce. A student can memorize a formula or follow a worked example without internalizing why a method applies, when it fails, or how to reason about its output in an unfamiliar context. This gap between procedural familiarity and conceptual fluency is a persistent challenge in analytics education, and it is precisely the gap that instructional design must address if students are to apply these tools flexibly after the course ends while they stay engaged with the course material.

Podcasts have emerged as one promising supplementary format for closing this gap. As an instructional medium, podcasts allow students to engage with course material outside the constraints of the traditional classroom such as during commutes, while exercising, or in other moments otherwise unavailable for study. This flexibility has been associated with increased engagement and more consistent exposure to material, and podcasts have increasingly been adopted across higher education as a low-friction supplement to lectures and readings, valued for extending learning time without adding to students' scheduled course load. Rather than replacing core instruction, podcasts function best as reinforcement: a second exposure to material in a different cognitive mode, which can aid retention and give students a chance to revisit difficult concepts at their own pace.

Podcasts used for instructional purposes could vary considerably in format. Some can be constructed as narrated summaries that closely mirror a lecture content; others can take a conversational, dialogue-driven approach that models the reasoning process itself, walking through how a problem is framed and solved rather than simply stating conclusions. The secondary dialogue-driven format may be particularly well suited to abstract analytics topics, since hearing a concept explained, questioned, and clarified in conversational form can make statistical reasoning feel less like a fixed procedure and one that students can imagine applying themselves. In short, podcasts have the potential to render abstract quantitative reasoning more accessible through an accessible and repeatable audio format. And recently AI-generated podcasts have been a popular medium across various higher education fields.

Generative AI substantially changes the economics and flexibility of producing various contents including podcasts. Traditionally, high-quality instructional podcasts require significant time investment from instructors or instructional designers, along with recording, editing, and production expertise that many faculty members simply do not have the bandwidth to sustain across an entire course, let alone across multiple sections or semesters. AI-generated podcasts remove much of this barrier: they can be produced quickly, at low marginal cost, and updated easily whenever course content, datasets, or examples change, without requiring a new recording session. This also enables a level of customization that is difficult to achieve with traditionally produced podcasts, where content can be tailored to a specific week's material, aligned tightly to an instructor's syllabus and terminology, or adjusted in tone and pacing for a particular student population, all with far less overhead than conventional production. Because they are inexpensive to generate, instructors can also reiterate: producing multiple versions of an explanation, testing different framings of the same abstract concept, or supplementing a single lecture with several short, targeted episodes rather than committing one polished but static artifact. It is this combination of pedagogical potential and practical scalability that motivates the present study's investigation of AI-generated podcasts as a supplementary tool in business analytics education.

2. Literature Review

2.1 Landscape and Adoption of Podcasts in Business and Management Education

Podcasting has been adopted as a learning resource in MBA and business courses, with studies reporting positive attitudes, perceived learning gains, and perceived flexibility from student participants. For example, Carmi (MBA course podcasting in online information review) reports that MBA students perceived podcasts as offering learning flexibility and examined learning effectiveness, with differences observed between mid-term assignments and full lectures, though the study notes variability in attitudes based on prior experience with podcasts (Carmi, 2023). These findings are echoed by additional business-education-oriented reviews that highlight learner flexibility, asynchronous access, and enhanced engagement as core benefits of podcast use in business curricula (Kurtböke et al., 2023; Carmi, 2023). Moreover, student and faculty perceptions of podcasts as a medium in professional education broader than business, including medical or leadership domains, underscoring perceived benefits such as accessibility, convenience, and the potential for learner-centered pedagogy that leverages learner-generated or instructor-generated audio (Mason et al., 2024; Devies et al., 2025). These studies collectively position podcasts as a viable supplement for traditional lectures in business education, especially in settings requiring asynchronous and flexible delivery.

2.2 Traditional vs. AI-Generated Podcasts in Business Education

Traditional podcasts (hosted, interviewer-led, or student-produced) appear in business and management education literature as a means to deliver content, support reflective practice, and foster engagement. Literature emphasizes practical benefits such as convenience, portability, and the ability to revisit content for reinforcement, which aligns with the needs of working professionals in MBA programs seeking flexible learning modalities (Carmi, 2023; Mason et al., 2024). Furthermore, AI-generated podcasts are an emerging alternative. Early investigations exploring generative AI podcasts for learning in clinical and biomedical education indicate a movement toward AI-assisted podcast production, and positive perceptions of AI-created audio within health-professional education. In business education, the recent literature has been focused on the potential of AI-enabled podcasting to enable instructors to tailor content to learning objectives and experimenting with new formats such as AI-created summaries or content generation for case discussions. For instance, (Graham & Stevens, 2026). Although not business-specific, these findings signal analogous opportunities and cautions for business programs experimenting with AI-generated audio for case discussions, guest interviews, or module recaps.

Protocols and pilot studies have begun to compare AI-generated podcast formats with traditional text-based or human-delivered content in professional education, with anticipated investigations into learning outcomes, cognitive load, and motivation (Stadler et al., 2025). They propose a protocol to compare AI-generated podcast summaries to traditional reading for medical articles, a design that offers a transferable blueprint for business education researchers seeking to compare AI-generated audio against standard learning materials in business case studies or managerial theory modules (Stadler et al., 2025). While focused on medical literature, the protocol contributes methodological guidance for evaluating AI podcasts in business education contexts.

2.3 AI-Enabled Podcasting Modalities and the Implications for Pedagogy

2.3.1 AI-Enabled Podcasting Encompasses Two Primary Modalities in Education

(i) AI-generated podcasts created by language models or text-to-speech systems to deliver instructor-selected content, and (ii) AI-assisted workflows that help produce or curate podcast content (e.g., transcripts, summaries, or scripts). The literature notes that AI podcasts can increase efficiency, broaden access, and support personalized or just-in-time learning, while also raising concerns about accuracy, trustworthiness, and the need for alignment with assessment tasks. In nursing and neurology education, studies report that students perceive AI-generated podcasts as convenient and engaging but highlight quality and trust concerns, suggesting careful integration with explicit quality controls, sources, and alignment with learning objectives (Graham & Stevens, 2026; (Mason et al., 2024; Albin et al., 2026). While these studies are not business-specific, their findings offer cautionary guidance for business educators piloting AI podcasts for executive education, entrepreneurship, or leadership modules.

In leadership education, sound pedagogical frameworks are described to guide the use of podcasts (including AI-enabled formats) to augment classroom learning or deliver leadership insights. Middleton's synthesis on studio-learning paradigms and the disruptive potential of audio-based approaches argues for considering learner agency, co-production, and the co-design of podcast content with students, including the use of learner-generated podcasts to deepen learning and reflection (Middleton, n.d.; . This suggests a productive pathway for business programs to implement AI-assisted or student-generated podcasts in capstone projects, leadership simulations, or entrepreneurship seminars.

Egilmez and Schneider (2025) recently investigated whether AI-generated, topic-specific podcasts could improve learning outcomes and student engagement in an undergraduate Business Analytics course. A quasi-experimental design compared an intervention section that received AI-generated podcast supplements with a control section using traditional instructional materials. Quantitative analyses were complemented with surveys and focus group interviews. The study concludes that AI-generated educational podcasts can positively influence both academic performance and student engagement when used as supplemental instructional materials. The findings suggest that generative AI can support more accessible, flexible, and personalized learning experiences while reducing content development effort for instructors. The authors recommend further research involving larger samples, longer intervention periods, different disciplines, and additional learning outcomes such as long-term retention and student motivation.

2.4 Practical and Methodological Considerations for Using Podcasts in Business Education

2.4.1 Pedagogical Design

The literature emphasizes choosing podcast formats aligned to intended learning outcomes, including instructor-led, learner-produced, or hybrid formats. For business education, this translates into thoughtful decisions about when to substitute traditional lectures with podcast segments, when to use podcasts as pre-briefs or post-briefs in case discussions, and when to employ learner-produced podcasts to demonstrate applied business concepts (e.g., strategy pitches, market analyses, or leadership reflections) (Carmi, 2023).

2.4.2 Assessment and Evaluation

Several studies advocate rigorous evaluation of podcast-based learning using validated instruments when possible, noting that much of the existing evidence relies on descriptive analytics or qualitative feedback. In business education, researchers should incorporate pre/post assessments, knowledge checks related to case content, and measures of critical thinking or reflective practice to determine the educational value of podcast interventions (Carmi, 2023; Mason et al., 2024; Stadler et al., 2025).

2.4.3 Quality Control and Ethics

Given concerns about the quality and trustworthiness of AI-generated content, business educators should implement clear sourcing, attribution, and quality assurance processes. The neurology and medical education literature recommends transparency about AI-assisted outputs and careful integration with evidence-based content, which is directly applicable to business education where case content and managerial concepts must be accurately sourced and cited (Mason et al., 2024; Graham & Stevens, 2026).

2.5 Summary of the Literature Review

2.5.1 Traditional Podcast Use in Business Education

Podcasts in MBA courses are associated with perceived learning gains and flexible learning, with students recognizing benefits of asynchronous access and the ability to revisit material (Carmi, 2023). Business education literature also reports learner satisfaction and perceived usefulness of podcasting as part of blended or online approaches, including the potential for learner-generated podcasts to foster engagement (Carmi, 2023; Mason et al., 2024).

2.5.2 AI-Generated Podcasting in Business Education

Studies in related health-professional education report that AI-generated podcasts can be convenient and engaging but raise concerns about quality and trust; these insights inform business educators about potential benefits and risks when adopting AI-generated formats for course content or summaries (Graham & Stevens, 2026; Mason et al., 2024). Protocol-level research suggests formal experimental designs to compare AI-generated podcast summaries with conventional reading, which can guide business educators in evaluating AI podcast interventions with rigorous methods (Stadler et al., 2025).

2.5.3 Pedagogical and Methodological Guidance

Analyses of learner-generated podcasting and the broader podcast literature in higher education emphasize the importance of designing learning outcomes, including learner agency, co-design, and the potential for podcasts to support active and reflective learning in leadership and entrepreneurship contexts (Middleton, n.d.; Devies et al., 2025). Reviews of podcasting in other professional education domains (e.g., medical, neurology, radiation sciences) underscore the need to integrate podcasts thoughtfully with curricula, to manage content quality, and to consider cognitive load and accessibility as part of instructional design, offering transferable considerations for business education (Mason et al., 2024; Shah et al., 2025; Arruzza et al., 2026).

3. Methods

3.1 Study Design

This study was conducted in an undergraduate Business Analytics course that incorporates both knowledge-based and application-based assessments. Knowledge-oriented evaluations included quizzes, midterm examinations, and final examinations featuring multiple-choice, true/false, and essay-format questions. Application-oriented assessments consisted of weekly homework assignments and a semester-long project requiring students to apply business analytics techniques in practical contexts.

The course is offered in both online asynchronous and face-to-face formats. Historically, students enrolled in the online asynchronous sections have demonstrated lower levels of engagement, likely due to the absence of live instruction and the challenging nature of topics such as optimization, simulation, and statistical analysis. To enhance student engagement and support academic success, AI-generated learning resources were introduced. Specifically, students were provided with AI-generated audio podcasts and videocasts, both created using NotebookLM. Quiz assessments served as the primary focus of this investigation. Each quiz consisted of approximately 10–15 questions randomly selected from a larger question bank (e.g. 20-25 questions).

The primary aim of the study was to examine the impact of AI-generated podcasts and videocasts on student academic performance and engagement within an online learning environment. A mixed method research design is employed. Quantitative analyses will compare quiz performance among students who reported using AI-generated audio only podcasts and both audio podcasts and videocasts. To facilitate this comparison,

students are asked on each quiz whether they had listened to the podcast, watched the videocast. Responses will be collected without any grade incentives and are subsequently used for statistical analysis.

To complement the quantitative findings, a post-course survey is administered through Qualtrics during Week 7 of the semester, and voluntary student interviews will be conducted. These qualitative data explored students' perceptions of the AI-generated learning materials, usage patterns, engagement levels, and recommendations for future improvements. By integrating quantitative and qualitative evidence, the study sought to better understand both the magnitude of the intervention's effects and the underlying mechanisms through which those effects occurred.

3.2 AI Podcast and Videocast Generation

Although several platforms currently support the creation of AI-generated educational media, this study utilizes Google NotebookLM, a Gemini 2.0-powered research and writing assistant. NotebookLM is selected because of its user-friendly interface, availability of a free version, and ability to rapidly generate high-quality learning content.

As illustrated in Figure 1, the process begins by accessing NotebookLM and signing in with a Google account. Users then create a new notebook and upload source materials, which may include PDFs, webpages, YouTube videos, audio recordings, Google Docs, text documents, and Google Slides. For this study, the complete pool of quiz questions and answers, together with detailed chapter presentation slides, was uploaded as the primary source material used to generate the learning resources.

Once the materials were uploaded, users could interact with the content through NotebookLM's chat interface to ask questions, generate summaries, and explore key concepts. Alternatively, they could access NotebookLM Studio, which provides a variety of AI-generated outputs, including audio podcasts, videocasts (video overviews), interactive or non-interactive discussions, mind maps, reports, flashcards, and practice quizzes.

Using these capabilities, both AI-generated audio podcasts and AI-generated videocasts are created to supplement student learning. The generated resources synthesize course content into accessible and engaging formats designed to reinforce key concepts and improve comprehension. The final step involves saving and distributing podcasts, videocasts, and supporting study materials to students, thereby providing an integrated and interactive learning experience that could be accessed outside of the traditional classroom environment. Lastly, the text transcriptions of the podcasts are also made available for ADA-compliance purposes.

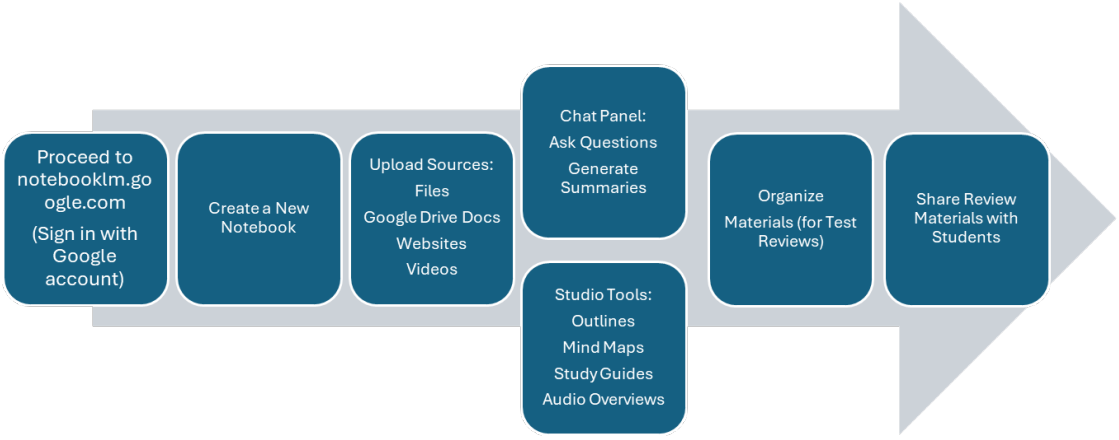


Figure 1: Steps to develop an AI Generated Podcast (Egilmez and Schneider, 2025)

3.3 Assessment: Empirical Survey

The survey was a single-administration, self-report questionnaire delivered via Qualtrics to students in two sections of Business Analytics I (Fall 2025 and Spring 2026), designed to assess how AI-generated podcasts (and, in Spring 2026, companion videocasts) affected academic performance and course engagement. It opened with an IRB-approved consent/information sheet (Lindenwood University) and took an estimated 10–15 minutes to complete. The instrument contained 15 substantive items, organized into four blocks:

1. Respondent background (Q1–Q2, Q5–Q7) - major, year in program, race/ethnicity (select all that apply), gender identity, and age group, used to characterize the sample rather than as study variables.
2. Prior experience and usage patterns (Q3, Q4, Q8, Q9) - whether students had used podcasts as a learning tool before this course, self-rated familiarity with AI-generated content (5-point scale), how often they listened to the podcasts before weekly quizzes (5-point frequency scale), and average weekly listening time (banded minute ranges).
3. Perceived impact (Q10–Q12) - two 5-point Likert agreement matrices (Strongly disagree–Strongly agree): a 4-item academic-performance block (felt more prepared, performed better on quizzes, understood concepts better, retained key concepts) and a 3-item engagement block (material more engaging, more motivated to study, would recommend in other courses), plus a single comparative item (Q11) rating podcast effectiveness relative to traditional study methods (5-point scale from "Much less effective" to "Much more effective").
4. Open-ended feedback (Q13–Q15) - three free-text questions on what students liked most, what could be improved, and suggestions for integrating the podcasts more effectively into the course.

4. Results

This supplementary analysis directly compares the Fall 2025 section, where students had access to AI-generated audio podcasts only, to the Spring 2026 section, where the same audio podcasts were supplemented with companion videocasts. Thus, Fall 2025 is the control and Spring 2026 is the intervention (treatment) group. The goal is to see whether adding a video component changed students' experience and perceived value relative to audio alone.

Because responses to the 5-point agreement and frequency items are ordinal rather than continuous, group differences were tested with the Mann-Whitney U test (two-tailed, $\alpha = .05$) rather than a t-test. Effect sizes are reported as the rank-biserial correlation (r), where values around .10 are considered small, .30 medium, and .50 large. Open-ended responses (Q13–Q15) were re-coded by semester using the same thematic categories as the main report, so the theme counts here are broken out by section rather than pooled.

Caveat: with only 19–22 valid responses per semester, this comparison has limited statistical power - real differences of a moderate size may not reach statistical significance. Results should be read as suggestive evidence, not a definitive causal test of "audio-only vs. audio + video."

4.1 Quantitative Comparison: Fall 2025 vs. Spring 2026

Mean scores and significance tests for every rated item, ordered as they appear in the survey, and depicted in Table 1. Additionally, a visual comparison of Fall 25 and Sprin 26 sections were provided in Figure 2. The interpretation of the quantitative findings are provided in the upcoming sub-sections referencing Table 1 and Figure 2.

Table 1: Mean Scores and Significance Tests

Item	Fall n	Fall mean	Spring n	Spring mean	p-value	Effect size (r)	Higher group	Significant (p<.05)
AI-content familiarity (Q4)	22	2.86	20	3.3	0.156	0.245	Spring 2026	No
Listening frequency (Q8)	22	3.5	20	3.5	0.948	0.014	Tie	No
Time spent/week (Q9)	21	2.05	20	1.75	0.402	0.143	Fall 2025	No
Felt prepared for quizzes (Q10.1)	21	3.57	19	4	0.273	0.19	Spring 2026	No
Performed better on quizzes (Q10.2)	21	3.57	19	4.05	0.27	0.195	Spring 2026	No
Understood concepts better (Q10.3)	21	3.52	19	3.95	0.271	0.195	Spring 2026	No

Item	Fall n	Fall mean	Spring n	Spring mean	p-value	Effect size (r)	Higher group	Significant (p<.05)
Retained key concepts (Q10.4)	21	3.71	19	3.84	0.836	0.038	Spring 2026	No
More effective than trad. methods (Q11)	20	3.65	19	3.68	0.893	0.026	Spring 2026	No
Material more engaging (Q12.1)	21	3.76	19	4	0.523	0.113	Spring 2026	No
More motivated to study (Q12.2)	21	3.43	19	3.37	0.864	0.033	Fall 2025	No
Would recommend in other courses (Q12.3)	21	3.43	19	4.26	0.026	0.398	Spring 2026	Yes

p-values from two-tailed Mann-Whitney U tests. Effect size (r) is the rank-biserial correlation. Only "Would recommend in other courses" (Q12.3) reached conventional statistical significance (p = .026), with Spring 2026 rated higher.

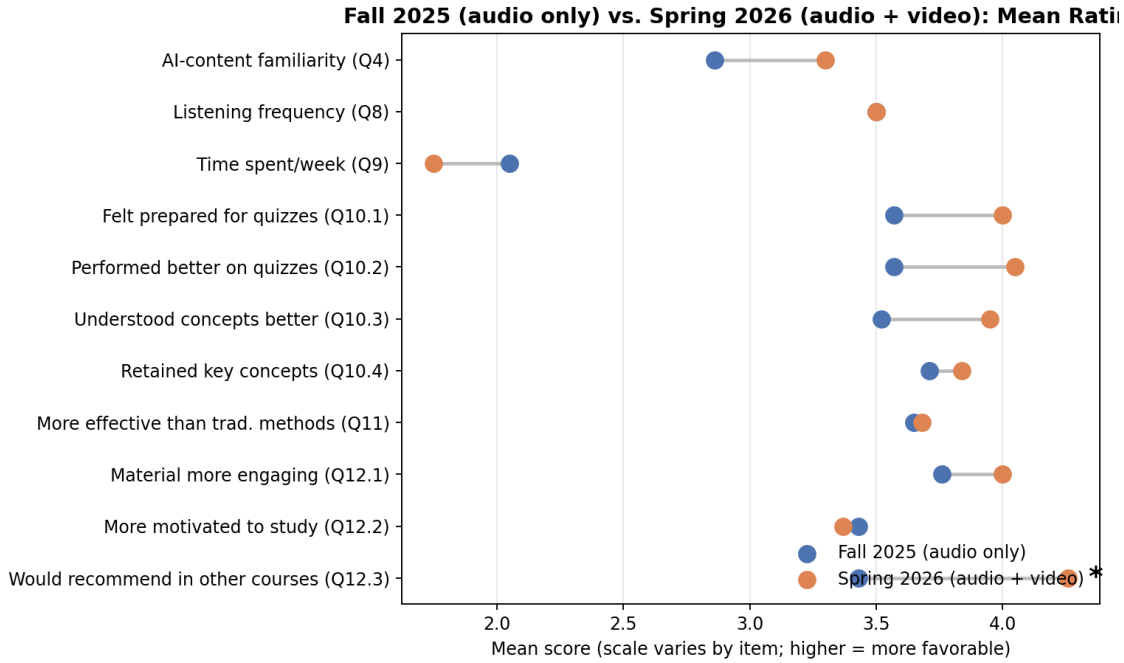


Figure 2: Dumbbell chart of mean ratings by semester. (An asterisk (*) marks the one statistically significant difference (Q12.3))

4.1.1 Familiarity and Usage Patterns (Q4, Q8, Q9)

Self-reported familiarity with AI-generated content was somewhat higher in Spring 2026 (mean 3.30 vs. 2.86), though not statistically significant (p = .156). This may reflect growing general familiarity with AI tools over time rather than the effect of the video addition. Listening frequency was essentially identical between semesters (mean 3.50 in both, p = .948). Time spent per week was slightly lower in Spring 2026 (mean 1.75 vs. 2.05, p = .402, not significant) consistent with a videocast potentially delivering the same content more efficiently, though the difference is not statistically reliable.

4.1.2 Academic Performance Impact (Q10)

All four academic-performance items were rated more favorably in Spring 2026 than Fall 2025 (e.g., "felt more prepared for quizzes": 78.9% top-2-box in Spring vs. 61.9% in Fall), but none of these differences reached statistical significance at this sample size (all p > .27). Moreover, the most consistent gap was on "understood

chapter concepts/terms better" (73.7% vs. 57.1% top-2-box) and "performed better on quizzes" (78.9% vs. 61.9%) — both point in the direction of the combined audio + video format helping comprehension, though this is a directional trend, not a confirmed effect.

4.1.3 Effectiveness vs. Traditional Study Methods (Q11)

63.2% of Spring 2026 respondents rated the format as more effective than traditional study methods, versus 55.0% in Fall 2025, which indicates a modest, non-significant difference (p = .893).

4.1.4 Course Engagement (Q12)

"Made course material more engaging" and "would recommend in other courses" were both rated higher in Spring 2026. "Would recommend using AI-generated podcasts/videocasts in other courses" was the one statistically significant difference in the entire survey: 84.2% of Spring 2026 respondents gave a top-2-box rating vs. 47.6% in Fall 2025 (p = .026, medium-to-large effect r = .40). This is the clearest quantitative signal that the audio + video combination was more well-received than audio alone. "Felt more motivated to study" was the only item rated (marginally) higher in Fall 2025 (mean 3.43 vs. 3.37), though the difference is negligible and not significant (p = .864).

4.2 Qualitative Comparison (Open-Ended Feedback)

Open-ended responses were re-coded by semester using the same thematic categories as the main report. Percentages are the share of each semester's respondents who mentioned a theme (respondents could mention more than one). Question 13 (See Figure 3) investigated what students liked the most and Question 14 asked for qualitative feedback on what to improve (See Figure 4). And, lastly, students were asked if they would like this concept to be integrated to other parts of the course beyond quizzes (See Figure 5).

Better understanding/retention" was mentioned far more often in Spring 2026 (53.8% vs. 25.0% of Fall respondents), and Spring 2026 was the only semester where students spontaneously mentioned the video component (15.4%) as a reason they liked the format - e.g., one student wrote the videos let them "hear what I have learned in a different way," and another cited having "four different ways to take in the information.

As seen in Figure 3, most of the Spring 2026 respondents (53.8%) said nothing needed improvement, compared to 20.0% in Fall 2025 - consistent with higher overall satisfaction once video was added. "Reduce AI-generated feel" appeared only in Spring 2026 (15.4%), suggesting the video format made the AI-generated quality more noticeable to some students even as overall satisfaction rose. As seen in Figure 4, Spring 2026 students were more likely to ask for greater emphasis or promotion of the podcasts/videocasts in the course (16.7% vs. 0%), a further indirect sign of stronger buy-in for the combined format.

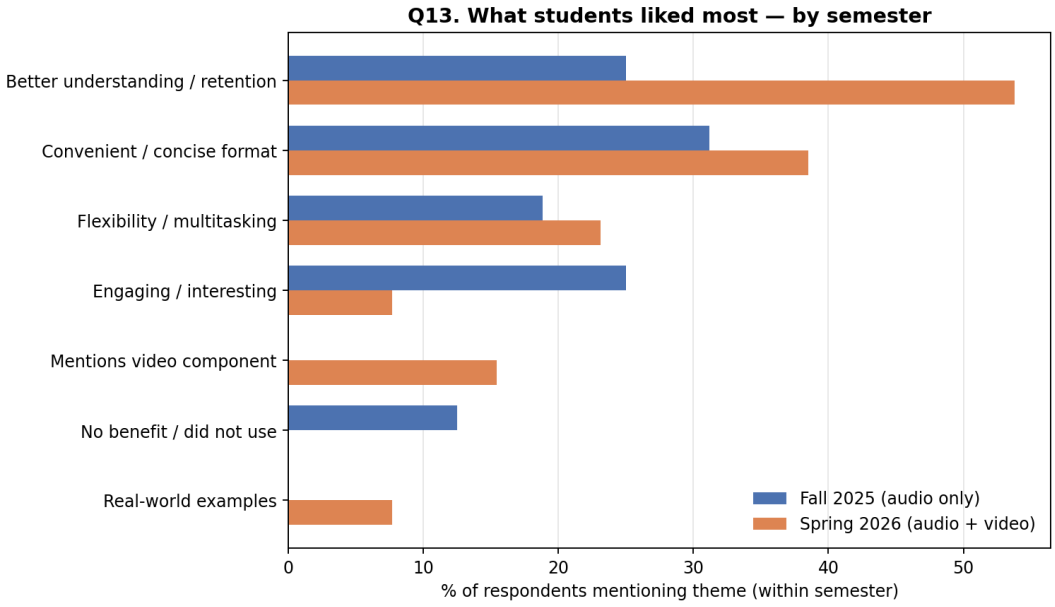


Figure 3: Q13: What did you like most?

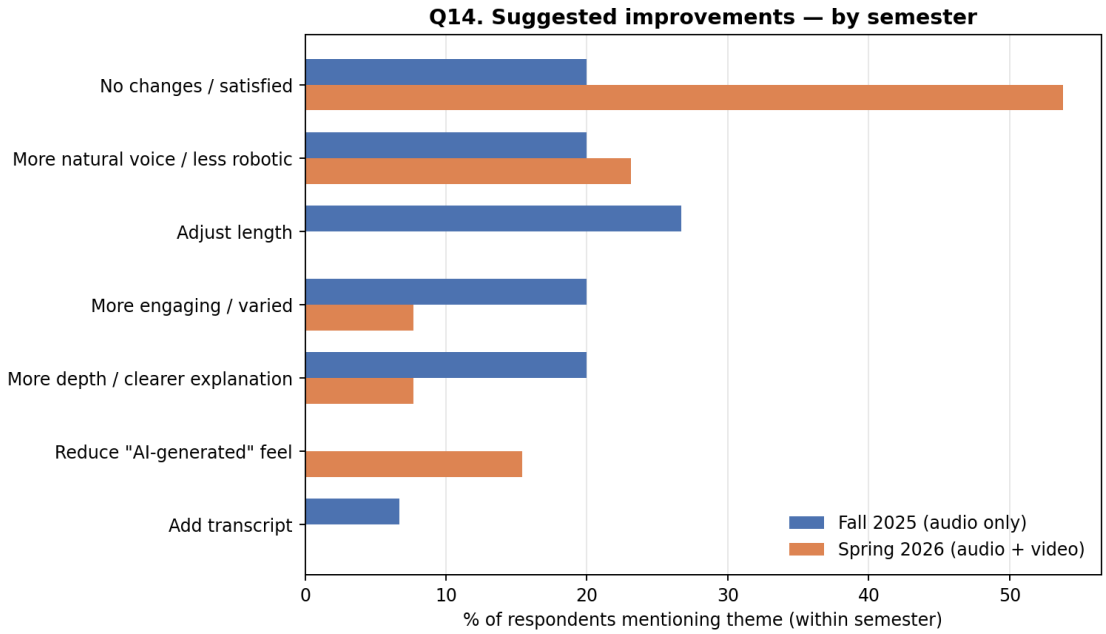


Figure 4: Q14. What could be improved?

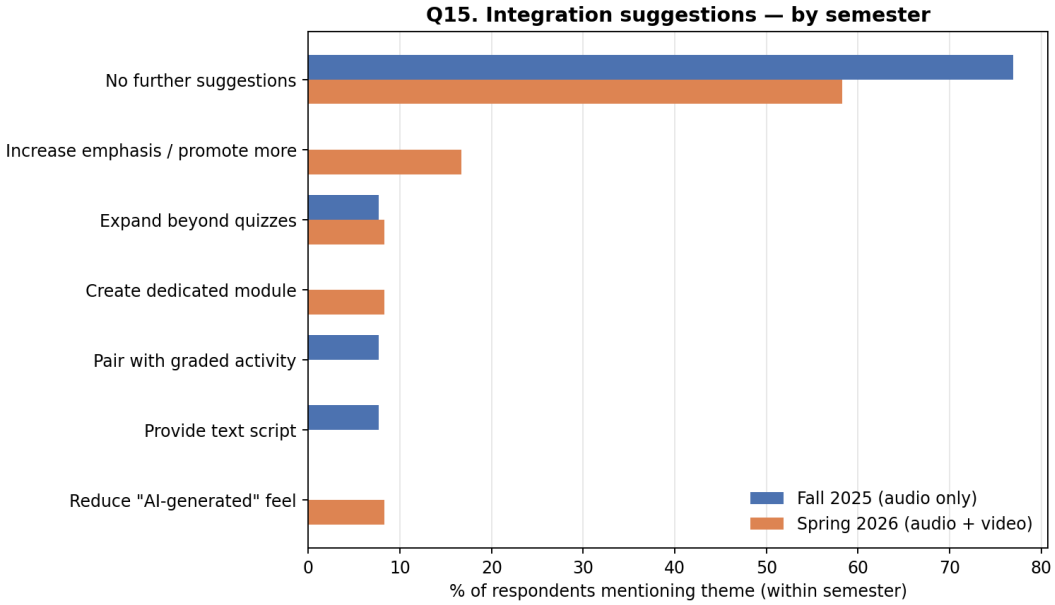


Figure 5: Q15. Suggestions for integration

4.3 Which Format Do Students Prefer?

Taken together, the evidence consistently, though not always statistically significantly and given the small sample, favors the Spring 2026 audio + video combination over audio alone offered in Fall 2025. The detailed highlights of the empirical survey findings are as follows:

- Every rated item except one ("felt more motivated to study," a negligible, non-significant difference) was scored higher in Spring 2026 than Fall 2025.
- The single statistically significant difference in the whole survey, willingness to recommend the format in other courses, favored Spring 2026 by a wide margin (84.2% vs. 47.6% top-2-box, p = .026).

- Qualitatively, Spring 2026 students spontaneously praised the multi-modal format (having "four different ways to take in the information") and reported comprehension benefits more often than Fall 2025 students.
- Spring 2026 students were also more likely to say nothing needed improvement and to ask for the format to be used more, both signs of stronger satisfaction.

The main countersignal is that the video addition made the "AI-generated" quality more salient for a few Spring 2026 students, who asked for a more human-sounding voice or less "AI-like" feel - a consideration for future production quality rather than evidence against the multi-modal approach itself.

Overall: the data are directionally consistent with audio + video outperforming audio alone from the students' perspective, anchored by one statistically significant result (recommendation intent) and a uniform pattern of higher point estimates across every other measure. However, with n ≈ 20 per section, most individual comparisons do not reach statistical significance on their own, so this should be treated as encouraging preliminary evidence rather than a confirmed effect - a larger sample (e.g., combining several future sections) would be needed to test it more rigorously.

5. Conclusion

This study compared student-reported outcomes across two consecutive offerings of Business Analytics I: a Fall 2025 section that used AI-generated audio podcasts alone (n = 24), and a Spring 2026 section that paired the same audio podcasts with companion videocasts (n = 20). Across nearly every measured dimension - perceived exam preparation, comprehension, retention, course engagement, and effectiveness relative to traditional study methods - the combined audio + video condition produced higher mean ratings than audio alone. The single statistically significant difference, a markedly higher willingness to recommend the format in other courses (84.2% vs. 47.6% top-2-box, p = .026, r = .40), together with converging qualitative evidence (students in Spring 2026 spontaneously cited the value of having "four different ways to take in the information" and reported comprehension gains more often), provides a coherent, if preliminary, picture: supplementing AI-generated audio with a video component appears to improve students' subjective experience of the material without introducing new complaints beyond a modestly increased sensitivity to the content's AI-generated origin. Taken together, the evidence favors continuing and refining the multi-modal (audio + video) approach over reverting to audio alone.

Limitations of this study should be understood as follows. Non-experimental, non-randomized design: Students were not randomly assigned to the audio-only or audio + video condition; the two groups are different cohorts taking the course in different terms. Any pre-existing differences between the Fall 2025 and Spring 2026 cohorts (prior exposure to podcasts, general familiarity with AI tools, major composition, motivation levels) are confounded with the treatment itself and cannot be ruled out as alternative explanations for the observed differences. Moreover, small sample size and limited statistical power: With 19–22 valid responses per semester, the comparison can reliably detect only large effects. Most of the observed differences, while consistently favoring Spring 2026, did not reach conventional statistical significance and should not be interpreted as null findings as they may simply reflect insufficient power. Thirdly, self-reported outcomes only: All academic-performance items (Q10) are students' perceptions of their own preparation and performance, not objective measures such as quiz scores or course grades. Perceived and actual performance can diverge, particularly for a novel study tool that may increase confidence without a proportional increase in mastery. And, lastly, single institution, single course, single instructor: All data come from one course (Business Analytics I) at one institution. Findings may not generalize to other subjects, student populations, or instructors.

Future work could be strategized in the following ways. Future iterations should merge (de-identified) survey responses with actual quiz and exam scores to test whether the higher self-reported preparation and comprehension in the audio + video condition correspond to measurable performance gains. Use a within-subject or crossover design: Rotating audio-only and audio + video modules within the same section (e.g., alternating by chapter) would let each student serve as their own control, removing the cohort-level confounds inherent in the current between-section comparison. Increase sample size across multiple terms: Pooling additional Fall and Spring sections over two or more years would substantially increase statistical power and allow the currently non-significant but consistently favorable trends (e.g., quiz preparation, comprehension) to be tested more conclusively. Adopt validated instruments: Replacing ad hoc Likert items with established, validated scales for engagement, motivation, and self-efficacy (with reported reliability coefficients) would strengthen the measurement quality of future surveys, and a pre-registered analysis plan would guard against

post hoc interpretation. Investigate the "AI-generated feel" concern directly: Since a subset of Spring 2026 students specifically requested a more natural, less "AI-like" narration voice, future work should experimentally vary voice/production quality (e.g., synthetic vs. human-recorded narration) to determine whether this is a meaningful driver of student satisfaction or a minor concern. Conduct follow-up qualitative interviews or focus groups: A small number of semi-structured interviews would allow deeper exploration of why students in the audio + video condition were more willing to recommend the format, and whether this reflects the video content itself, increased modality choice, or other course-level changes between terms. Examine differential effects by subgroup: Larger future samples would allow exploration of whether the benefit of adding video differs by prior podcast experience (Q3), self-reported AI familiarity (Q4), or academic major - questions the current sample is too small to address with adequate power.

Ethics Declaration

The IRB approval was received prior to conducting the empirical survey.

AI Declaration

Claude AI and MS Copilot were used for language proficiency improvement purposes.

References

Aguilar, F. R. (n.d.). Evaluating the use of mobile technologies for language learning purposes. https://doi.org/10.4995/thesis/10251/163675

Albin, C. S. W., Martindale, J. M., Marawar, R., Gottlieb-Smith, R., Paul, A., Strelzik, J., Christy, A., Xixis, K., & Goldstein, J. (2026). Let's Talk Tech: The Role of Technology in Neurology Education. Seminars in Neurology, 46(03), 296–310. https://doi.org/10.1055/a-2795-8556

Arruzza, E., Ofori-Manteaw, B., Nocum, D. J., Nelson, T., Awwad, D. A., Ago, J. L., Chau, M., Zaidi, T., & Jimenez, Y. (2026). Exploring the Use of Podcasting in the Medical Radiation Sciences. Health Science Reports, 9(4). https://doi.org/10.1002/hsr2.72237

Carmi, G. (2023). Students' attitude and learning effectivenes for two types of podcasts in MBA course. Online Information Review, 47(7), 1320–1339. https://doi.org/10.1108/oir-12-2021-0647

Devies, B., Bullock, L. M., Jenkins, D., Allen, S. J., & Stanberry, J. (2025). Sound Leadership: Harnessing the Power of Podcasts in Leadership Development. New Directions for Student Leadership, 2025(185), 45–52. https://doi.org/10.1002/yd.20653

Egilmez, G., & Schneider, K. (2025). *Could AI podcasts improve students' academic performance and engagement?: An empirical assessment*. In *Proceedings of the 6th International Conference on AI Research (ICAIR 2025)* (Vol. 6, No. 1, pp. 710–718). Academic Conferences International Limited. https://doi.org/10.34190/icair.5.1.4455

Graham, A. C., & Stevens, E. C. (2026). Generative AI Podcasts for Learning: Perceptions of Faculty and Graduate Nursing Students. Journal of Nursing Education, 65(7). https://doi.org/10.3928/01484834-20251029-02

Kurtböke, D. İ., Lacey, M., Smith, D. L., Grando, D., Winter, G., Ketheesan, N., Andronicos, N. M., Clarke, D., Jabbar, A., Waller, K. L., Zhang, Y. Y., Chaminda, A., Ranadheera, S., & Chaminda, S. (2023). "Students' perspectives on a podcast-based assignment exercise, while learning to communicate about infections and vaccine", Volume 44 Number 3. Microbiology Australia, 44(3), 118–172. https://doi.org/10.1071/mav44n3

Mason, F. E., Mayer, J., O'Toole, B. I., Williamson, J., & Russell, A. R. (2024). How to use podcasts in neurology. Practical Neurology, 24(5), 387–392. https://doi.org/10.1136/pn-2024-004109

Middleton, A. J. W. (n.d.). The situated agency of studio learning and its value for non-studio disciplines. https://doi.org/10.7190/shu-thesis-00459

Shah, G., Zier, C., Reich, A. J., Johnson, R. F., Dang, Q. L., Agan, A. D., Isaacson, B., & Mitchell, R. B. (2025). Medical Education Strategies for Otolaryngology in the Online World and the Role of Podcasting. Laryngoscope Investigative Otolaryngology, 10(4). https://doi.org/10.1002/lio2.70231

Stadler, M., Richters, C., Fischer, M. R., & Hutmacher, F. (2025). Evaluating AI-Generated Podcasts Versus Traditional Reading for Learning From Medical Articles: Protocol for a Mixed-Design Study Among Resident Physicians. Jmir Research Protocols, 14, e78505. https://doi.org/10.2196/78505

Artificial Intelligence as a Mediating Tool in Higher Education: An Exploratory Study in Quantitative Methods

Teresa Florentino
Escola Superior de Actividades Imobiliárias, Lisboa, Portugal

mariaflorentino@esai.pt

Abstract: This study investigates the role of Artificial Intelligence (AI), with a focus on generative tools such as ChatGPT, in higher education through the lens of pedagogical mediation. The analysis contends that the educational impact of AI is not direct or automatic, but rather depends on instructional design, teaching practices, assessment structures, and instructor guidance. An exploratory mixed-methods approach was employed in a Quantitative Methods I course involving undergraduate students from Real Estate and Construction programmes. A questionnaire was administered to 89 students, resulting in 24 valid responses and a response rate of 27%. Data from both closed- and open-ended questions were analysed using descriptive and interpretive methods, reflecting the voluntary and self-selected nature of the sample. The findings reveal that students primarily use AI to clarify concepts, assist with solving exercises, and organise their study routines. A significant proportion also reported encountering incorrect or misleading outputs, underscoring the necessity for validation, critical interpretation, and explicit pedagogical guidance. The results suggest that AI functions mainly as a complementary learning resource rather than a replacement for cognitive engagement, especially when integrated into self-regulated learning processes (Kasneci et al., 2023; Panadero et al., 2023). This study introduces a Pedagogical Mediation Framework to analyse how AI use is shaped by the interplay among curriculum, assessment, teaching mediation, and students' critical engagement.

Keywords: Artificial Intelligence in Education, Higher Education Pedagogy, Self-Regulated Learning, Quantitative Methods Teaching, Educational Innovation.

1. Introduction

The rapid proliferation of generative artificial intelligence (AI), particularly ChatGPT, has intensified discussions regarding learning, teaching, assessment, and academic integrity in higher education. Recent scholarship suggests that the implications of AI cannot be fully captured by a simple dichotomy of benefit versus risk. Rather, the educational significance of AI is contingent upon the pedagogical conditions under which it is introduced, interpreted, and assessed by both students and instructors (Kasneci et al., 2023; Tlili et al., 2023). This consideration is particularly pertinent in quantitative and applied disciplines, where students may employ AI not only to obtain answers but also to clarify concepts, explore procedures, and support independent study.

Research on ChatGPT in higher education demonstrates that both students and educators recognise opportunities as well as limitations. AI tools can facilitate study organisation, concept clarification, and autonomous learning; however, concerns remain regarding overreliance, misinformation, and academic integrity (Dos, 2025; George-Reyes et al., 2025). These findings reinforce a pedagogical perspective on AI integration, indicating that the value of such tools is not inherent but is determined by the tasks, expectations, and guidance associated with their use.

In this study, Artificial Intelligence is conceptualised as a mediating resource within the learning process, rather than as an independent determinant of learning outcomes. This position relates to Brousseau's (1997) concept of the didactic contract, which foregrounds the relationships among teacher, student, and knowledge. Generative AI modifies aspects of these relationships by giving students access to explanations and procedural support beyond the classroom. The central issue is therefore not simply whether students use AI, but how this use is framed, monitored and connected to disciplinary learning objectives.

This issue is particularly relevant in Quantitative Methods I (QMI), where students frequently struggle not only with calculations but also with interpretation and application to professional contexts. In this type of course, generative AI may help students clarify procedures, test their reasoning, and overcome initial barriers to problem-solving. At the same time, it may encourage answer reproduction when the assessment is limited to final results. The integration of AI in QMI therefore requires a pedagogical frame that values explanation, interpretation, justification and critical verification.

This paper advances current debates on AI in higher education in three key ways. First, it analyses AI from a pedagogical perspective, examining how teaching practices and assessment conditions influence students' use of AI. Second, it presents exploratory empirical evidence from a Quantitative Methods course, focusing on students' reported practices, perceptions, and critical awareness. Third, it introduces a Pedagogical Mediation

Framework that differentiates between AI use as support for understanding and AI use as a mechanism for answer production.

2. Theoretical Review

2.1 Artificial Intelligence in Higher Education

Research on ChatGPT in higher education indicates heterogeneous patterns of student use. Some students use it to save time or clarify doubts; others seek additional explanations when they lack confidence or are reluctant to ask questions in class. This diversity matters because AI use cannot be reduced to a simple opposition between misconduct and innovation. Its educational value depends on the nature of the task, the learning context, and students' capacity to critically interpret the outputs. Earlier reviews of AI in higher education also highlight the need to examine not only technological applications, but also the pedagogical role of educators in AI-supported learning environments (Zawacki-Richter et al., 2019).

For this reason, debates on AI in higher education should move beyond questions of access, prohibition or technological novelty. A more productive focus is whether pedagogical practices enable students to distinguish a valid explanation from a plausible but weak response, a justified procedure from a mechanical answer, and a meaningful conclusion from a disconnected result. In quantitative courses, this distinction is central because a correct numerical answer is insufficient evidence of learning if students cannot explain the procedure, interpret the result or relate it to the problem context.

Artificial Intelligence is frequently discussed in higher education as a form of technological innovation. Nevertheless, its educational relevance depends on how it is embedded in teaching and learning contexts. AI may support concept clarification, alternative explanations and study organisation (Kasneci et al., 2023), but it also raises concerns about superficial learning, technological dependence and the reliability of generated responses. Generative AI intensifies these long-standing tensions because it can produce immediate, persuasive answers that may still be incomplete, inaccurate, or contextually inappropriate.

The emergence of AI also exposes limitations in pedagogical models that remain centred on lectures, procedural repetition and assessment of final answers. In such contexts, AI becomes disruptive by challenging evaluation methods that do not clearly distinguish conceptual understanding from answer reproduction. The issue is therefore not only technological, but pedagogical. When integrated into tasks that require reflection, interpretation and feedback, AI may support understanding, autonomy and analytical engagement. When used mainly to accelerate answer production, it may reinforce superficial learning strategies.

2.2 Self-Regulated Learning and Pedagogical Framing

This distinction is important because AI can create an appearance of autonomy without genuine self-regulation. A student may seem independent because answers are obtained quickly; however, this independence is fragile if it is not accompanied by monitoring, verification and reflection. Genuine autonomy requires students to decide when to use the tool, how to question it, how to compare its outputs with course materials and when to seek teacher support.

Pedagogical framing, therefore, concerns not only rules about AI use but also the intellectual work students are expected to undertake with it. Students may, for example, be asked to identify errors in an AI-generated solution, compare alternative explanations, justify why a statistical or mathematical answer is valid, or explain how a result should be interpreted in a real estate or construction context. Such activities reposition AI from an answer generator to an object of analysis.

Self-regulated learning offers a useful lens for examining AI-supported learning environments. It refers to learners' capacity to plan, monitor and evaluate their own learning processes (Zimmerman, 2002; Panadero et al., 2023). AI tools may contribute to these processes by providing alternative explanations, rapid feedback and opportunities to revisit concepts. Used uncritically, however, they may promote cognitive delegation rather than engagement. The key distinction is therefore between cognitive support and cognitive substitution.

In this paper, pedagogical framing refers to the ways in which teaching practices, assessment models, and instructor guidance structure students' use of technological tools. The same AI tool may produce different educational effects depending on whether students are required to justify reasoning, compare outputs, validate answers and reflect on errors. Meaningful integration of AI, therefore, requires alignment with curriculum goals,

explicit guidance on validation and interpretation, and assessment environments that prioritise conceptual understanding over answer production.

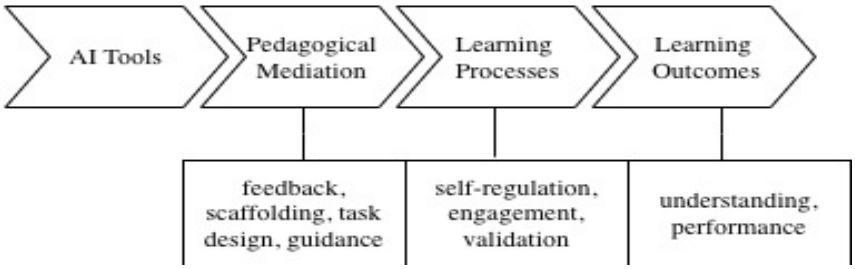


Figure 1: Pedagogical mediation framework of AI in Higher Education

As represented in Figure 1, AI is positioned as an element within a mediated pedagogical system rather than as a direct cause of learning outcomes.

2.3 Quantitative Methods, Context and Meaning

The historical comparison with calculators helps to clarify this point. Calculators did not eliminate the need for mathematical understanding; rather, they required educators to reconsider what should be taught and assessed. Similarly, AI should not displace quantitative reasoning, but should encourage stronger emphasis on interpretation, justification and decision-making. If assessment remains centred on final answers, students will tend to use AI to produce them. If assessment values are explained and applied, AI can support the learning process.

This shift is particularly relevant in applied programmes. Students in Real Estate and Construction need to understand how quantitative information informs decisions on prices, investments, risks, markets, buildings, and territories. Quantitative methods should therefore be connected to professional meaning. AI may help generate examples, clarify procedures and support data exploration, but teacher mediation remains necessary to connect these outputs to disciplinary knowledge and professional judgement.

The issue is especially visible in quantitative courses. Research in mathematics didactics has long shown that students do not learn mathematics through the memorisation of isolated procedures, but through the construction of meaning in relation to context, language, problems and feedback (Brousseau, 1997). Despite this, teaching in quantitative disciplines often remains formula-centred and procedural. Students may complete calculations while still struggling to interpret, analyse and apply results in context.

In QMI, these difficulties are intensified by the diversity of students' mathematical backgrounds. Some students have fragmented prior knowledge or long gaps since their last formal mathematics learning, which may generate anxiety and a sense of inadequacy. AI can provide additional explanations, examples and practice opportunities, but it can also create an illusion of learning when outputs are accepted without understanding the reasoning behind them. This risk is consistent with research showing that minimal guidance is insufficient for complex learning tasks (Kirschner et al., 2006).

For students in Real Estate and Construction programmes, quantitative methods acquire meaning when they are connected to concrete data, markets, territories and professional decision-making. Project-based methodologies, real data analysis, and contextualised problems may enable AI to support exploration, visualisation, and the exploration of alternative scenarios. Without such pedagogical redesign, AI may compensate for weaknesses in traditional instruction without transforming the learning process.

2.4 Ethical and Institutional Responsibility

The integration of AI in higher education also requires ethical and institutional reflection. Guidelines on academic honesty and integrity are necessary, but they are insufficient without pedagogical support. Ethics in AI use is not limited to plagiarism prevention; it also involves the construction of learning environments that promote understanding, critical thinking and intellectual autonomy. In this context, the instructor's role becomes more central, not less: the teacher acts as mediator, facilitator and designer of learning experiences, clarifying expectations and aligning AI use with curriculum goals.

At the institutional level, debates on AI should be accompanied by revisions of curricula, teaching strategies and assessment models. It is inconsistent to demand pedagogical innovation while maintaining organisational conditions that restrict active methodologies, such as excessive class sizes or teaching loads incompatible with close monitoring. AI does not resolve these contradictions; it makes them more visible. The exploratory study presented below is therefore situated within a descriptive research logic aimed at understanding students' actual practices and perceptions in QMI.

3. Methodology

This study adopts an exploratory mixed-methods design, combining quantitative and qualitative data to examine how undergraduate students use generative AI tools in Quantitative Methods I. The exploratory orientation is justified by the evolving nature of AI use in higher education and by the need to understand students' practices in a specific teaching context. The objective is not statistical generalisation, but the identification of initial patterns, perceptions and pedagogical issues that may inform further research.

This approach is consistent with exploratory mixed-methods research, in which quantitative and qualitative evidence are combined to provide a broader understanding of an emerging educational phenomenon (Creswell & Creswell, 2018).

The mixed-methods design was appropriate because the research question required both a descriptive overview of the frequency and contexts of AI use and an interpretation of the meaning's students attributed to that use. Quantitative data identified response patterns, whereas qualitative answers helped clarify students' concerns about errors, dependence and the need for guidance. The study was conducted in an authentic teaching context rather than in an experimental setting, and the evidence should therefore be interpreted as a situated first approximation to an emerging phenomenon.

The study population consisted of 89 students enrolled in QMI during the first semester of undergraduate programmes in Real Estate Management and Construction and Building Management. The questionnaire was made available to all students. Participation was voluntary, anonymous and not associated with assessment. A total of 24 valid responses were obtained, corresponding to a response rate of approximately 27%. The sample is therefore self-selected and non-probabilistic.

The modest response rate should be interpreted in context. Data collection occurred at the end of the semester, close to the examination period, when students were primarily focused on assessment and exam preparation. In addition, because participation was not linked to immediate academic benefit, some students may have perceived the questionnaire as having limited relevance. Survey research recognises that timing, perceived relevance, respondent fatigue and voluntary participation affect response rates (Dillman et al., 2014; Cohen et al., 2018).

The questionnaire items were developed by the author specifically for this study and for the context of Quantitative Methods I. The wording was informed by teaching experience in the course, familiarity with previous educational questionnaires and the aim of identifying four dimensions: contexts of AI use, perceived support for learning, awareness of errors and expectations regarding pedagogical guidance. The instrument combined closed questions, used to describe response patterns, and open-ended questions, used to capture students' explanations and concerns.

No formal pilot test, external review or psychometric validation was conducted before application. This is a methodological limitation. However, because the study is exploratory and situated in a specific teaching context, the questionnaire should be understood as an initial instrument for identifying relevant issues rather than as a validated measurement scale. Future work should include expert review, cognitive pre-testing with students and, where appropriate, reliability analysis.

Given the small sample and voluntary participation, findings should be interpreted with caution. Respondents may include students with a greater interest in, or awareness of, AI tools. Consequently, the findings are not statistically generalisable. Their relevance is analytical and pedagogical: they identify issues that may be compared with other courses, cohorts and institutional settings in future research.

4. Results

Overall, the results indicate that AI was used mainly as a complementary learning support resource rather than as a systematic substitute for study, reasoning or instructor guidance. Given the exploratory design, the results

are interpreted descriptively and are not used to infer statistically significant relationships. The programme distribution showed that 75% of respondents were enrolled in Real Estate Management and 25% in Construction and Building Management, broadly reflecting the composition of the QMI class.

4.1 Contexts and Frequency of AI Use

The distribution of responses suggests that students use AI at different points in the learning cycle. Some uses are directly associated with problem-solving, such as checking steps in exercises. Others relate to preparation and consolidation, including study organisation and clarification of doubts after class. This variety shows that AI use does not have a single pedagogical meaning; its role depends on when it enters the learning process and whether it is used before, during or after students' own reasoning.

The presence of non-users is also relevant. It indicates that AI adoption should not be assumed to be universal, even when tools are widely available. Non-use may reflect limited familiarity, a lack of trust, low perceived usefulness, or a preference for traditional study practices. This reinforces the need for explicit classroom discussion, since students may begin from very different levels of AI literacy.

As shown in Figure 2, the most frequent context of AI use was support in solving exercises, selected by 11 respondents. This result is relevant because QMI involves quantitative reasoning, procedures, and practical problem-solving. The second most common use was understanding theoretical concepts, mentioned by 9 respondents. Other uses included clarification of doubts after classes, study organisation, diagrams and planning, each selected by 7 respondents. These findings suggest that students do not use AI exclusively to obtain final answers; they also use it to access alternative explanations and organise autonomous study. However, 6 respondents reported not using AI in QMI, indicating uneven adoption.

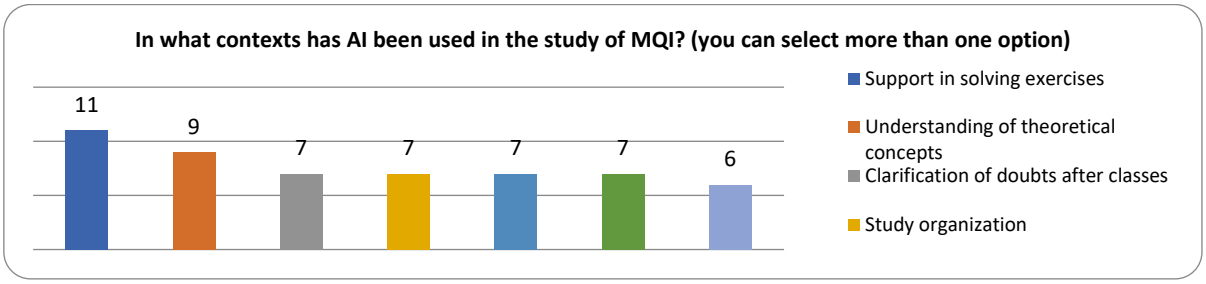


Figure 2: Contexts of AI use in the study of QMI

As shown in Figure 3, AI use was mostly occasional. Thirteen respondents reported using AI sometimes, 5 reported using it often, 5 reported never using it, and 1 reported using it rarely. This pattern suggests that AI is present in students' study practices but does not dominate their routines. It appears to function as an additional resource activated in specific situations, particularly when students require help with exercises, concepts or clarification of doubts.

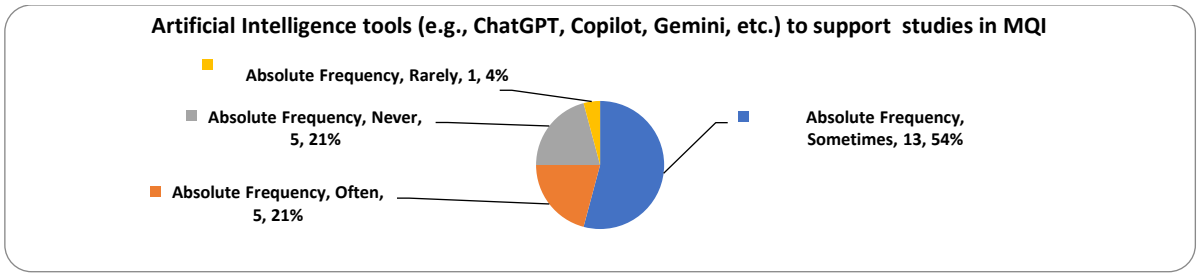


Figure 3: Frequency of AI use in QMI

4.2 Understanding, Autonomy and Awareness of Limitations

Student awareness of AI limitations was one of the main findings in this section. The results should therefore be read not only as evidence of tool use, but also as evidence of how students judge, question and verify AI outputs in a quantitative learning context.

In QMI, this may involve asking students to compare AI-generated solutions with manual calculations, identify missing assumptions, test whether a formula was correctly applied and explain why a result is meaningful in

context. Such practices align with the broader objective of developing quantitative literacy, understood not only as the capacity to compute but also to evaluate and communicate results.

A central analytical distinction in the study concerns AI use oriented toward conceptual understanding, and AI use aimed primarily at obtaining answers. The results suggest that many students recognise this distinction. AI was perceived as a tool that can clarify reasoning, provide alternative explanations and help organise thinking. In a course such as QMI, where students must connect formulas, procedures and interpretation, this function is pedagogically relevant.

Regarding autonomy, several students reported that AI contributed, at least partially, to independent study. This does not imply that AI replaces the instructor or the structured learning environment. Rather, it suggests that students use AI as additional support between classes, particularly when immediate teacher support is unavailable. From a self-regulated learning perspective, students mobilise available resources to monitor understanding, manage difficulties and obtain feedback (Zimmerman, 2002; (Zimmerman, 2002; Molenaar et al., 2023)).

However, 75% of respondents reported encountering situations in which AI generated incorrect answers or confusing explanations. This finding is important because it indicates critical awareness of AI limitations. Students do not appear to attribute unquestioned authority to AI; instead, they recognise the need to validate, compare and verify generated information. This result is consistent with the literature, which emphasises that AI outputs may be incomplete, inaccurate, or misleading, and that students require verification practices and critical judgement (Kasneci et al., 2023; Miao & Holmes, 2023).

4.3 Pedagogical Guidance and Future Integration

Most respondents considered it important for instructors to provide explicit guidance on the ethical and critical use of AI in courses. When asked how AI should be integrated into QMI, 12 students (50%) supported integration as a pedagogical aid, 7 (29%) supported its use with clear limits, 4 (17%) expressed no opinion, and 1 (4%) rejected integration. These results indicate a broadly favourable position, but not unconditional acceptance. Students appear to expect organised guidance, explicit boundaries and clarification of appropriate use in learning and assessment contexts.

Open-ended responses supported this interpretation. Students described AI as useful for understanding processes, obtaining alternative explanations and organising study, while also identifying the risks of excessive dependence and the need to validate responses. The recurrent idea is that AI should not replace the instructor, but may complement face-to-face teaching when its use is explicitly framed.

5. Discussion and Pedagogical Implications

The results indicate that students expect the teacher to establish boundaries for AI use. This expectation is significant because it challenges the assumption that students necessarily prefer unregulated access to AI. In this sample, acceptance of AI was associated with guidance, clarity and limits. Students recognised its usefulness, but also the need to connect its use to course objectives and assessment criteria.

This observation has practical implications for QMI. AI could be incorporated into controlled learning activities, for example, by asking students to document prompts, evaluate the quality of generated explanations, identify errors in AI-assisted solutions or explain how they verified a result. These tasks would make AI use visible and assessable without replacing students' reasoning. They would also reduce ambiguity around academic integrity by clarifying what is appropriate and what is not.

At the same time, the findings suggest the need for caution regarding purely technological interpretations of educational change. Access to AI does not automatically improve learning. Without changes in task design and assessment, AI may simply accelerate existing superficial strategies. The pedagogical challenge is therefore to design tasks in which AI use requires explicit reasoning, validation of outputs and interpretation of results.

The exploratory evidence suggests that students' use of AI in QMI does not necessarily involve uncritical delegation of learning. Respondents distinguished between understanding and merely obtaining answers, recognised limitations of AI tools and valued the instructor's role as mediator. This finding is relevant because public and institutional debates often associate AI with the decline of critical thinking. The data support a more nuanced interpretation: students mainly mobilise AI when they need additional explanations, examples, time or support beyond what traditional instructional models provide.

The educational value of AI, therefore, depends on the pedagogical conditions in which it is used. With appropriate guidance, AI may support reflection, autonomy and conceptual clarification. Without such guidance, it may encourage instrumental uses focused on answer production. In quantitative disciplines, assessment should therefore move beyond final results and require explanation, interpretation, validation of outputs and application to real contexts.

From the perspective of mathematics didactics, the findings support the need to re-centre pedagogical models on meaning construction and contextualised application. In QMI, learning should be understood as the capacity to mobilise quantitative concepts to interpret real phenomena in real estate and construction contexts, rather than as isolated mastery of procedures. Project-based methodologies, real data and contextualised analysis may enable AI to function as support for reasoning rather than as a shortcut.

The discussion also has institutional implications. Active pedagogies, close monitoring and contextualised learning require adequate organisational conditions. It is difficult to promote critical AI use in environments characterised by excessive class sizes, rigid curricula, or assessment models centred mainly on final products. AI integration should therefore be connected to broader reflection on curriculum, workload, assessment and pedagogical innovation.

Although the study was conducted in the context of Quantitative Methods I, its findings may be relevant to other disciplinary areas in higher education where students use AI to clarify concepts, generate explanations and organise study practices. Transferability depends on the similarity between contexts, particularly in relation to assessment design, teacher guidance, student AI literacy and the type of disciplinary tasks involved. The results should therefore not be read as directly generalisable, but as analytically useful for instructors facing comparable pedagogical questions.

The Pedagogical Mediation Framework can be operationalised through simple teaching and assessment practices. In curriculum design, students may be asked to formulate prompts, compare AI-generated explanations with course materials, identify errors in AI outputs and connect results to disciplinary concepts. In assessment, students should not be evaluated only on final answers, but also on reasoning, justification, validation and interpretation. These practices make AI use pedagogically visible and help distinguish legitimate support from substitution of student work.

6. Conclusion

This study examined the role of AI in higher education through a pedagogical mediation perspective, drawing on exploratory evidence from a Quantitative Methods I course. The results indicate that students mainly use AI as a complementary support resource, particularly for solving exercises, clarifying concepts and organising their study. At the same time, students reported awareness of AI limitations, including incorrect or confusing outputs, and expressed the need for instructor guidance and clear limits.

The study's main contribution lies in the proposed Pedagogical Mediation Framework. Although modest and exploratory, this contribution is relevant because it shifts the analysis from the general question of whether AI is beneficial or harmful to the more precise question of how pedagogical design shapes its educational effects. AI should therefore be analysed less as an isolated technological factor and more as a resource whose meaning depends on instructional design, assessment practices, teacher guidance and students' critical engagement.

The low response rate and the self-selected nature of the sample constitute important limitations. Data collection took place at the end of the semester, close to the examination period, which may have reduced students' availability and willingness to participate. Respondents may also have been more interested in AI, more engaged with the course, or more willing to reflect on their learning practices than non-respondents. Consequently, the findings may overrepresent students with greater awareness of AI tools or stronger views about their use. The results should therefore be interpreted as exploratory and descriptive, rather than as representative of the whole student population.

Future research should adopt longitudinal and comparative designs, following students over a semester or academic year and comparing different classes, quantitative and non-quantitative courses and, at a later stage, different institutions. Further studies should examine how teacher guidance influences students' prompting, verification and interpretation practices. A future version of the instrument should also be reviewed by experts, piloted with students and refined before broader application. Given the methodological difficulty of measuring the direct impact of AI on academic performance, initial research may focus on students' practices, perceptions, critical awareness and quality of reasoning.

In sum, the study suggests that the educational relevance of AI depends on the pedagogical work built around it. In QMI and similar courses, AI is most productive when it supports questioning, explanation, verification and contextual interpretation, rather than when it is used merely to obtain final answers.

Ethics Declaration

Ethical clearance was not formally required for this study under the institutional procedures applicable to this context. Participation in the questionnaire was voluntary, responses were anonymised, and no personally identifiable information was used in the analysis.

AI Declaration

AI tools were used in a limited way to support language revision, text refinement and organisational editing. The conceptual framing, interpretation of findings, argumentation and final academic responsibility remain entirely with the author.

References

Brousseau, G. (1997). Theory of didactical situations in mathematics. Dordrecht, Netherlands: Kluwer Academic Publishers.

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method (4th ed.). Wiley.

Dos, I. (2025). A systematic review of research on ChatGPT in higher education. The European Educational Researcher. https://doi.org/10.31757/euer.824

George-Reyes, C. E., Avello-Martínez, R., & Buenestado-Fernández, M. (2025). Perceptions of ChatGPT and the complexity of its impact among higher education students: Evidence from 10 Latin American and European countries. Educational Process: International Journal, 15, Article e2025171. https://doi.org/10.22521/edupij.2025.15.171

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Molenaar, I., Mooij, S. de, Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior, 139*, 107540. https://doi.org/10.1016/j.chb.2022.107540

Panadero, E., Andrade, H. L., & Brookhart, S. M. (2023). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. Educational Psychologist, 58(1), 1–19. https://doi.org/10.1080/00461520.2022.2147053

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., Loglo, P., Mishra, S., Molinaro, M., Ossiannilsson, E., Ramírez-Montoya, M. S., Shon, J. G., & Burgos, D. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Effect of Open and Distance Learning Programmes on Inclusivity in Higher Education

Vishaka Nanayakkara, Savin Gunawardena, Buddhika Karunarathne and Dhanushima Gamage
University of Moratuwa, Sri Lanka

vishaka@cse.mrt.ac.lk
savin.20@cse.mrt.ac.lk
buddhika@cse.mrt.ac.lk
dhanushima@bit.uom.lk

Abstract: Ensuring equitable access to higher education remains a persistent challenge in Sri Lanka, where limited intake capacity in conventional university programs restricts opportunities for many qualified learners. The Centre for Open and Distance Learning (CODL) at the University of Moratuwa, Sri Lanka provides a substantial opportunity for a wider cohort of students by offering the Bachelor of Information Technology (BIT) external degree, delivered in the Open and Distance Learning (ODL) mode. This study aims to investigate the extent to which the BIT program enhances inclusivity by expanding access to higher education and supporting diverse learner participation. A comparative analysis was conducted between students enrolled in the BIT external degree and those in the internal degree programmes. Student profiles were examined across multiple dimensions including age, gender, ethnicity, geographic distribution, and educational background. This analysis aimed at identifying how ODL broadens access beyond traditional student populations. Particular emphasis was placed on geographic and ethnic inclusivity by evaluating the representation of students, highlighting the program's role in reducing regional and ethnic disparities in higher education access. Beyond access, the study evaluated inclusivity in terms of participation and academic success. Retention rates and completion rates were analysed to assess sustained student engagement within the BIT programme. Academic performance indicators, including grade distributions and pass rates, were examined across different demographic groups to determine whether expanded access through ODL is accompanied by comparable learning outcomes. In addition, the study considered the role of programme flexibility in enabling participation among non-traditional learners, particularly working professionals and individuals with family or other commitments. The findings of this paper indicate that the ODL mode attracts a more diverse student population, who may otherwise lack access to higher education. Furthermore, the results suggest that this expanded access does not significantly compromise academic performance, demonstrating that inclusivity can be achieved alongside quality. The learnings from the external degree programme conducted in the ODL mode provide a scalable and practical model for institutions seeking to enhance inclusivity in large-scale education systems.

Keywords: Open and Distance Learning; Inclusive Higher Education; Information Technology Education; External Degree Programme; Widening Participation; Sri Lanka

1. Introduction

1.1 Higher Education Access and Inclusivity

Equitable access to higher education remains a challenge in Sri Lanka, where qualified students compete for limited places in state university programmes. In technology-related disciplines, access may be further shaped by programme-specific entry requirements, particularly prior preparation in mathematics or science. Consequently, learners seeking IT education may face barriers arising from both limited intake capacity and conventional entry structures. Inclusivity extends beyond admission. An inclusive higher education pathway should attract learners from varied geographic, ethnic, gender, age and educational backgrounds, while enabling meaningful academic progression. Open and distance learning (ODL) is significant in this context because it can widen participation by reducing dependence on continuous campus attendance and providing flexible routes into higher education. Studies in Sri Lanka similarly emphasise that inclusive online and distance learning requires both access and meaningful learner engagement (Yang et al., 2022).

1.2 The BIT External Degree as an ODL Pathway

The Bachelor of Information Technology (BIT) external degree programme, offered by the Centre for Open and Distance Learning (CODL) of the University of Moratuwa, provides an ODL-based route into IT higher education. The programme is structured across three academic years, with progressive certification: Diploma after Year 1, Higher Diploma after Year 2 and the BIT degree after Year 3. Its admission route is open to students who have obtained three simple passes in the G.C.E. Advanced Level (A/L) examination in any stream, together with a

credit pass in G.C.E. Ordinary Level (O/L) Mathematics. This structure makes the BIT programme a suitable case for examining inclusivity. Unlike several internal computing-related routes whose entry conditions are linked to specific A/L mathematics or physics related preparation, BIT creates an opportunity for students from a wider range of educational backgrounds to enter IT higher education.

1.3 Aim and Contribution of the Study

This study analyses the 23S2 BIT intake, enrolled in July 2024, comprising 719 students. Inclusivity is examined through age, gender, ethnicity, geographical residence and prior A/L stream. Since the cohort was in Semester 4 at the time of analysis, final degree or Higher Diploma completion could not be evaluated. Instead, achievement of the Diploma-level milestone, defined as completion of all Year 1 requirements, is used as an indicator of early academic progression.

The study first examines diversity at enrolment and then considers whether it remains visible among the 157 students who had achieved the Diploma-level milestone, supported by selected foundational-module outcomes. The comparison with internal degree programmes is contextual: the BIT cohort is analysed directly, while conventional internal pathways are considered through official admission requirements and related evidence.

The study addresses the following questions:

- What demographic, geographic and educational-background diversity is evident among BIT students?
- How does the BIT entry pathway widen access relative to conventional internal computing-related routes?
- Is diversity visible among students achieving the Diploma-level milestone?
- Do selected academic outcomes indicate meaningful early progression among these students?

Through this analysis, the paper examines BIT as an ODL-based pathway for widening participation in IT higher education while supporting early academic progress.

2. Background and Related Work

2.1 Inclusive Higher Education Through ODL

Inclusive higher education aims to provide meaningful opportunities for learners whose participation may be constrained by conventional entry pathways, geographical location, age or personal obligations. ODL is relevant in this context because it can reduce dependence on continuous physical attendance and provide alternative routes for learners whose circumstances make conventional campus-based participation difficult. In this sense, ODL has been associated with widened opportunity and social inclusion in higher education (Tait, 2013; Butcher and Rose-Adams, 2015).

Inclusivity, however, should not be assessed only by the diversity of students who enrol. It should also consider whether learners from varied backgrounds remain represented as they progress academically. Building on this perspective, the present study examines diversity among students entering an ODL-based IT programme and among those achieving its first formal academic milestone. Sri Lanka also has an established history of using ODL to expand higher education opportunities beyond conventional campus-based provision (Jayatilleke and Kulasekera, 2020).

Recent international studies on inclusive online and distance higher education emphasises that widening participation should be understood as more than expanding enrolment. Inclusive ODL environments must also support participation, engagement and progression, particularly for learners whose access to conventional higher education may be shaped by location, prior educational opportunity, employment, caring responsibilities or other personal circumstances. Cathy Stone's work on online and distance learner success emphasises that flexibility can widen opportunity, but student progression depends on sustained engagement, institutional support and course design (Stone, 2019).

2.2 Conventional Computing-Related Entry Pathways in Sri Lanka

Admission to Sri Lankan state universities is competitive and programme-specific. Internal Computer Science programmes require mathematics, physics or related A/L preparation, while Computer Science and Engineering at the University of Moratuwa is offered through the Engineering route. The internal BSc Honours in Information Technology permits eligibility from different A/L streams, but requires at least a credit pass in Higher

Mathematics, Combined Mathematics, Mathematics or Physics. The BIT external degree is not a replacement for these internal programmes. Rather, internal pathways provide a contextual benchmark for examining access to computing-related higher education. BIT complements these routes by enabling participation from students with a broader range of prior educational backgrounds.

3. Methodology

3.1 Research Design and Comparative Approach

This study adopts a comparative case-study approach to examine the contribution of the BIT external degree programme to inclusive IT higher education. The primary empirical analysis is based on student data from one BIT cohort offered through CODL, University of Moratuwa. This analysis is compared with conventional internal computing-related degree pathways through their published admission structures and subject-based entry requirements. Equivalent individual-level demographic and academic records from internal programmes were not available for the present study. Therefore, all comparisons with internal programmes in this paper should be understood as structural comparisons of access pathways, based on published admission requirements, rather than direct empirical comparisons of student profiles or academic outcomes across external and internal cohorts.

Accordingly, the comparison focuses on how the BIT external degree differs from conventional internal pathways in the breadth of educational backgrounds from which students can enter IT higher education, while analysing diversity and early progression within the BIT cohort directly. Internal Computer Science and Information Technology pathways commonly require mathematics or physics related A/L preparation, whereas the BIT external degree admits applicants with three simple passes in any A/L stream together with a credit pass in O/L Mathematics. This design enables examination of whether an ODL-based external programme provides an expanded route into IT higher education for learners whose educational background, geographic location or personal circumstances may limit participation through conventional pathways.

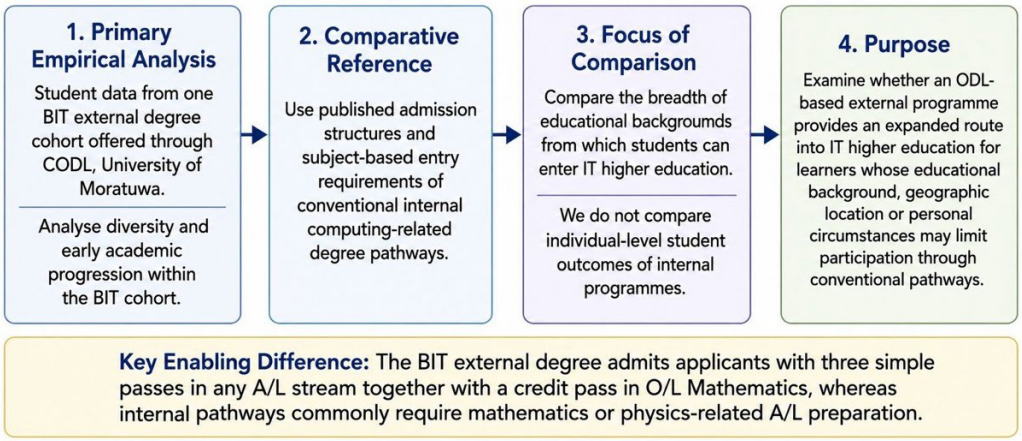


Figure 1: Comparative case-study approach used to examine inclusive access through the BIT external degree programme.

3.2 Dataset and Study Cohort

The dataset relates to the 23S2 intake of the BIT external degree programme, enrolled in July 2024. Following the removal of identifiers, records for 719 students were analysed. The dataset included demographic, geographical, educational background and early academic progression variables, as summarised in Table 1.

Table 1: Description of the study cohort and dataset

Programme | Delivery institution	Bachelor of Information Technology (BIT) external degree programme | Centre for Open and Distance Learning (CODL), University of Moratuwa
Study cohort | Enrolment period	23S2 intake | July 2024
Records analysed	719 students
Demographic variables	Age at enrolment, gender and ethnicity

Geographic variable	District and province
Educational background variables	A/L stream and O/L Mathematics grade
Early academic progression variables	Grades in Mathematics & Statistics for IT and Fundamentals of Programming
Progression variable	Achievement of the Diploma-level milestone

The BIT degree is structured across three academic years and six semesters, with a defined academic milestone associated with each year of study. These milestones and their relevance to the present analysis are shown in Table 2.

Table 2: BIT programme structure and progression indicator used in the study

Programme Stage	Academic Period	Academic Milestone	Status of the 23S2 Cohort at the Time of Analysis	Use in this Study
Year 1	Semesters 1–2	Diploma-level milestone	Eligible to achieve	Used as the indicator of early academic progression
Year 2	Semesters 3–4	Higher Diploma	Cohort was still in Semester 4	Not used as a completion outcome
Year 3	Semesters 5–6	BIT Degree	Cohort had not reached this stage	Not used as a completion outcome

At the time of analysis, the 23S2 cohort was in Semester 4 and was therefore not yet eligible to complete the Higher Diploma or the BIT degree. Accordingly, achievement of the Diploma-level milestone, defined as completion of all Year 1 subjects, was used as an indicator of successful early academic progression. Of the 719 students in the cohort, 157 students had achieved this milestone.

3.3 Dimensions of Analysis

Inclusivity was examined through age, gender, ethnicity, geographical location and prior educational background. Age was grouped into five categories: below 20, 20–24, 25–29, 30–39, and 40 years and above. Ethnicity was analysed using the categories Sinhala, Tamil and Muslim. Geographical reach was analysed by district and province, with participation from outside the Western Province used as an additional indicator because the University of Moratuwa is located within that province. Educational-background diversity was analysed using students' A/L streams. Early progression was examined through the demographic and geographical profile of students achieving the Diploma-level milestone. Academic achievement among these students was summarised using grades in two foundational modules: Mathematics & Statistics for IT and Fundamentals of Programming.

3.4 Data Analysis and Ethical Considerations

Descriptive statistics, primarily frequencies and percentages, were used to analyse three dimensions of inclusivity: diversity at enrolment, diversity among students achieving the Diploma-level milestone, and selected academic achievement indicators among milestone achievers. Detailed demographic comparisons of Diploma-level milestone-achievement rates are reserved for a separate focused study. All findings are reported in aggregated form, with student identifiers excluded. Ethnic and geographical data are used to describe representation and reach, not to make judgements about particular communities. The analysis identifies patterns of participation and early progression, but does not establish the reasons for enrolment or academic outcomes.

4. Inclusivity of Access in the BIT External Degree Programme

The findings are presented in two stages. First, the analysis examines inclusivity of access by considering the diversity of the full BIT cohort in relation to prior educational background, geographical location, gender, ethnicity and age. Second, the paper considers whether this diversity remains visible among students who achieved the Diploma-level milestone, supported by selected academic achievement indicators. This structure enables the study to distinguish between widened access and visible early progression, while avoiding claims about final programme completion.

4.1 Diversity of Prior Educational Background

A key indication of inclusivity in the BIT cohort is the diversity of students' prior A/L streams. Of the 719 students analysed, only 155 students (21.6%) entered from the Physical Science stream, while the remaining students came from Arts, Technology, Commerce, Bio Science and Common streams.

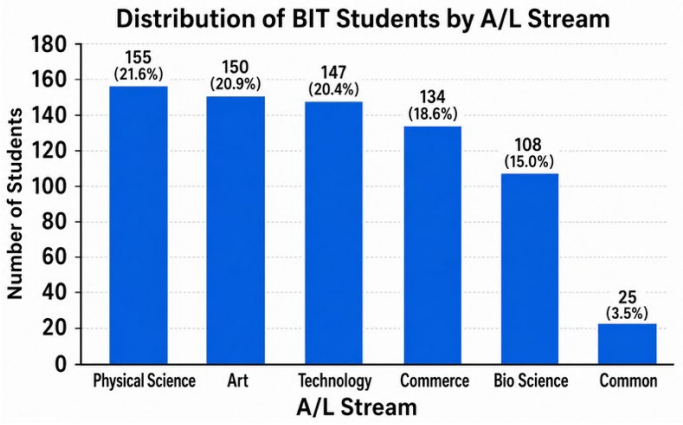


Figure 2: Distribution of BIT students by prior educational background, 23S2 cohort (N = 719).

Overall, 564 students (78.4%) entered from streams other than Physical Science. This finding is significant because it demonstrates that BIT participation is not concentrated among students from a conventional mathematics and physical science-oriented background. This contrasts with the published admission structures of several internal computing-related entry pathways, where eligibility depends on specified mathematics, physics or related subject preparation at A/L. By permitting entry from any A/L stream, the BIT external degree creates an alternative route into IT higher education for students from a wider range of prior educational backgrounds. Thus, the programme contributes to inclusivity by widening access beyond the educational pathways more commonly associated with internal computing-related programmes.

4.2 Geographical Reach

Geographical representation provides the strongest evidence of inclusive access within the BIT cohort. Students were represented from all nine provinces and 25 districts of Sri Lanka. Although the Western Province accounted for the largest share, it represented only 232 students (32.3%) of the cohort.

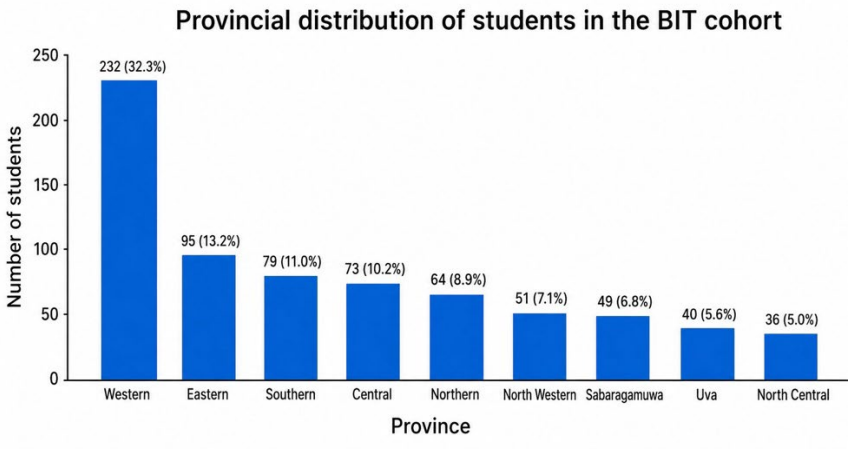


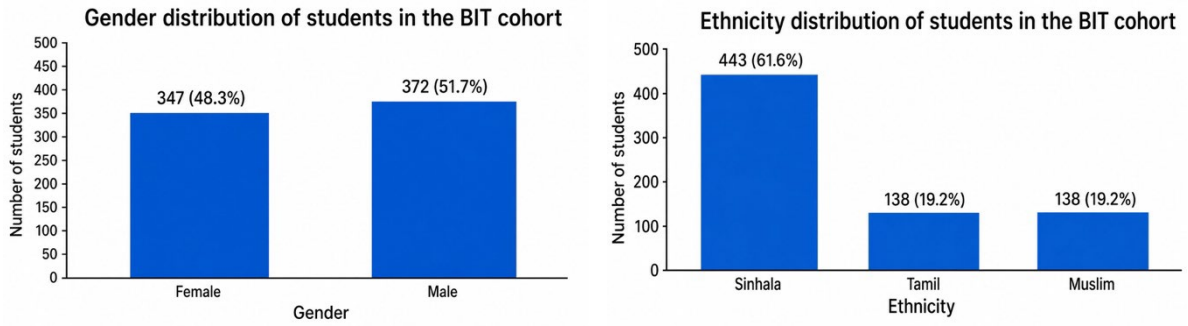
Figure 3: Distribution of BIT students by province of residence, 23S2 cohort (N = 719)

Within the cohort, 487 students (67.7%) resided outside the Western Province. This demonstrates that participation extends well beyond the province in which the University of Moratuwa is located. The district distribution further supports this finding. While Colombo was the most represented district, accounting for 123 students (17.1%), the remaining students were distributed across 24 other districts covering every province of Sri Lanka. This national distribution illustrates the potential of ODL to enable participation in IT higher education without requiring students to be located near the delivering institution.

These findings do not imply that all geographical or personal barriers have been removed. Although ODL reduces the need for continuous campus-based participation, students living at greater distances from the institution or managing other personal obligations may still experience challenges in sustaining study. Nevertheless, the BIT cohort demonstrates substantial national reach and provides strong evidence of geographically inclusive access.

4.3 Gender and Ethnic Representation

The gender distribution of the cohort was close to balanced. Of the 719 students, 372 were male and 347 were female, representing 51.7% and 48.3% respectively.



Figures 4 & 5: Gender and ethnic representation among BIT students, 23S2 cohort (N = 719)

This distribution is notable within an IT-related degree pathway. Gender participation patterns across fields of higher education in Sri Lanka have previously been identified as an important equity concern (Withanage, 2021). While the study does not include equivalent internal programme-level gender data, the substantial participation of both female and male students indicates that the ODL-based external pathway attracts learners across gender groups. The finding should be interpreted as evidence of representation, rather than as proof that gender-related barriers have been eliminated.

The cohort also included students from all three recorded ethnic categories. Sinhala students represented 61.6% of enrolment, while Tamil and Muslim students each represented 19.2%. Together, Tamil and Muslim students accounted for 38.4% of the cohort. This multi-ethnic participation, combined with representation from every province, demonstrates that access to the BIT programme extends across different communities and regions.

4.4 Age Representation

The cohort was predominantly composed of younger learners, with 576 students (80.1%) aged between 20 and 24 years at the beginning of BIT study. Nevertheless, 122 students (17.0%) were aged 25 years or above, including 52 students (7.2%) aged 30 years or above.

Age provides supporting rather than central evidence of inclusivity. The distribution does not indicate broad balance across age categories, since most students were within a conventional undergraduate-age range. However, the participation of students aged 25 years and above indicates that the BIT pathway also accommodates a meaningful group of adult learners.

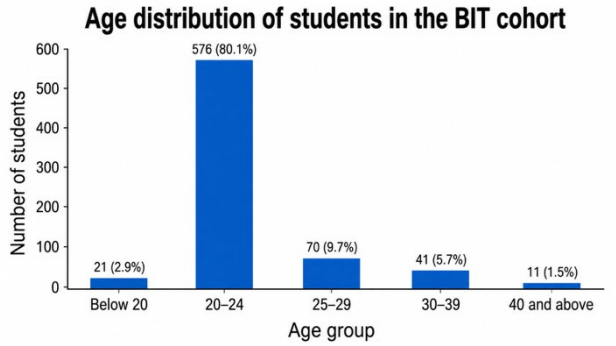


Figure 6: Age distribution of BIT students at enrolment, 23S2 cohort (N = 719)

5. Inclusive Early Progression and Academic Achievement

Having established the diversity of access within the full BIT cohort, this section considers whether inclusivity is also visible in early academic progression. The analysis does not compare achievement rates across demographic groups; rather, it examines whether students from varied backgrounds are represented among those achieving the first formal academic milestone of the programme.

5.1 Diploma-Level Milestone Achievement

As the 23S2 cohort was in Semester 4 at the time of analysis, Higher Diploma and full degree completion could not yet be assessed. Accordingly, this study uses achievement of the Diploma-level milestone, defined as completion of all Year 1 subjects, as an indicator of early academic progression. At the time of analysis, 157 of the 719 students (21.8%) had achieved this milestone. The 21.8% achievement rate should be interpreted as an interim early-progression indicator rather than a final measure of cohort completion. It represents students who had completed all Year 1 requirements either within the allocated first two semesters or after completing outstanding Year 1 subject requirements during Semester 3. Since the cohort was still active, more students may complete the remaining Year 1 requirements in later semesters and subsequently become eligible for the Diploma-level certificate. Therefore, the Diploma-level milestone is used to identify students who had achieved early completion of Year 1 requirements by the time of analysis, rather than to classify the remaining students as unsuccessful.

5.2 Diversity Among Diploma-Level Milestone Achievers

Diversity observed at enrolment remained evident among milestone achievers. Of the 157 students reaching the Diploma-level milestone, 87 were female (55.4%) and 70 were male (44.6%).

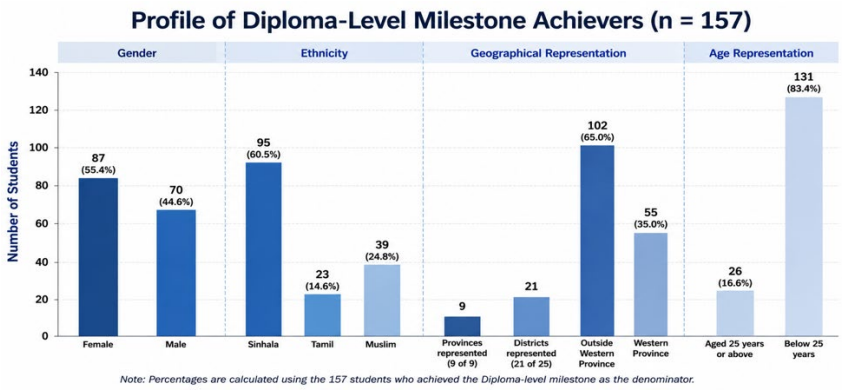


Figure 7: Profile of students achieving the Diploma-level milestone at the time of analysis (n = 157)

All three recorded ethnic groups were represented: 95 Sinhala, 39 Muslim and 23 Tamil students. Geographical representation was also sustained, with milestone achievers drawn from all nine provinces and 21 districts. Of these students, 102 (65.0%) resided outside the Western Province. Adult learners were represented to a lesser extent, with 26 students (16.6%) aged 25 years or above.

These results provide preliminary evidence that demographic and geographical diversity in the BIT programme is not limited to enrolment, but is also visible among students achieving early academic progression.

5.3 Selected Academic Achievement Indicator

Academic achievement among milestone achievers was examined through two foundational modules: Mathematics & Statistics for IT and Fundamentals of Programming. In these modules, 120 students (76.4%) and 121 students (77.1%), respectively, obtained grades within the A or B ranges. In addition, 33 milestone achievers (21.0%) had obtained C or D grades in O/L Mathematics. This suggests that early academic progression in the BIT programme was not restricted only to students entering with the strongest prior mathematics background.

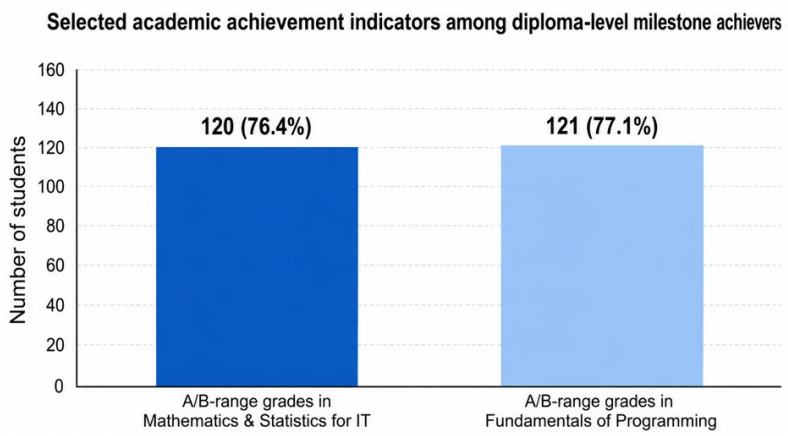


Figure 8: Selected academic achievement indicators among Diploma-level milestone achievers (n = 157)

5.4 Representation Summary

Table 3: Diversity at enrolment and among Diploma-level milestone achievers.

Dimension	Full Cohort, N = 719	Diploma-Level Milestone Achievers, n = 157	Interpretation
Gender	Female: 347, 48.3% Male: 372, 51.7%	Female: 87, 55.4% Male: 70, 44.6%	Both genders were substantially represented at enrolment and among milestone achievers.
Ethnicity	Sinhala: 443, 61.6% Tamil: 138, 19.2% Muslim: 138, 19.2%	Sinhala: 95, 60.5% Tamil: 23, 14.6% Muslim: 39, 24.8%	All recorded ethnic groups were represented among students achieving early progression.
Province	9 provinces represented	9 provinces represented	National representation remained visible among milestone achievers.
District	25 districts represented	21 districts represented	District-level diversity remained visible among milestone achievers.
Residence outside Western Province	487, 67.7%	102, 65.0%	Early progression was visible among students outside the institution's province.
Age 25 and above	122, 17.0%	26, 16.6%	Adult learners remained represented among milestone achievers.
A/L stream	6 streams represented	6 streams represented	Students from all recorded A/L streams achieved the Diploma-level milestone.

Table 3 shows that the diversity observed at enrolment was also visible among students achieving the Diploma-level milestone. Both female and male students, all recorded ethnic groups, students from all provinces, students from outside the Western Province, adult learners and students from all recorded A/L streams were represented among milestone achievers. This does not measure group-specific achievement rates, but it demonstrates that early progression was not limited to a single demographic, geographical or educational-background category.

6. Discussion

6.1 Broadening Access to IT Higher Education

The findings indicate that the BIT external degree widens access to IT higher education, particularly through the diversity of students' prior educational backgrounds. More than three-quarters of the cohort entered from A/L streams other than Physical Science, including Arts, Commerce, Technology and Bio Science. In contrast to internal computing-related pathways that commonly require mathematics or physics related A/L preparation, BIT enables entry from any A/L stream subject to its O/L Mathematics requirement. The programme therefore complements conventional internal provision by creating an IT education pathway for a broader academic population. This interpretation is limited to differences in entry structures and access pathways and it does not imply that the present study compares academic outcomes between BIT students and internal-degree students.

6.2 National Reach and Diverse Participation

Geographical reach provides the strongest evidence of inclusivity in the programme. Students were represented from all nine provinces and 25 districts, while 67.6% of those with valid residence data lived outside the Western Province. This demonstrates the relevance of ODL in extending access beyond learners located near the university. The same geographic spread remained visible among students achieving the Diploma-level milestone. Milestone achievers represented all nine provinces and 21 districts, with 65.0% residing outside the Western Province. This suggests that national reach is evident not only at enrolment, but also among students achieving early academic progression.

6.3 Gender and Ethnic Representation

The BIT cohort also demonstrated broad participation across gender and ethnicity. Female students represented 48.3% of enrolment, while male students represented 51.7%, indicating near-balanced gender participation in an IT-related degree pathway. Female and male students were also both substantially represented among Diploma-level milestone achievers. Students from all three recorded ethnic groups were represented in the cohort, with Tamil and Muslim students together accounting for 38.4% of enrolment. Furthermore, Sinhala, Tamil and Muslim students were all represented among those achieving the Diploma-level milestone. These findings indicate that participation and early progression in BIT extend across multiple recorded ethnic communities.

6.4 Beyond Enrolment: From Access to Early Progression

Inclusivity should be reflected not only in enrolment, but also in students' ability to make academic progress. As the cohort had not yet reached eligibility for full degree completion, the Diploma-level milestone provides an appropriate early indicator. Diversity remained present among milestone achievers, while more than three-quarters achieved grades in the A or B ranges in both selected foundational modules. These results provide initial evidence that wider access through BIT is accompanied by meaningful early progression. The importance of supporting successful progression in ODL has also been identified in studies involving Sri Lankan open and distance higher education provision (Zuhairi, Karthikeyan and Priyadarshana, 2020).

The analytical summary of enrolment and milestone-achiever profiles strengthens the paper's inclusivity argument. Diversity was not only evident at entry; it also remained visible among students achieving the first formal academic milestone. This is important because inclusive access alone does not necessarily imply inclusive progression. In the present cohort, students from varied gender, ethnic, geographical, age and educational-background categories were represented among those who achieved the Diploma-level milestone. However, the paper does not interpret these patterns as group-specific achievement rates, as detailed demographic comparisons are reserved for a separate focused study.

6.5 Implications for ODL Provision

The BIT experience suggests that ODL can complement conventional internal higher education by widening entry routes, supporting nationally distributed participation and enabling early progression among diverse learners. For institutions seeking scalable approaches to inclusive technology education, the findings highlight the value of flexible admission pathways together with continued monitoring of student progression and academic achievement.

7. Limitations and Future Research

This study examines an active cohort and therefore assesses enrolment diversity and Diploma-level early progression rather than final degree completion. Its comparison with internal degree programmes is contextual, based on published admission conditions rather than equivalent student-level data. In addition, the dataset does not capture personal circumstances that may shape access and progression, such as employment, family responsibilities, disability or reasons for selecting ODL. Future research should follow the cohort through Higher Diploma and degree completion, examine Diploma-level milestone achievement across demographic groups, incorporate students' experiences of access and progression, and conduct direct comparisons with internal degree cohorts where comparable data become available.

8. Conclusion

This study examined inclusivity in the BIT external degree programme offered through ODL by CODL, University of Moratuwa, using data from the 719 students of the 23S2 intake. The programme demonstrated strong geographical and educational-background diversity: 78.4% of students entered from A/L streams other than Physical Science, learners represented all nine provinces and 25 districts, and more than two-thirds of the students lived outside the Western Province. Near-balanced gender participation and representation from all recorded ethnic groups further demonstrate broad participation within the programme.

Inclusivity was also visible in early progression. Although the cohort had not yet reached eligibility for final degree completion, 157 students had achieved the Diploma-level milestone by completing all Year 1 requirements. These students represented both genders, all recorded ethnic groups and every province, while selected foundational-module outcomes provide preliminary evidence of meaningful early academic progression. The comparison with internal pathways is contextual rather than student-level. Nevertheless, the findings indicate that the BIT external degree complements conventional provision by extending access to IT higher education for learners from broader educational backgrounds and geographically distributed communities.

Ethics Declaration

This study used anonymised institutional administrative data analysed at aggregate level. No personally identifiable student information is reported, and the analysis involved no direct interaction with participants or intervention in their learning. Ethical clearance was not required for this descriptive analysis under the applicable institutional requirements. Permission to use the anonymised programme data was obtained through the relevant institutional authority.

AI Declaration

Generative AI tools were used to assist with language refinement, structural organisation and editorial review of the manuscript. All research decisions, data interpretation, verification of results, selection of references and final approval of the manuscript remain the responsibility of the authors. No AI tool was used to generate, alter or fabricate the underlying institutional dataset or reported numerical findings.

References

Butcher, J. and Rose-Adams, J. (2015) 'Part-time learners in open and distance learning: Revisiting the critical importance of choice, flexibility and employability', Open Learning: The Journal of Open, Distance and e-Learning, 30(2), pp. 127–137. doi:10.1080/02680513.2015.1055719.

Jayatilleke, B.G. and Kulasekera, G.U. (2020) 'Forty years of distance education: Challenges and implications at the Open University of Sri Lanka', Open Praxis, 12(3), pp. 311–327. doi:10.5944/openpraxis.12.3.1122.

Stone, C. (2019) 'Online learning in Australian higher education: Opportunities, challenges and transformations', *Student Success*, 10(2), pp. 1–11. doi:10.5204/ssj.v10i2.1299.

Tait, A. (2013) 'Distance and e-learning, social justice, and development: The relevance of capability approaches to the mission of open universities', The International Review of Research in Open and Distributed Learning, 14(4), pp. 1–18. doi:10.19173/irrodl.v14i4.1526.

Withanage, D.P. (2021) 'Gender equality in education in Sri Lanka', *Sri Lanka Journal of Educational Research*, 19(1), pp. 48–69.

Yang, D., Tang, Y.M., Hayashi, R., Ra, S. and Lim, C.P. (2022) 'Supporting inclusive online higher education in developing countries: Lessons learnt from Sri Lanka's university closure', Education Sciences, 12(7), article 494. doi:10.3390/educsci12070494.

Zuhairi, A., Karthikeyan, N. and Priyadarshana, S.T. (2020) 'Supporting students to succeed in open and distance learning in the Open University of Sri Lanka and Universitas Terbuka Indonesia', Asian Association of Open Universities Journal, 15(1), pp. 13–35. doi:10.1108/AAOUJ-09-2019-0038.

An AI-Supported Simulation-Based Serious Game for Digital Health Management Education: Design and Formative Evaluation

Annika Hering[1a], Gülsen Otman[1a], Annika Viedenz[1] and Jan-David Liebe[1,2]

[1]Center for Medical, Social and Health Informatics (MESH), Osnabrück University of Applied Sciences, Osnabrück, Germany

[2]UMIT - Private University of Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

[a]Authors contributed equally to this paper

a.hering@hs-osnabrueck.de
g.otman@hs-osnabrueck.de
annika.viedenz@hs-osnabrueck.de
j.liebe@hs-osnabrueck.de

Abstract: Background Teaching digital transformation in healthcare management requires students to address complex organizational challenges, multiple stakeholder perspectives, and decision-making under uncertainty. Traditional case-based teaching provides limited experience of real-world decision-making complexity and feedback. Objectives This study presents the design and formative evaluation of the first version of an artificial intelligence (AI)-supported digital serious game (DSG) in Digital Health Management (DHM) education. It examined students' evaluations of the integrated management methods, generative AI (GenAI)-supported stakeholder interactions, educational design, and AI-generated model solutions. Methods Design of the DSG was informed by the co.LAB generic framework and resulted in a scenario-based learning environment integrating management methods, collaborative problem-solving, GenAI-supported stakeholder interactions, and AI-generated model solutions. The formative evaluation used an embedded mixed-methods design across bachelor's and master's degree courses at two higher education institutions. Twenty-one students participated in the DSG, of whom 15 completed the interim evaluation, while the number of method-specific evaluations varied across learning activities. Qualitative reflections were collected at group level. Results The browser-based DSG combined healthcare digitalization case scenarios, sequential management methods, GenAI-supported stakeholder interactions, collaborative problem-solving, and AI-generated model solutions. The management methods received consistently positive ratings (M = 4.19–4.55/5). Students also positively evaluated the GenAI-supported stakeholder interactions (M = 5.95/7), their usability (M = 4.18/5), the teaching-learning experience (M = 4.98/6), and the AI-generated model solutions (M = 3.98/5). Internal consistency ranged from acceptable to excellent (α = .72 to .91), while Spearman-Brown coefficients for two-item subscales ranged from .66 to .92. The qualitative findings suggest that the AI-generated model solutions supported validation, extension, and critical comparison of students' own analyses. The Net Promoter Score (NPS) was 40. Discussion The findings suggest that GenAI can support scenario-based learning in DHM when GenAI-supported stakeholder interactions and feedback are embedded in a coherent, collaborative learning design. Future studies should use a prospectively defined evaluation framework to compare different forms of AI-generated feedback and assess reflection, objective learning outcomes, competency development, and transfer in larger samples. Conclusion The study provides an initial design and empirical basis for the further development of AI-supported DSGs in DHM education. The findings indicate that the AI-supported components may support perspective taking, comparison, reflection, and method-based decision-making within a coherent learning design.

Keywords: Digital Health Management, Digital Transformation, Serious Games, Education, Artificial Intelligence

1. Background

Higher education is increasingly challenged to prepare students for professional contexts characterized by digital transformation, uncertainty and complex decision-making. In response, artificial intelligence (AI), simulation-based learning and digital serious games (DSGs) are gaining importance as didactic approaches that can support more adaptive, experiential and practice-oriented forms of learning.

AI has been applied in education in various ways, including automated assessment and feedback, personalized learning plans, adaptive instructional support and intelligent tutoring systems (Chen *et al.*, 2020). In particular, AI-driven adaptive learning systems can dynamically adapt learning content, feedback, and learning pathways to individual learners' needs and preferences, thereby supporting personalized learning experiences (Strielkowski *et al.*, 2025). Generative AI (GenAI) extends these possibilities by generating context-sensitive content and feedback in response to specific learning situations. However, its educational value depends on both technical capabilities and pedagogical integration into the learning environment. AI-generated feedback can therefore be embedded in hybrid feedback processes in which it is reviewed, supplemented, and contextualized through human judgment (Banihashem *et al.*, 2026). This is particularly important in competence-oriented higher education, where feedback should not only indicate deficiencies but also stimulate reflection, reasoning, and transfer to complex professional situations.

This need for pedagogical integration is also central to the design of serious games. Previous research has emphasized that the educational effectiveness of serious games depends on integration into a broader learning design that combines gameplay with preparation, facilitation, collaboration, and reflection (Jaccard *et al.*, 2022). DSGs and other experiential learning approaches can connect academic learning with professional practice. In operations and supply chain management education, experiential learning activities, including DSGs and university challenges, were associated with a closer alignment between students' competencies and professional role profiles. Students reported gains in soft skills such as leadership, teamwork and adaptability (Fantozzi and Schiraldi, 2026). Moreover, game-based learning environments can enable learners to experience and reflect on the consequences of their own actions. In serious games, this can support the development of professional, methodological, social, and personal competencies and contextual understanding (Schwägele, 2015).

Such approaches appear particularly relevant in management contexts. Managers face conflicting objectives, limited resources, and difficult-to-predict long-term consequences. Traditional teaching provides limited opportunities to experience the complexity of real-world decision-making. Simulation-based learning is particularly relevant for management education because DSGs expose learners to such situations and support communication, assertiveness, and networked strategic thinking. Such games should be grounded in educational requirements and domain-specific management challenges rather than the game format alone. Relevant design elements include a realistic organizational scenario, differentiated management roles, interdependent decision situations, and a structured didactic framework including preparation, facilitation, theoretical input and discussion or presentation phases (Häberle *et al.*, 2006). In project management education, participation in a simulation-supported course was associated with higher scores in an external certification exam, particularly in competence areas related to teamwork, change management and project design (Opas and Ćwil, 2026).

In Digital Health Management (DHM), these requirements become even more pronounced. Managers require a broad set of capabilities and competencies that enable data-driven decision-making, digital health literacy, planning and managing digital technologies, and leading organizational transformation processes (Brommeyer and Liang, 2022). At the same time, digital health innovations are embedded in complex healthcare systems and are shaped by technology, organizational readiness, integration into existing structures, management and maintenance responsibilities, financing, and the wider system context (Hellstrand Tang *et al.*, 2024). Students therefore need learning environments that allow them to apply management methods in realistic situations, rather than only learning about such methods in abstract or static formats. The integration of AI may further enhance these learning environments by supporting dynamic stakeholder interactions and providing context-sensitive feedback through AI-generated model solutions. Against this background, this study presents the design and formative evaluation of an AI-supported DSG for DHM education. The evaluation was guided by the following research questions (RQs):

> RQ1: How do students evaluate the instructional implementation, applicability, and perceived contribution of management methods?
>
> RQ2: How do students evaluate the GenAI-supported stakeholder interactions, their usability, and the educational design of the DSG?
>
> RQ3: How do students evaluate the AI-generated model solutions as a form of feedback and learning support?

2. Methods

2.1 Design and Development of the Digital Serious Game

The first version of the DSG was developed to provide a practice-oriented learning environment for DHM education. The design process was informed by the co.LAB framework (Jaccard *et al.*, 2021). Typical healthcare digitalization management challenges were translated into scenario-based cases, and management methods were sequenced from current-state analysis through target-concept development to implementation planning. The case scenarios, including the organizational context and stakeholder constellations, were generated with the support of ChatGPT (OpenAI, 2026) and subsequently reviewed and revised by the research team. The resulting learning design was implemented as a browser-based application integrating case-specific content, role-based GenAI-supported stakeholder interactions, task instructions, progress-dependent access, and AI-generated model solutions.

2.2 Formative Evaluation Design and Participants

The formative evaluation used an embedded mixed-methods design. Group-based qualitative reflections complemented quantitative evaluations by providing additional insight into how students compared and interpreted the AI-generated model solutions. Reporting of the quantitative survey component followed the Checklist for Reporting of Survey Studies (CROSS) (Sharma *et al.*, 2021).

Data collection was embedded in master's and bachelor's degree courses in the field of DHM at Osnabrück University of Applied Sciences (HSO) and Medical School Hamburg (MSH) between April and June 2026. The study population comprised students enrolled in these courses. All these students were eligible to participate, and no exclusion criteria were applied. Participants therefore constituted a convenience sample of the participating course cohorts. Quantitative data was collected at the individual level. In contrast, qualitative reflections on the AI-generated model solutions were collected at the group level because the cases had been collaboratively developed and discussed within student teams. Depending on the method, approximately 120–180 minutes were allocated to each methodological activity, followed by 20–30 minutes for subsequent discussion and reflection.

In total, 21 students participated in the DSG, including 13 at HSO and eight at MSH. As participation in the evaluation was voluntary and integrated into ongoing coursework, not all participants completed every measurement point. Responses were collected anonymously, therefore reasons for non-participation could not be systematically recorded. Of the 21 students, 15 completed the interim evaluation, corresponding to a response rate of 71.4%.

2.3 Instruments and Data Collection

To address RQ1, a self-developed method evaluation instrument comprising twelve items was used. Items were rated on an agreement scale, with higher values indicating stronger agreement.

RQ2 was examined through an interim evaluation conducted after the process analysis method. The evaluation used modified items from established instruments: nine items from the Persona Perception Scale (PPS) to assess GenAI-supported stakeholder interactions (Salminen *et al.*, 2020), seven items from the Chatbot Usability Scale (BUS-11) to evaluate usability aspects (Borsci *et al.*, 2023), and 13 items (ten multi-item scale items and three additional single items) from the questionnaire developed by the Centre for Management Simulations (CMS) to assess the educational design of the DSG (Alf and Trautwein, 2023). Two researchers independently translated the PPS and consolidated their versions. Instrument items were selected based on their relevance and adapted to the educational context. Reliability for the instruments was examined post hoc (Eisinga *et al.*, 2013).

Regarding RQ3, a self-developed nine-item scale and group-based free-text reflections assessed the AI-generated model solutions. Groups compared each model solution with their own artifact and documented a joint assessment.

The interim evaluation assessed willingness to recommend the DSG using the 11-point Net Promoter Score (NPS) item (Reichheld, 2003). Student characteristics were also collected, including prior experience with serious games, professional background, experience in developing or implementing digitalization strategies or projects in healthcare, and initial interest in the DSG.

Surveys were administered via LimeSurvey (LS) through a DSG link available only to enrolled students. The LS questionnaire was reviewed by two students to assess comprehensibility and technical functionality prior to data collection, but the instruments were not subjected to formal or psychometric pretesting. Data was collected after predefined learning activities, so assessments referred directly to experienced methods. Table 1 provides an overview of the instruments used.

Table 1: Overview of the instruments

No.	Instrument	Timepoint of Data Collection	Construct Definition	Items (n)	Response Scale
1	Method evaluation	After each management method	Perceived instructional support, clarity and structure of the method implementation, understanding of its purpose, independent applicability, and contribution to analyzing and planning the digitalization project	12	5-point scale (Strongly disagree – Strongly agree)

No.	Instrument	Timepoint of Data Collection	Construct Definition	Items (n)	Response Scale
2	Modified PPS	Interim evaluation	Perceived realism, consistency and usefulness of the GenAI-supported stakeholder interactions	9	7-point scale (Completely disagree – Completely agree)
3	Modified BUS-11	Interim evaluation	Usability of the GenAI-supported stakeholder interactions	7	5-point scale (Strongly disagree – Strongly agree)
4	Modified CMS Teaching-Learning Experience	Interim evaluation	Perceived structure, comprehensibility, realism, feedback opportunities, learning support and motivation	10	6-point scale (Strongly disagree – Strongly agree)
4a	CMS Pace	Interim evaluation	Perceived pace of the course	1	3-point scale (a bit too fast – just right – a bit too slow)
4b	CMS Time	Interim evaluation	Perceived time available per method	1	3-point scale (a bit too short – just right – a bit too long)
4c	CMS Scope	Interim evaluation	Perceived scope of joint discussion phases	1	3-point scale (a bit too short – just right – a bit too long)
5	AI-generated model solutions	After each management method	Validation of students' own work, extension and refinement of students' analyses, and critical reflection and contextual adaptation	3	Free text
		Interim evaluation	Perceived quality, completeness, comprehensibility and reflective value of the AI-generated model solutions	9	5-point scale (Strongly disagree – Strongly agree)
6	NPS	Interim evaluation	Willingness to recommend the DSG	1	11-point scale (0–10)

Note: PPS = Persona Perception Scale; BUS-11 = Chatbot Usability Scale; CMS = Centre for Management Simulations; DSG = Digital serious game.

2.4 Data Analysis

Quantitative data was analyzed using R (v. 4.4.1). Descriptive analyses included mean scores (M) and standard deviations (SD) with scale scores calculated as the mean of the respective items. Responses were included in the dataset if all items of the respective measure had been completed. Due to variation in participation across evaluation points, the number of valid responses was reported separately for each analysis. As students could evaluate multiple management methods, the method-specific data included repeated observations. Analyses were therefore descriptive and method-specific. Cronbach's α or Spearman-Brown was calculated for the PPS, BUS-11, CMS Teaching–Learning Experience, and for the AI-generated model solutions instrument. Reliability was not assessed within the management method evaluation ratings because the items represented distinct evaluative criteria rather than interchangeable indicators of a single latent construct. Nevertheless, an equally weighted mean score was calculated as a descriptive summary of students' overall evaluation across these criteria. The NPS was calculated as the proportion of promoters (9–10) minus detractors (0–6) and ratings of 7–8 were classified as passive.

Free-text responses regarding the AI-generated model solutions were analyzed using qualitative content analysis following Kuckartz and Rädiker (2024). Initial main categories were derived deductively from the reflection tasks, while additional subcategories were developed inductively from the data. A meaningful quote within a group-based reflection response constituted the coding unit. The material was coded by two researchers, and coding was iteratively refined during the analysis. Ambiguous quotes were discussed until consensus was reached.

3. Results

3.1 Learning Design and Case Studies of the Digital Serious Game

The DSG combined scenario-based challenges, role-based stakeholder interactions, rule-based sequential task progression, collaborative decision-making, method-based problem-solving, and structured feedback and reflection. The first version implemented simulation-based learning through qualitative, scenario-based decision-making and GenAI-supported stakeholder interactions rather than numerical simulation of system

consequences. The case studies addressed the introduction of (1) a patient portal, (2) a digital medication management system, (3) a telemonitoring program for chronic heart failure, (4) a mobile documentation and route-planning system in long-term care, (5) an AI-supported risk-identification system in child and youth welfare, and (6) a digital platform for cross-sectoral case coordination. Although fictionalized for educational purposes, they were based on typical digitalization projects in healthcare organizations.

At the beginning of the intervention (1), the lecturer introduced the learning objectives, case studies, and task sequence. Each group worked on one case throughout the intervention. The learning path followed three guiding questions: “Where do we stand?”, “Where do we want to go?”, and “How do we get there?”. These questions structured the gradual development of a method-based implementation strategy for the digital transformation project in healthcare. After familiarizing themselves with the selected case study and the assignment logic, students worked through the management methods. Figure 1 provides an overview of the learning design and instructional sequence of the DSG.

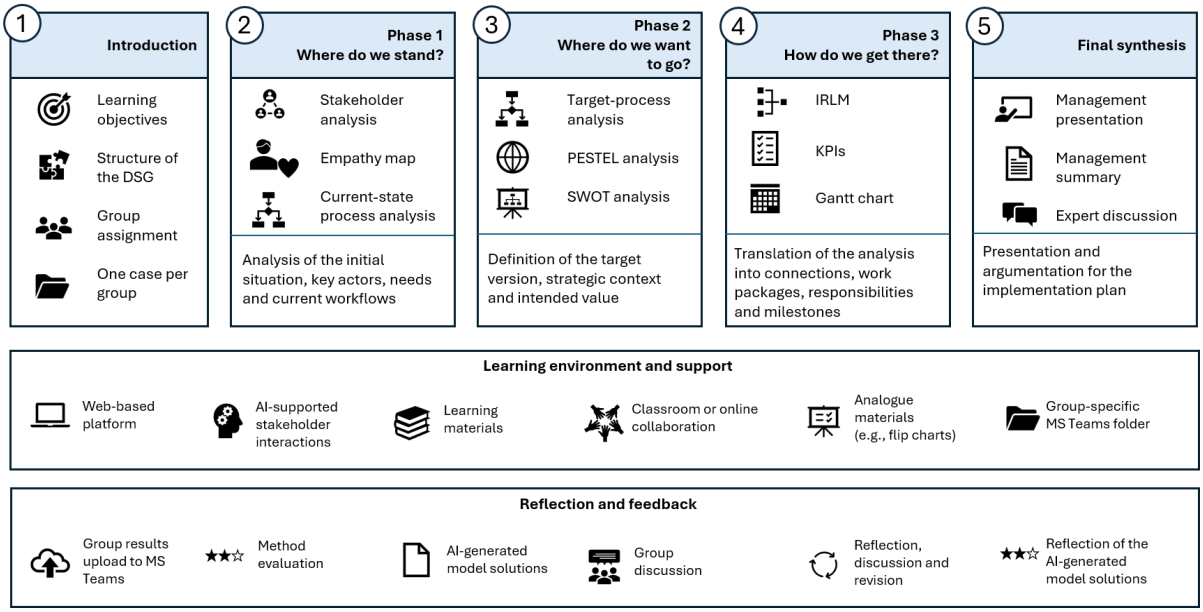


Figure 1: Learning design and instructional sequence of the GenAI-supported digital serious game (DSG)

In the first phase (2), students analyzed the initial situation, underlying problem, and contextual conditions of their case. They applied stakeholder analysis and empathy mapping and conducted a current-state process analysis to examine existing analogue or inefficient workflows.

In the second phase (3), students developed a target concept for the digitalization project. This included modeling a target-state process, analyzing external influencing factors through Political, Economic, Social, Technological, Environmental, and Legal (PESTEL) analysis, and synthesizing internal strengths and weaknesses as well as external opportunities and threats through Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. This phase supported students in defining the intended value of the project for the healthcare organization and its stakeholders.

In the third phase (4), students translated their previous analyses into an implementation logic and a project management plan. They developed an Implementation Research Logic Model (IRLM), including key performance indicators (KPIs), to specify relevant determinants, implementation strategies, and intended outcomes. Based on these results, students created a Gantt chart in which they derived work packages, activities, responsibilities, and milestones for their case.

For the final synthesis (5), students integrated the results of the applied methods and their Gantt chart into a management presentation and prepared a summary for senior management. During their presentation they responded to questions from the lecturer about their implementation plan and justified the approach they had taken. The DSG formed the digital component of a blended learning arrangement. Students used it to access case information, GenAI-supported stakeholder interactions, learning materials, task-specific instructions, and interactive checklists. The management methods were mainly applied collaboratively in the physical classroom using analogue materials such as flip charts and sticky notes. The resulting artifacts were documented and stored

in group-specific Microsoft Teams (MS Teams) folders. After applying each management method, students discussed their results in class and evaluated the respective methods.



Figure 2: "Empathy map" as exemplary project artifact of the digital serious game (DSG)

3.2 Technical Implementation

The DSG was implemented as a browser-based web application using client-server architecture. This supported flexible integration into higher education teaching and location-independent browser access. A central component of the platform was the GenAI-supported stakeholder simulation. Students could interview virtual stakeholders through a conversational interface using natural-language questions. Stakeholder responses were generated via the Anthropic Messages API using Claude 3.5 Sonnet (Anthropic, 2026). Stakeholder-specific profiles of responsibilities, interests, objectives, and concerns informed the prompting process. Students could save stakeholder statements as favorites, add notes, and revisit conversations for subsequent project work.

Progress was tracked within the DSG, and subsequent sections were unlocked only after all tasks within the current method had been completed. This mechanism guided students systematically through the sequence of management methods. Figure 3 illustrates the interface for the use case of an AI-supported risk-identification system in child and youth welfare.

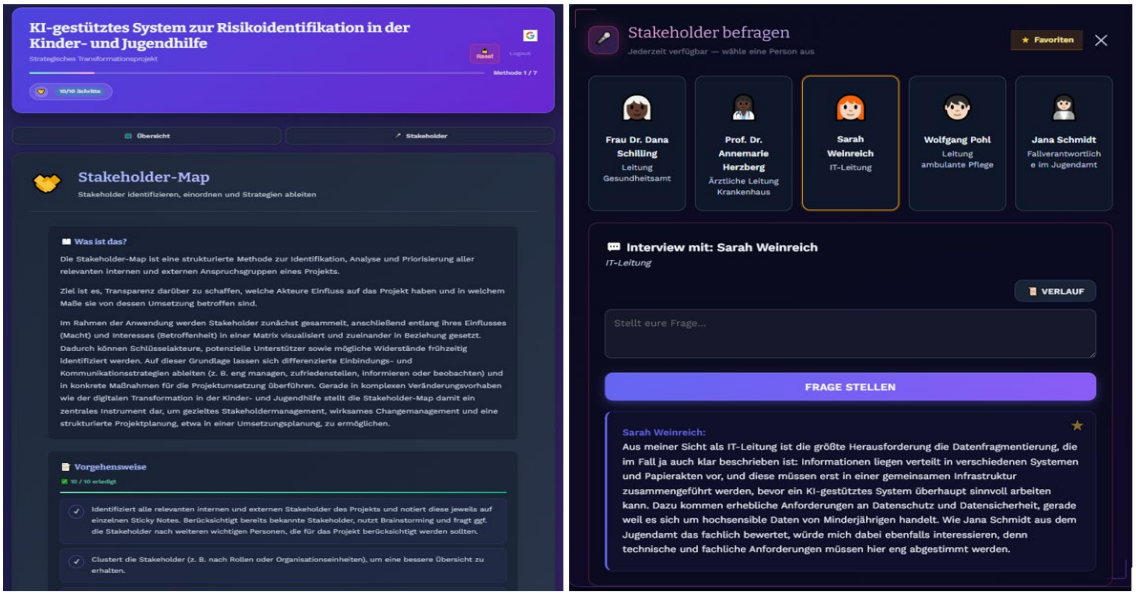


Figure 3: Exemplary "stakeholder analysis" interface with access to GenAI-supported stakeholders in the digital serious game (DSG)

For reflection and self-assessment, AI-generated model solutions were produced for each management method within the platform using Claude (Anthropic, 2026). The model solutions drew on each case and stakeholder interviews and followed the method's predefined steps and guiding questions. Although method-structured, the results were primarily textual rather than visual method-specific artifacts. They were not presented as single

correct answers but as possible approaches against which students could critically compare their own results and reflect on alternative implementation strategies. After completing a method and uploading their results to MS Teams, the student groups received the corresponding model solution from the lecturer.

3.3 Evaluation of Management Methods

Students evaluated all management methods positively, with mean ratings above 4.0 on the five-point agreement scale. SWOT analysis received the highest mean rating (M = 4.55), followed by the empathy map (M = 4.34). PESTEL analysis received the lowest mean rating, although its evaluation was still positive overall (M = 4.19). The results are presented in Table 2. Detailed item-level results are available on the project website (Hochschule Osnabrück, 2026).

Table 2: Descriptive analysis of management methods; n indicates responses summarized across HSO and MSH

Management Method	Responses (n)	M (SD)
Stakeholder analysis	20	4.26 (0.84)
Empathy map	18	4.34 (0.86)
Process analysis (current-state and target-state analysis)	33	4.24 (0.79)
PESTEL analysis	16	4.19 (0.79)
SWOT analysis	14	4.55 (0.74)
IRLM and KPIs	3	4.22 (0.64)

Note: M = Mean score; SD = Standard deviation. Rating on a five-point agreement scale. For process analysis, evaluations of current-state and target-state analyses were pooled.

3.4 Interim Evaluation

3.4.1 Student Characteristics

As shown in Table 3, the interim evaluation included seven HSO and eight MSH students. Most participants had no prior experience with serious games (11/15). Professional backgrounds were heterogeneous across both institutions, with medicine (3/15), healthcare administration, and social services (2/15 each) representing the most common fields. Experience with healthcare digitalization projects was more common among MSH students (7/8) than HSO students (1/7). The initial interest in the DSG was generally positive (12/15).

Table 3: Student characteristics of the interim evaluation; n = 15

Student Characteristics	HSO (n)	MSH (n)	Total (N)
Prior Participation in Serious Games			
Not at all	7	4	11
1-2 times	-	2	2
3-4 times	-	2	2
More than 4 times	-	-	-
Professional Background			
Medicine (e.g. doctors)	3	-	3
Care (e.g., healthcare and nursing, care for the elderly)	-	1	1
Therapeutic professions	-	1	1
Health Management / Health Economics	-	1	1
Administration in the healthcare sector	-	2	2
Computer Science / IT	-	1	1
Social services (e.g., social work, case management)	2	-	2
Business Administration / Management	-	1	1

Student Characteristics	HSO (n)	MSH (n)	Total (N)
Other	2	1	3
Experience with Digitalization Projects in Healthcare			
Yes	1	7	8
No	6	1	7
Initial Interest in the DSG			
Very low	-	-	-
Low	-	-	-
Rather low	3	-	3
Rather high	3	2	5
High	1	4	5
Very high	-	2	2

Note: IT = Information technology; DSG = Digital serious game.

3.4.2 Instrument Results

Students evaluated the GenAI-supported stakeholder interactions, their usability, the CMS teaching-learning experience, and the AI-generated model solutions positively. The findings are presented in Table 4. Detailed item-level results are available on the project website (Hochschule Osnabrück, 2026).

Table 4: Descriptive analysis and reliability of the quantitative interim evaluation; individual rating per student summarized across HSO and MSH (n = 15 for each instrument)

Instrument	Dimension	M (SD)	Reliability
PPS	-	5.95 (1.01)	.89
BUS-11	-	4.18 (0.70)	.72
CMS teaching-learning experience	-	4.98 (1.03)	-
	Comprehensibility	5.17 (0.75)	.66*
	Realism	5.27 (0.94)	.87*
	Teaching design	5.11 (0.80)	.80
	Learning outcomes	4.53 (1.29)	.91
CMS Pace	-	2.00 (0.38)	-
CMS Time	-	1.67 (0.49)	-
CMS Scope	-	1.80 (0.41)	-
AI-generated model solutions	-	3.98 (0.66)	-
	Completeness	4.10 (0.55)	.82*
	Traceability	3.90 (0.55)	.80*
	Justification	3.84 (0.67)	.74
	Overall evaluation	4.13 (0.82)	.92*

Note: M = Mean score; SD = Standard deviation. Reliability for the CMS teaching-learning experience and AI-generated model solutions is reported at the subscale level; reliability was not calculated for CMS Pace, CMS Time, or CMS Scope because each construct was assessed using a single item. Unmarked coefficients represent Cronbach's α, coefficients marked with an asterisk represent Spearman-Brown coefficients.

The dimension "Realism" was the highest-rated CMS subscale (M = 5.27), while "Completeness" (M = 4.10) and "Overall evaluation" (M = 4.13) received the highest ratings among the AI-generated model solutions. Course pace was rated as broadly appropriate (M = 2.00), whereas the time available per method tended to be perceived as somewhat short (M = 1.67). The scope of the joint discussion phases was rated as largely appropriate, with a slight tendency toward being too short (M = 1.80).

The NPS stood at 40. Overall, 53.3% (n = 8) of the students were classified as “promoters”, 33.3% (n = 5) as “passives”, and 13.3% (n = 2) as “detractors”.

Following George and Mallery (2003), internal consistency for the instruments ranged from acceptable to excellent across the respective scales and subscales (α = .72–.91). For the two-item subscales of the CMS teaching-learning experience, Spearman-Brown coefficients ranged from .66 to .87, with Comprehensibility showing the lowest coefficient. For the two-item subscales of the AI-generated model solutions instrument, Spearman-Brown coefficients ranged from .80 to .92.

3.5 Reflection of AI-Generated Model Solutions

As shown in Table 5, the qualitative analysis was based on the three overarching reflection tasks of the AI-generated model solutions: validation of students’ own work, extension and refinement of students’ analyses, and critical reflection and contextual adaptation. The qualitative analysis comprised 343 coded meaning units. Confirmation was the most frequently coded function, followed by extension and concretization. Critical comparison, alternative perspectives, and the identification of omissions and blind spots occurred less frequently. All qualitative results are available on the project website (Hochschule Osnabrück, 2026).

Table 5: Reflection of AI-generated model solutions; group reflections are summarized over the management methods across HSO and MSH

Feedback Area	Feedback Function	Coded Quotes (n)	Illustrative Quotes
Validation of students’ own work	Confirmation	119	*“All details have been confirmed.”* (HSO, IRLM) *“Our list of stakeholders matches the model solution almost 100 %. The clusters we have selected (management vs operational users vs end users) also closely resemble those in the model solution.”* (MSH, Stakeholder analysis)
Extension and refinement of students’ analyses	Extension and concretization of analysis	108	*“The characters’ thoughts were more detailed.”* (HSO, Empathy map) *“The model solution is much more detailed.”* (MSH, PESTEL analysis)
	Identification of omissions and blind spots	27	*“We have not provided a detailed answer to Step 4.”* (HSO, Empathy map) *“We have not identified any key findings.”* (MSH, Process analysis)
Critical reflection and contextual adaptation	Alternative perspective	59	*"Further key stakeholders identified: patients, IT/software providers and payers (holistic approach)"* (MSH, Stakeholder analysis) *"We have chosen a different stakeholder […] For us, this was the more appropriate choice in order to gain two different perspectives."* (HSO, Empathy map)
	Critical comparison	30	*“In general, we find our solution more suitable because it is somewhat more detailed; […].”* (HSO, PESTEL analysis) *“Our visual presentation makes our solution much clearer.”* (MSH, SWOT analysis)

Note: Coded quotes (n) refer to the number of coded meaning units; a single free-text response could contribute to more than one coded quote.

4. Discussion

This study presents the design and formative evaluation of an AI-supported DSG for DHM education under authentic teaching conditions. By combining GenAI-supported stakeholder interactions and AI-generated model solutions with established management methods, the study examines a configuration for which empirical evidence remains limited.

With respect to RQ1, students positively evaluated the instructional implementation, applicability, and perceived contribution of the management methods (M = 4.19–4.55/5). The item-level evaluations further indicate that instructions and methodological steps were generally perceived as clear and comprehensible and that students largely felt able to apply the methods independently. However, ratings were somewhat lower for translating individual method results into subsequent project-planning activities, particularly the derivation of actions for the Gantt chart, and for the perceived applicability of some methods in professional practice. This suggests that while the methods were generally perceived as well implemented within the DSG, their connection to subsequent implementation planning could be made more explicit.

Regarding RQ2, students positively evaluated the GenAI-supported stakeholder interactions, usability, and educational design. Realism was the highest-rated CMS dimension, suggesting that the learning environment was perceived as practice-oriented and plausible, whereas perceived learning outcomes were comparatively lower. Course pace was considered appropriate, but time per method and, to a lesser extent, discussion phases tended to be perceived as too short. Additional time for method application and reflection may therefore strengthen the learning design. The findings indicate that GenAI-supported stakeholder interactions can be integrated into a coherent DHM learning design and were positively perceived by students. The positive NPS further indicates overall acceptance of the first DSG version.

In relation to RQ3, students evaluated the AI-generated model solutions positively as a form of feedback and learning support (M = 3.98/5). Within this evaluation, completeness and the overall assessment received somewhat higher ratings, whereas traceability and justification were rated slightly lower. Students used the model solutions to validate and extend their own work, while they also identified alternative perspectives, omissions, and opportunities for critical comparison. Rather than being treated as authoritative answers, the model solutions therefore served as reference points for comparison and reflection. This feedback function aligns with the role of reflection and debriefing in serious games (Crookall, 2010) and suggests that AI-generated model solutions may support reflective processes. At the same time, confirmation and extension occurred more frequently than critical comparison, indicating that more critical forms of engagement were less prominent in the reflections.

Students' critical engagement with the model solutions further suggests that the pedagogical design of AI-generated feedback, including how directive, staged, and open it is, may be as important as its availability. This interpretation is consistent with recent research showing that the design of AI-generated feedback can influence students' revision behaviour (Alsaiari *et al.*, 2026). Future iterations could move beyond primarily textual model solutions by allowing students to upload their artifacts and receive interactive, artifact-specific feedback. Instead of presenting a complete reference solution, the AI could identify gaps, ask critical follow-up questions, and prompt students to justify or rethink their results.

Several limitations should be considered. The study was conducted in two university courses with a small convenience sample, and voluntary participation resulted in incomplete measurement points. The findings are therefore exploratory and not representative of DHM students more broadly. Participants differed in professional background, prior digitalization experience, and exposure to serious games, particularly between institutions, which may have influenced evaluations. Response numbers also varied across management methods, with very few evaluations for IRLM and KPIs. The adapted and newly developed instruments were not formally validated or psychometrically pretested, and the prior review by two students addressed only comprehensibility and technical functionality. Reliability estimates should therefore be interpreted cautiously. The study assessed perceptions rather than objective learning outcomes or competency development. Because the AI-supported components were embedded in the overall learning design, their specific effects cannot be isolated. Their quality also depended on the stakeholder profiles, prompts, and case information used. Finally, the model solutions were not reviewed before release; students' reflections therefore indicate engagement with the generated content but do not constitute formal validation of its factual accuracy.

Future research could examine whether and how AI-supported components in serious games affect learning outcomes and competency development using larger samples, and whether more explicit prompts for questioning and contextualizing AI-generated feedback strengthen critical reflection. DSGs should incorporate documented practice-based cases to increase contextual depth and organizational realism. A prospective evaluation framework should define validated instruments, outcome measures, measurement points, and procedures for assessing the contribution of individual DSG components.

5. Conclusion

The design of the AI-supported DSG demonstrates how scenario-based healthcare digitalization cases, sequential management methods, GenAI-supported stakeholder interactions, collaborative problem-solving, and AI-generated model solutions can be integrated into a coherent learning environment for DHM education. The educational value of the AI components appears to lie not in providing authoritative answers but in supporting comparison, reflection, and method-based decision-making. The study provides an empirically informed basis for the further development of AI-supported DSGs in DHM education.

Acknowledgements

The authors would like to thank the students who participated in the study and contributed their perspectives during the DSG.

Funding

This study is part of the research project "SMiLeH – Strategic Management in Learning Health Systems" (Project No. 3004-0241), funded by the Stiftung Innovation in der Hochschullehre (StIL). The funder had no role in the study design, data collection, analysis, interpretation of data, or in the writing of the manuscript.

Ethics Declaration

An ethics waiver was issued by the Ethics Committee of Osnabrück University of Applied Sciences (Approval No. W_HSOS/2026/1/3), 21 April 2026, Prof. Dr. Frank Ollermann. All students were informed about the purpose and voluntary participation in the surveys and provided informed consent prior to participation.

AI Declaration

The authors used DeepL (DeepL, 2026) to support language refinement and ChatGPT (OpenAI, 2026) for structural revision during the development of the manuscript. The AI-generated suggestions were used to improve clarity, coherence, and wording. All suggestions were critically reviewed, edited, and approved by the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Data Availability Statement

The complete survey items, quantitative results, and qualitative findings are available on the project website (Hochschule Osnabrück, 2026) or from the authors upon request.

Author Contributions

AH: Conceptualization, Methodology, Writing – Original Draft. GO: Conceptualization, Validation, Writing – Original Draft. AV: Formal analysis, Data Curation, Writing – Review & Editing. JDL: Conceptualization, Supervision.

References

Alf, T. and Trautwein, F. (2023), "Planspielbasierte Lehrveranstaltungen: Ein Vergleich zwischen Präsenz- und Onlinelehre", *Zeitschrift Für Hochschulentwicklung*, Vol. 18 No. Sonderheft Planspiele, pp. 345–363, doi: 10.21240/zfhe/SH-PS/19.

Alsaiari, O., Baghaei, N., Lodge, J.M., Noroozi, O., Gašević, D., Boden, M. and Khosravi, H. (2026), "Directive, metacognitive, or a blend of both? A comparison of AI-generated feedback types on student engagement, confidence, and outcomes", *Computers and Education: Artificial Intelligence*, Vol. 10, p. 100553, doi: 10.1016/j.caeai.2026.100553.

Anthropic. (2026), "Claude 3.5 Sonnet. version: claude-3-5-sonnet-20241022".

Banihashem, S.K., Noroozi, O., Khosravi, H., Schunn, C.D. and Drachsler, H. (2026), "Pedagogical framework for hybrid intelligent feedback", *Innovations in Education and Teaching International*, Vol. 63 No. 2, pp. 554–570, doi: 10.1080/14703297.2025.2499174.

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A. and Van Der Velde, F. (2023), "A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation", *Personal and Ubiquitous Computing*, Vol. 27 No. 2, pp. 317–330, doi: 10.1007/s00779-022-01690-0.

Brommeyer, M. and Liang, Z. (2022), "A Systematic Approach in Developing Management Workforce Readiness for Digital Health Transformation in Healthcare", *International Journal of Environmental Research and Public Health*, Vol. 19 No. 21, p. 13843, doi: 10.3390/ijerph192113843.

Chen, L., Chen, P. and Lin, Z. (2020), "Artificial Intelligence in Education: A Review", *IEEE Access*, Vol. 8, pp. 75264–75278, doi: 10.1109/ACCESS.2020.2988510.

Crookall, D. (2010), "Serious Games, Debriefing, and Simulation/Gaming as a Discipline", *Simulation & Gaming*, Vol. 41 No. 6, pp. 898–920, doi: 10.1177/1046878110390784.

DeepL. (2026), "DeepL", 15 July, available at: https://www.deepl.com/de/write.

Eisinga, R., Grotenhuis, M.T. and Pelzer, B. (2013), "The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?", *International Journal of Public Health*, Vol. 58 No. 4, pp. 637–642, doi: 10.1007/s00038-012-0416-3.

Fantozzi, I.C. and Schiraldi, M.M. (2026), "Bridging education and industry: university challenges and serious games for professional skill development", *The TQM Journal*, Vol. 38 No. 3, pp. 595–610, doi: 10.1108/TQM-12-2024-0511.

George, D. and Mallery, P. (2003), *SPSS® for Windows® Step by Step: A Simple Guide and Reference*, 4th ed., Vols. 1-11.0 Update, Allyn & Bacon, Boston.

Häberle, O., Baume, M. and Krcmar, H. (2006), "Realitätsnahes Entscheidungstraining für zukünftige IT-Manager: Ein computerunterstütztes Planspiel: Aufbau und Lessons Learned".

Hellstrand Tang, U., Smith, F., Karilampi, U.L. and Gremyr, A. (2024), "Exploring the Role of Complexity in Health Care Technology Bottom-Up Innovations: Multiple-Case Study Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability Complexity Assessment Tool", *JMIR Human Factors*, Vol. 11, p. e50889, doi: 10.2196/50889.

Hochschule Osnabrück. (2026), "SMiLeH - Strategic Management in Learning Health Systems", 4 September, available at: https://www.hs-osnabrueck.de/smileh/ergebnisse-und-publikationen/.

Jaccard, D., Bonnier, K.E. and Hellström, M. (2022), "How might serious games trigger a transformation in project management education ? Lessons learned from 10 Years of experimentations", *Project Leadership and Society*, Vol. 3, p. 100047, doi: 10.1016/j.plas.2022.100047.

Jaccard, D., Suppan, L., Sanchez, E., Huguenin, A. and Laurent, M. (2021), "The co.LAB Generic Framework for Collaborative Design of Serious Games: Development Study", *JMIR Serious Games*, Vol. 9 No. 3, p. e28674, doi: 10.2196/28674.

Kuckartz, U. and Rädiker, S. (2024), *Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz*, 6. Auflage., Juventa Verlag, Weinheim.

Opas, M. and Ćwil, M. (2026), "Simulation-based Learning and Project Management Certification Outcomes: Evidence from IPMA Level D Training", *Electronic Journal of E-Learning*, Vol. 24 No. 3, pp. 79–91, doi: 10.34190/ejel.24.3.4740.

OpenAI. (2026), "ChatGPT", GPT 5.5, available at: https://chatgpt.com/.

Reichheld, F.F. (2003), "The one number you need to grow", *Harvard Business Review*, Vol. 81 No. 12, pp. 46–54, 124.

Salminen, J., Santos, J.M., Kwak, H., An, J., Jung, S. and Jansen, B.J. (2020), "Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas", *International Journal of Human-Computer Studies*, Vol. 141, p. 102437, doi: 10.1016/j.ijhcs.2020.102437.

Schwägele, S. (2015), *Planspiel – Lernen – Lerntransfer Eine Subjektorientierte Analyse von Einflussfaktoren*, Monografie, Otto-Friedrich-Universität Bamberg, Bamberg.

Sharma, A., Minh Duc, N.T., Luu Lam Thang, T., Nam, N.H., Ng, S.J., Abbas, K.S., Huy, N.T., *et al.* (2021), "A Consensus-Based Checklist for Reporting of Survey Studies (CROSS)", *Journal of General Internal Medicine*, Vol. 36 No. 10, pp. 3179–3187, doi: 10.1007/s11606-021-06737-1.

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G. and Vasileva, T. (2025), "AI -driven adaptive learning for sustainable educational transformation", *Sustainable Development*, Vol. 33 No. 2, pp. 1921–1947, doi: 10.1002/sd.3221.

Digital Serious Games in Healthcare Management: Comparative Analysis and Design Recommendations

Tobias Höllmer[1], Annika Hering[1], Gülsen Otman[1], Uwe Buddrus[1] and Jan-David Liebe[1,2]
[1]Center for Medical, Social and Health Informatics (MESH), Osnabrück University of Applied Sciences, Osnabrück, Germany
[2]UMIT - Private University of Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

tobias.hoellmer@hs-osnabrueck.de
a.hering@hs-osnabrueck.de
g.otman@hs-osnabrueck.de
u.buddrus@hs-osnabrueck.de
j.liebe@hs-osnabrueck.de

Abstract: Introduction Digital serious games (DSGs) enable healthcare management students to safely navigate the systemic complexities of the sector. However, domain-specific comparative frameworks remain scarce. Objectives: This study addresses three research questions (RQs): (RQ1) What comparative characteristics can be identified for the analysis of serious games in healthcare management? (RQ2) What similarities and differences emerge between the serious games based on these characteristics? and (RQ3) Which recommendations can be derived for the design of DSGs for teaching? Methods: To address RQ1, a literature review across four databases identified five sources describing game characteristics, design principles, and pedagogical principles. Following a relevance assessment by five researchers, the identified characteristics were synthesized into a comparison matrix comprising five clusters: formal game structure and interaction; learning objectives (LOs) and measurability of results; game rules, game dynamics, and decision-making structures; resources and game environments; and aesthetic aspects and didactics. To address RQ2, the matrix was applied to five healthcare management DSGs. The analysis drew on publicly available materials, demo versions, structured interviews with providers or experts, and supplementary provider materials. Based on the comparative findings, competency-oriented design recommendations (DRs) were derived to address RQ3. Results: The comparison revealed several recurring characteristics across the analyzed DSGs. Considerable variation was identified in player interaction, game duration, strategic flexibility, feedback mechanisms, and the representation of healthcare-specific complexity. Based on these findings, ten DRs were developed, emphasizing multi-role structures, LO-aligned game mechanics and feedback, flexible strategic decision-making, multidimensional performance indicators, explicit resource constraints, increasing complexity, and risk-free experimentation. Discussion: The findings indicate that competency-oriented DSG design requires alignment between LOs, game mechanics, decision structures, and feedback mechanisms. However, the study was limited to five specialized DSGs and examined design characteristics rather than learner experiences or effectiveness. Future studies should assess the effects of the proposed DRs on learning outcomes, competency development, strategic decision-making, and learner engagement. Conclusion: The study provides an initial, literature-informed comparison matrix and ten provisional DRs for competency-oriented DSGs in healthcare management. The DRs offer a basis for developing future learning environments that reflect the specific complexity of healthcare management.

Keywords: Simulation Games, Serious Games, Healthcare Management, E-Learning, Design

1. Introduction

Digital serious games (DSGs) are increasingly recognized as a crucial e-learning approach, enabling students to interact with complex, practice-oriented scenarios in a risk-free, interactive environment. DSGs can be viewed as innovative educational tools that provide students with interactive, motivating, and engaging learning experiences, thereby fostering cognitive and social skills (Anastasiadis *et al.*, 2018). Using a game-based approach in teaching incorporates playful elements in a friendly competitive environment to boost learners' motivation and interest, support them in achieving learning objectives (LOs), and improve skills (Cavus *et al.*, 2023). This prepares learners for professional practice by developing the competencies they will need in their future roles. The potential applications of serious games are diverse, ranging from educational games for children to simulation environments for industry or for use in education to raise awareness about managing illness (Dörner *et al.*, 2016).

This approach is particularly relevant for healthcare management students, given that organizations in this sector have highly complex systems involving diverse stakeholder groups and competing objectives. Today, healthcare managers must navigate complex, multidimensional systems, coordinate numerous stakeholders, implement new technologies and regulations, stay within budgetary constraints, and simultaneously ensure the highest levels of patient safety and quality of care (Spanos *et al.*, 2024). The complexity of the system is also evident in

the changing demands placed on healthcare leaders. Accordingly, leadership in this system requires not only an understanding of roles and competencies, but also a focus on goals, the promotion of a corporate culture, and the targeted promotion of innovation (Spanos *et al.*, 2024).

Various generic frameworks exist for the comprehensive analysis of serious games. However, generic models for analyzing serious games, such as the Learning Mechanics – Gaming Mechanics Model (LM-GM) model (Arnab *et al.*, 2015) and the Serious Game Design Assessment Framework (SGDAF) (Mitgutsch and Alvarado, 2012), are insufficient for this purpose, as they fail to consider the distinctive features of the healthcare sector and the contemporary understanding of the role of healthcare managers.

To address this critical gap, this study aims to conduct a sector-specific analysis of existing DSGs, evaluate these DSGs, and develop design recommendations (DRs) for future development. This is essential to ensure that future healthcare managers are equipped with evidence-based, competency-oriented digital learning tools that reflect real-world institutional complexities. Based on these topics, the following research questions (RQs) have arisen:

RQ1: What comparative characteristics can be identified for the analysis of serious games in healthcare management?

RQ2: What similarities and differences emerge between serious games based on these characteristics?

RQ3: Which recommendations can be derived for the design of DSGs for teaching?

2. Theoretical Background

2.1 Digital Serious Games in E-Learning

Dörner et al. (2016) define a serious game as “a digital game created with the intention to entertain and to achieve at least one additional goal (e.g., learning or health).” In the context of serious games in healthcare management, this additional goal is learning. Consequently, the game development team needs subject matter experts to create serious games. Developing serious games remains a challenge, even though their benefits have long been recognized (Dörner *et al.*, 2016).

2.2 Healthcare Management as Learning Context

This raises the question of how the learning context of healthcare management can be situated in this context. Healthcare management is essentially a distinct branch of training within healthcare professions. It focuses on developing competencies at systemic level, such as quality management, policy, and crisis communication. The field of healthcare management differs fundamentally from general medical training, which primarily focuses on clinical competencies. The study of hospital management examines organizational performance and how decisions influence systems and organizations (Zhang *et al.*, 2026).

2.3 Existing Frameworks for Serious Games Analysis

In the field of serious games analysis, there are frameworks that address this topic more broadly. In this context, Mitgutsch & Alvarado (2012) developed a framework that analyzes serious games based not on their entertainment value or content, but rather on the formal connection between the game’s elements and the LOs. This framework is named “SGDA-Framework”, and it defines six dimensions that must be integrated into a serious game: Purpose; Content and Information; Game Mechanics; Fiction and Narrative; Aesthetics and Graphics; Framing; Coherence and cohesiveness of the game (Mitgutsch and Alvarado, 2012). Another framework for the analysis of serious games is the “Learning Mechanics-Game Mechanics Framework” by Arnab *et al.* (2015). It is designed to support the analysis and design of serious games by mapping the relationship between pedagogical elements and game mechanics. The model consists of two main systems: Learning Mechanics (LMs) and Game Mechanics (GMs). These are connected through a descriptive framework that enables users to generate maps and tables illustrating static and dynamic relationships between educational and gaming elements (Arnab *et al.*, 2015).

Although these frameworks are a theoretical basis for analyzing serious games, there is a lack of a systematic and comparative analytical tool in the field of serious games in healthcare management. Furthermore, it remains unclear how existing serious games represent specific systemic complexity of the healthcare system. This study addresses this gap by developing a literature-based comparison matrix that was directly applied to five

established DSGs in healthcare management (Elwacon Business Partner, 2026; HFX Training, 2026; MEGA Learning, 2026; SimInstitute, 2026; TopSim, 2026).

3. Methods

3.1 Framework Development

Regarding the first research question (RQ1), a generic literature review was conducted across four databases to identify relevant dimensions for serious games in healthcare management. The databases used for this purpose were SpringerLink, the internal database of Osnabrück University of Applied Sciences ("scinos"), PubMed and Google Scholar. Search strings were developed using Boolean operators (AND, OR). The database search was performed combining keywords such as "game", "game design", "serious games", "digital serious games", and "learning by design".

Five sources were found that define and describe the dimensions of a game (Fullerton, 2024; Gee, 2004; Hunicke *et al.*, 2004; Schell, 2008; Zagal and Bruckman, 2009). In the next step, frameworks were extracted from the literature. Overall, categories were established that encompass the general framework of a game, the design principles, and the pedagogical principles. These initial frameworks were discussed with five researchers to ensure validity and applicability to DSGs in healthcare management.

To cluster the frameworks and dimensions in the next step, we looked for commonalities in their content so that groups with similar thematic focuses could be formed. Categories with identical content were merged. Additionally, the frameworks were evaluated based on expert consensus. The research team evaluated all categories using a 3-point rating scale (1 = lowest relevance, 3 = highest relevance) to ensure a clear distinction between low, moderate, and high relevance without forced granular nuances. Frameworks and dimensions were assigned a score of 1 if not relevant to the development of management competencies, or if, although they were relevant, they appeared to be better understood when considered alongside other categories in a different context. A rating of 3 indicated a high level of relevance for teaching management competencies or the fundamental elements of a DSG. All categories were evaluated using this scale. Categories with a rating of 1 were then excluded. This led to the exclusion of the following categories: roles of players (Fullerton, 2024); agent goals (Zagal and Bruckman, 2009); object attributes (Schell, 2008); entity manipulation (Hunicke *et al.*, 2004; Zagal and Bruckman, 2009); checkpoint, cropping, expiration, Pseudo-physical Rules, save point, spawn point (Zagal and Bruckman, 2009); operational rules, foundational rules, behavioural rules, written rules, laws, official rules, advisory rules, house rules (Schell, 2008); information on demand or just in time (Gee, 2004); meaning as action image (Gee, 2004); Sensation, fantasy, narrative, Challenge, fellowship, discovery, expression, submission (Hunicke *et al.*, 2004); Cycles of expertise challenge (Gee, 2004) and Pleasantly frustrating (Gee, 2004).

3.2 Comparative Analysis

To answer the second research question (RQ2), 18 DSGs in the field of healthcare management were identified through an open-ended search on the "Google" search engine in February 2026. Search terms such as "healthcare management game", "serious games", "simulation games", "hospital business games" or "hospital management game" were used. The identified DSGs were then subsequently categorized according to teaching applications, competency-oriented designs, and exclusive use within the healthcare sector. The DSGs were ultimately analyzed based on the clusters and categories identified in the systematic analysis.

Data on the DSGs was gathered from four sources:

1. Information sourced from open-access web resources and collateral public materials;
2. demo version of the DSG;
3. interview with the provider or expert of the serious game and
4. exclusive materials from the providers.

Initially, a comprehensive review of publicly accessible informa

tion was conducted, which included open-access web resources and marketing or educational materials (1). To gain a deeper understanding of the DSGs, the providers were asked to provide access to a demo version and to participate in an interview. Not all DSGs were able to provide a demo version. Where possible, the demo versions were used to analyze design elements and content of a DSG (2). The interviews with the providers were

conducted using a structured interview guide (Helfferich, 2009) (3). In addition, the providers provided access to exclusive supplementary materials that were not freely available (4).

To ensure complete traceability and transparency throughout the analysis, a documentation system was established. Each piece of information collected was systematically assigned a reference number from (1) to (4) that corresponded directly to the specific data source from which it originated. This allowed for a clear and comprehensive mapping of information.

4. Results and Analysis

4.1 The Five Analysis Clusters

The dimensions of a game were identified based on various sources, resulting in five distinct clusters.

Cluster (a) “Formal Game Structure and Interaction” describes the general framework and control mechanisms of a game. The thematic focus of this cluster is on the number of players, how players or teams interact with or against each other, and how players’ actions are reflected in the game.

Cluster (b) “Learning Objectives and Measurability of Results” focuses on classification and types of goals as well as measurability and difficulty. To improve comparability, all LOs of the DSGs were first compiled, and then subcategories were formed based on content overlaps. These subcategories were then described using a keyword. This resulted in a total of eight general LOs. Furthermore, ways to win the game, and how to measure success in the game were analyzed.

Cluster (c) “Game Rules, Game Dynamics and Decision-Making-Structure” describes the defined set of rules, game structural flow, and decision-making logic within the game. The core dimensions of this cluster include the game world and specific aspects of gameplay, such as player briefings, the number of decision rounds, and the development of game strategy.

Cluster (d) “Resources and Game Environments” deals with spatial representation of the game, object structure, definition and significance of resources. Within these categories, the resource, their availability, and the space of the game were analyzed. The fifth Cluster (e) “Aesthetic Aspects and Didactics” places a central emphasis on game design and its educational applications. In particular, the analysis focuses on basic principles of game-based learning and the presentation of the game.

4.2 Comparative Analysis of Serious Games

The five clusters enabled a systematic analysis of the DSGs across several dimensions. To illustrate basic elements of a DSG, elements that are common to all DSGs were highlighted. This helped to illustrate the basic elements of a DSG in healthcare management.

4.2.1 Cluster (a): “Formal Game Structure and Interaction”

As Table 1 shows, an analysis of formal game structure and interaction in the simulation revealed that all DSGs were digital applications that employed an indirect manipulation method within the game. The differences lie primarily in number of players and in player interaction patterns. All forms of player interaction patterns were present here. The DSG analysis allowed for both single player and team to play against a computer, as well as team play against other teams. The DSG “Healthcare Hero” could even be configured to accommodate all player interaction patterns. In terms of the number of players, team games tended to involve groups of up to five people, divided into three to five teams.

Table 1: Formal Game Structure and Interaction - Matrix (Cluster a); abridged version

Design Elements	Comparison Dimensions	Comparative Features	Healthcare Hero (HFX Training)	Healthcare Strategy Challenge (MEGA Learning)	Hospital Business Management (Elwacon)	Hospital Operations (Sim-Institute)	Hospital Management (TopSim)
Players	Number of Players (Single- vs. Multiplayer)	Single-player, multiplayer, or teams?	Standard team size between 3-5, Cohort size unlimited but	Groups of 4–5, 3–5 teams	25 players max., at least 9 players, at	Single- and multiplayer	Multiplayer

Design Elements	Comparison Dimensions	Comparative Features	Healthcare Hero (HFX Training)	Healthcare Strategy Challenge (MEGA Learning)	Hospital Business Management (Elwacon)	Hospital Operations (Sim-Institute)	Hospital Management (TopSim)
			usually between 5-10 teams		least 3 groups		
Players	Player Interaction Patterns	Single player vs. Game, Multiple individual players vs. Game, Player vs. Player, Unilateral competition, multilateral competition, Cooperative play, Team competition	Single player vs. Game, Multiple individual players vs. Game, Player vs. Player, Unilateral competition, multilateral competition, Cooperative play, Team competition	Cooperative Play	Multilateral team competition	Single player vs. game; Cooperative Play	Multilateral team competition
Input Method	Locus of Manipulation	Direct and indirect manipulation	Indirect manipulation	Indirect manipulation	Indirect manipulation	Indirect manipulation	Indirect manipulation
	Manipulation Method	Analog vs. digital, gestic, buttons	Digital	Digital	Digital	Digital	Digital

4.2.2 Cluster (b): "Learning Objectives and Measurability of Results"

Analyzing LOs and the measurability of results provided a fundamental evaluation of instructional goals and an understanding of success within the games. Fullerton (2024) identified ten different game objectives in principle. The general objective of all DSGs in healthcare management was "construction" (Fullerton, 2024), meaning that the goal was to manage objects. The analysis of the LOs began by compiling all stated LOs gathered from various sources. All LOs identified through the research were grouped based on commonalities, resulting in eight LOs. However, when interpreting Table 2, it must be critically noted that the evidence for the respective LOs predominantly based of public materials (P) and the self-reported statements of providers during interviews (I). An objective verification of actual game mechanics using demo versions (D) was only possible in exceptional cases, such as partially for "Hospital Operations".

Table 2: Learning Objectives (LOs) and Measurability of Results - Matrix (Cluster b); abridged version

Learning Objective (LO)	LO Description	Healthcare Hero (HFX Training)	Healthcare Strategy Challenge (MEGA Learning)	Hospital Business Management (Elwacon)	Hospital Operations (Sim-Institute)	Hospital Management (TopSim)
LO1: Strategy Development	Strategic management in the context of strategy development and goal setting	I	P	P	P	I
LO 2: System Thinking	Comprehensive understanding of hospital operations and system thinking	I	P	P	P	I
LO 3: Data-driven Decision Making	Analysis of management-relevant data and data-driven decision-making	I	P	P	D	I
LO 4: Teamwork	Teamwork and interdisciplinary collaboration	I	P	P	D	I

Learning Objective (LO)	LO Description	Healthcare Hero (HFX Training)	Healthcare Strategy Challenge (MEGA Learning)	Hospital Business Management (Elwacon)	Hospital Operations (Sim-Institute)	Hospital Management (TopSim)
LO 5: Working efficiently	Optimizing Resource Management and Cost-Effectiveness	I	P	P	D	I
LO 6: Leadership skills	Gain an understanding of your own leadership role and the company's overall leadership role	I	P	P	P	I
LO 7: Multidimensional Impact Analysis	Understanding the impact of management decisions on patient outcomes, staff, and financial results	I	I	I	P	I
LO 8: Transfer effectiveness	Understanding of the effective use of simulation exercises to teach LOs in the strategic management of hospitals	I	I	I	P	I

Note: Letters in the table indicate the data source used to verify each LO; P – Information sourced from open-access web resources and collateral public materials, D – Demo version of the SG, I – Interview with the provider or expert of a SG.

Across available sources, evidence corresponding to each of the eight synthesized LO categories was identified for all five DSGs. To understand how these LOs were practically facilitated, the structural design of conflict was analyzed, specifically focusing on opponent structure. Here, differences became apparent. In DSGs, which were multilateral team competitions, the opponent in the game was another team. In a game with the player interaction pattern of a cooperative game, the opponent was a non-player opponent. One DSG even allowed for both opponent structures. In contrast, all DSGs were win/lose-games, meaning that at the end of the game, a winner was determined based on specific criteria. These criteria also varied widely and represented another key aspect of a DSG in healthcare management. There were three ways to win a DSG: the players with the highest overall score, winning criteria that compared each team against another, and selection criteria supplemented by an assessment of soft skills. Another factor that had to be considered when evaluating the winning criteria was the number of key performance indicators (KPIs) analyzed. It appears that between six and 20 KPIs were analyzed as part of the DSGs winning criteria. Performance was evaluated across all games based on financial criteria and stakeholder criteria, such as patient satisfaction. Results were presented via dashboards or tables. One simulation relied on a KPI matrix, while another used an AI-powered feedback system in a chat format.

4.2.3 Cluster (c) "Game Rules, Game Dynamics and Decision-Making Structure"

Next, the defined set of rules, the games structural flow, and decision-making logic within the games were analyzed. The borders of the simulation environment were the same for all DSGs and were limited to the analysis of a hospital at organizational level. Differences were evident in the size of the game world. The number of beds ranged from 207 to 300, and there were typically three departments within the game world. Special attention was paid to DSG structures. The overall playtime of the game showed that, despite many similarities, we are talking about fundamentally different game models. This was evident because typical duration ranged from 75 minutes to several days, though it is generally flexible. The briefing and debriefing structure illustrated how players were prepared for the game and guided through it as it progressed. Typically, players were introduced to the game using materials such as handouts, videos, or manuals; this was followed by a debriefing and discussion of how the game went. In addition, practice rounds could be played before the game began, to give players a feel for the game mechanics. For all DSGs, decisions were made at the end of a period, which typically spanned one year. Some decisions were made independently, while others had to be made within the framework of policies. The focus was on investments, salaries, and performance and capacity planning. Differences also emerged in the development of the game strategy. Games might dictate a strategy or allow players to develop their own strategy, which was then adapted as the game progressed. The tension usually stems from time pressure or pressure from opponents. The analysis also highlighted the importance of a systems-thinking approach in game design.

4.2.4 Cluster (d) "Resources and Game Environments"

The analysis of resources and game environments focused on the spatial representation of the game, the object structure, definition and significance of resources. DSGs in healthcare management were the same in their two-dimensional representation, their discrete spatial representation, and the availability of limited resources. The resources available to players in this context were similar and included, for example, financial budgets, staffing capacity, stakeholders or salary caps. In addition, the deployment method used in newer games is web-based, although one game still ran on a desktop. When examining the game objects within the context of resources and the game environment, it became clear that the games primarily relied on decision pages, dashboards, and reports. Other features here included news, reports and a glossary.

4.2.5 Cluster (e) "Aesthetic Aspects and Didactics"

An analysis of the game design revealed that none of the DSGs use audio content. The Learning by Design approaches are implemented in various ways. To begin with, players took an active role within learner agency. Roles that players could assume within the DSGs included strategic management HR, controlling and project management. Another principle of Learning by Design concerns the adaptability of difficulty levels. Some DSGs offered this, some did not. This was also the case with "well- ordered problems", namely whether players had to solve easier problems first and then move on to harder ones. The sandbox principle was also analyzed. This principle determines whether players can gain experience in a risk-free environment. The experiences players can gain through the DSGs include, for example, capacity planning, financial management, human resources, and marketing. According to some sources, the fishtank-principle, which was the simplified representation of a complex system, was applied here across all possible areas of a hospital.

4.3 Design Recommendations

The comparison matrix enabled a comparative analysis of five DSGs in healthcare management, revealing not only similarities and differences but also the fundamental principles of game design in this specific field. Based on these findings, in the next step DRs were developed to guide the future development of DSGs (RQ3).

4.3.1 Cluster (a): "Formal Game Structure and Interaction"

DR1: Use a multiplayer and multi-role structure.

Almost all games feature a multiplayer mode. However, complex stakeholder relationships can be presented not only by a team but also by different roles. The inclusion of different role perspectives should be considered to illustrate the dynamics within the healthcare system. Furthermore, DSGs could assign different roles within teams, which would be presented to players through different interfaces. For example, HR, financial control, and strategic management are distributed among the team members.

DR2: Use a digital platform and incorporate various supplementary materials.

A digital platform should allow players to access the game regardless of the device they use. Depending on the learning context, both digital and analog supplementary materials can be incorporated. When it comes to using digital tools, there are many possibilities. In addition to AI feedback tools, virtual reality content, for example, could also contribute to learning success. Given the limitless design possibilities, it is highly recommended to strategically select and integrate the specific supplementary materials that best align with the target LOs.

4.3.2 Cluster (b): "Learning Objectives and Measurability of Results"

DR3: Teach the competencies that hospital managers need in their daily work.

Teaching management competencies is the primary goal of DSGs in the healthcare sector. LOs should be oriented toward these competencies. The question of which competencies are essential for hospital managers remains contested. A systematic development of LOs from these competencies informed by literature and practical expertise is needed to align them with the DSG design and ensure their ongoing relevance in a rapidly changing healthcare environment.

DR4: Establish multidimensional performance metrics with LO-aligned feedback mechanisms

Evaluate success by deploying a balanced, multidimensional KPI matrix that integrates financial metrics, patient and staff needs, and quantified soft skills. It is crucial that the chosen feedback method be tailored to the specific

characteristics of each LO. Complex leadership skills cannot be taught only using quantitative, static data. Instead, they require conversational, analytical feedback.

4.3.3 Cluster (c) "Game Rules, Game Dynamics and Decision-Making Structure"

DR5: Implement a game flow that supports the achievement of LOs.

The flow of the game should be structured around the LOs, both in terms of time and subject matter. It may be useful to follow fiscal years or quarters but depending on context the game can also progress through project phases. Briefing and debriefing phases should be incorporated to evaluate progress and introduce new content. The ability to adjust the game's length allows for better alignment with individual LOs.

DR6: Provide flexible strategic decision-making.

Decision-making is one of the game's central challenges. There are various ways to structure this, and the right approach depends on the game's objective. Decision options can be either predetermined or freely chosen. Having a variety of decision options allows for greater flexibility.

4.3.4 Cluster (d) "Resources and Game Environments"

DR7: Represent resource constraints explicitly and adherence to actual hospital capacity metrics.

The hospital structure in the game should be based on real-world hospital metrics. Factors such as the number of beds and specialized departments provide a realistic framework for hospital planning. While budgets, staffing, and bed capacities act as strict operational limits during gameplay, they must function as dynamic, adjustable management levers that players can strategically expand and optimize over time.

DR8: The design must focus on illustrating the impact of decisions.

DSGs use dashboards or tables to visualize game results. The guiding principle here should be that DSGs are designed in such a way that players gain an understanding of the impact of their decisions. While dashboards and tables can certainly be used to illustrate these impacts, for example special reports, newspaper articles, or employee surveys can also serve this purpose.

4.3.5 Cluster (e) "Aesthetic Aspects and Didactics"

DR9*:* The complex dynamics of the healthcare system should be reflected in the design.

These complex dynamics pose a major challenge for healthcare managers. The representation of these complex dynamics should always be simulated in the game. Complexity can be illustrated, for example, by different stakeholder perspectives, roles with varying interests, or unforeseeable events.

DR10: Build in increasing complexity and enable risk-free experimentation.

Incorporating an increasing level of complexity helps keep players engaged (Gee, 2004). This can be achieved by starting with simple information or by limiting the game world at start. Test runs help players understand the game's dynamics and the consequences of their decisions.

5. Discussion

This study aimed to identify relevant characteristics for analyzing DSGs in healthcare management, to compare DSGs based on these characteristics, and to derive initial DRs for educational game development. The findings show that DSGs in healthcare management can be systematically compared using a literature-informed analytical structure. This matrix should be understood as an initial analytical heuristic that brings together relevant dimensions from existing serious game literature and applies them to a specific educational field. In doing so, the study provides a structured basis for making similarities and differences between existing DSGs visible.

The literature-based individual characteristics are generally known for describing the features of a DSG. The comparative analysis and adaptation, presented in form of a matrix, provided deep insight into sector-specific characteristics that consider the unique features of healthcare management. The final DRs represent a concrete outcome that goes beyond the generic principles of established frameworks. Across the five analyzed DSGs, recurring features were identified, including digital implementation, decision-based gameplay, limited resources, performance-oriented feedback, and competency-oriented LOs. At the same time, the comparison revealed considerable variation in interaction patterns, game duration, decision-making flexibility, feedback mechanisms,

and the extent to which healthcare-specific complexity is represented. The analysis suggests that existing DSGs in healthcare management share a common pedagogical core. All games address the same LOs and include elements related to teamwork, leadership, and resource optimization. From an LM-GM perspective, the mere presence of LOs is not sufficient. Educational value depends on whether these objectives are translated into meaningful game mechanics, decision structures, feedback loops, and opportunities for reflection (Arnab *et al.*, 2015). These findings indicate that DSGs are especially suitable for teaching competencies that require learners to connect analytical reasoning with strategic judgement under conditions of uncertainty. Not all games represent stakeholder perspectives, soft skills, or multidimensional consequences of decisions to the same extent. Future DSGs should therefore move beyond a narrow focus on financial or operational indicators and include feedback mechanisms that make process-related, technical, and organizational consequences visible. This finding resonates with the concept of serious games as 'reality-changing methods' (Sun and Zhang, 2025). Their participatory design study demonstrates that effective game design goes beyond simulation by translating complex system dynamics into experiential learning environments, thereby changing how participants perceive and act within real-world systems. A limitation is the small and highly specialized sample of five DSGs in a healthcare management context, which limits the transferability of the findings to other DSGs and educational contexts. In addition, the games were identified through an open-ended web search and access to available providers, meaning that unpublished, institution-specific, non-commercial, or non-English-language games may not have been included.

The study focused on game characteristics and design patterns rather than on learner experiences or long-term competency development. Moreover, the identification and assignment of LOs in the comparison matrix were constrained by the availability, scope, and level of detail of provider-supplied information and other publicly available sources. The study therefore did not examine whether learners achieved the identified LOs. The binary presentation in Table 2 merely captures whether LOs are addressed, but it does not allow for any conclusions regarding the depth or quality of its didactic implementation. Future research should investigate how competency requirements for healthcare managers can be systematically translated into observable and measurable LOs and subsequently aligned with DSG design decisions. Because the current study solely evaluates the formal presence of specific design and learning characteristics, the actual pedagogical effectiveness remains unexamined. Consequently, empirical effectiveness studies are needed to examine whether the proposed DRs are associated with improved learning outcomes, particularly competency development, learner engagement, and strategic decision-making.

6. Conclusion

This study provides a structured basis for analyzing and informing the design of DSGs in healthcare management education. By comparing five DSGs, it identified recurring design patterns and relevant differences in how LOs, decision-making structures, feedback mechanisms, resource constraints, and healthcare-specific complexity are represented in educational settings. The resulting DRs offer an initial orientation for educators and game developers seeking to derive and refine design decisions aligned with competency-oriented LOs. Therefore, the findings point to the importance of multi-role structures, flexible strategic decision-making, multidimensional feedback, explicit resource limitations, and the representation of process-related, technical, and organizational consequences. In this way, the study contributes to translating insights from existing DSGs into more concrete design considerations for future competency-oriented DSGs in healthcare management.

Data Availability Statement

The full dataset can be viewed on the website (SMiLeH - Comparison Matrix, 2026) or is available from the author upon request.

Funding

This study is part of the research project “SMiLeH - Strategic Management in Learning Health Systems” (Project No. 3004-0241), funded by the Stiftung Innovation in der Hochschullehre (StIL).

Ethics Declaration

All listed providers gave written informed consent and agreed to the use and publication of the collected data and textual interpretations for research purposes.

AI Declaration

DeepL was used to improve the language and grammar. Gemini was used to structure texts and adapt them to citation guidelines. Elicit and Perplexity were used to assist with literature reviews.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Author Contributions

TH: Conceptualization, Formal analysis, Writing – Original draft. AH: Methodology, Writing – Review & Editing. GO: Writing – Review & Editing. UB: Writing – Review & Editing. JDL: Conceptualization, Supervision.

Acknowledgements

We would like to thank the providers of the DSGs for their transparent and open communication and for sharing relevant information for this work: HFX Training, SA MEGA Learning, Elwacon – Businesspartner GmbH & Co. KG, and Sim-Institute.

References

Anastasiadis, T., Lampropoulos, G. and Siakas, K. (2018), "Digital Game-based Learning and Serious Games in Education", *International Journal of Advances in Scientific Research and Engineering*, Vol. 4 No. 12, pp. 139–144, doi: 10.31695/IJASRE.2018.33016.

Arnab, S., Lim, T., Carvalho, M.B., Bellotti, F., De Freitas, S., Louchart, S., Suttie, N., *et al.* (2015), "Mapping learning and game mechanics for serious games analysis", *British Journal of Educational Technology*, Vol. 46 No. 2, pp. 391–411, doi: 10.1111/bjet.12113.

Cavus, N., Ibrahim, I., Ogbonna Okonkwo, M., Bode Ayansina, N. and Modupeola, T. (2023), "The Effects of Gamification in Education: A Systematic Literature Review", *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, Vol. 14 No. 2, pp. 211–241, doi: 10.18662/brain/14.2/452.

Chatzea, V.E., Panagiotakis, C., Vidakis, N. and Fanourgiakis, J. (2026), "Be Ahead of the Game! A Systematic Review on Serious Games Aiming to Enhance Hospital Management Knowledge and Health Care Administrative Skills Acquisition", *Games for Health Journal*, Vol. 15 No. 1, pp. 28–41, doi: 10.1177/2161783X251361157.

Dörner, R., Göbel, S., Effelsberg, W. and Wiemeyer, J. (Eds.). (2016), *Serious Games*, Springer International Publishing, Cham, doi: 10.1007/978-3-319-40612-1.

Elwacon Business Partner. (2026), "Elwacon Business Partner", *Hospital Business Challenge*, available at: https://elwacon.de/.

Fullerton, T. (2024), *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*, Fifth edition., CRC Press, Boca Raton, FL.

Gee, J.P. (2004), "Learning by design: Games as learning machines", Wisconsin-Madison, Vol. Interactive Educational Multimedia No. 8, pp. 15–23.

Helfferich, C. (2009), *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews*, 3., überarbeitete Auflage., VS Verlag für Sozialwissenschaften, Wiesbaden, doi: 10.1007/978-3-531-91858-7.

HFX Training. (2026), "HFX Training", *Healthcare Hero*, available at: https://hfxtraining.com/.

Hunicke, R., LeBlanc, M. and Zubek, R. (2004), "MDA: A Formal Approach to Game Design and Game Research".

MEGA Learning. (2026), "MEGA Learning", *Healthcare Strategy Challenge*, available at: https://www.megalearning.com/.

Mitgutsch, K. and Alvarado, N. (2012), "Purposeful by design?: a serious game design assessment framework", *Proceedings of the International Conference on the Foundations of Digital Games*, presented at the FDG'12: International Conference on the Foundations of Digital Games, ACM, Raleigh North Carolina, pp. 121–128, doi: 10.1145/2282338.2282364.

Schell, J. (2008), *The Art of Game Design: A Book of Lenses*, Elsevier/Morgan Kaufmann, Amsterdam ; Boston.

SimInstitute. (2026), "SimInstitute", *Hospital Operations*, available at: https://www.siminstitute.com.

SMiLeH - Comparison Matrix. (2026), "SMiLeH - Strategic Management in Learning Health Systems - Comparison Matrix", 15 July, available at: https://website-smileh-comparison-matrix.onrender.com.

Spanos, S., Leask, E., Patel, R., Datyner, M., Loh, E. and Braithwaite, J. (2024), "Healthcare leaders navigating complexity: a scoping review of key trends in future roles and competencies", *BMC Medical Education*, Vol. 24 No. 1, p. 720, doi: 10.1186/s12909-024-05689-4.

Sun, Y. and Zhang, C. (2025), "Serious Game Design for Teaching University Students to Address Complexity Issues in the Healthcare Logistics System: Lessons from an Emergency Department Case Study", *Systems*, Vol. 13 No. 3, p. 197, doi: 10.3390/systems13030197.

TopSim. (2026), “TopSim”, *Hospital Management*, available at: https://topsim.com/wp-content/uploads/2021/03/KB_TOPSIM_HospitalManagement.pdf.

Zagal, J.P. and Bruckman, A. (2009), “Game Ontology Project”, Georgia, USA, 24 February, available at: https://www.gameontology.org.

Zhang, Y., Li, W., Zhang, C., Wang, Y. and Feng, S. (2026), “Reforming hospital management education with the P-MASE pedagogy: a randomized controlled trial”, *BMC Medical Education*, Vol. 26 No. 1, p. 601, doi: 10.1186/s12909-026-08948-8.

Inclusive Immersive E-Learning: Participatory Design of Accessible Virtual Reality. Supporting Diverse Learners Through Inclusive Medical Simulation Design

Tina Holmes[1], Teresa Crea[2], Xueqing Lu[3] and Graham Hannah[3]
[1]Faculty of Medicine & Health, University of New South Wales, Sydney, Australia
[2]Faculty of Arts, Design and Architecture, University of New South Wales, Sydney, Australia
[3]Immersive Technologies, University of New South Wales, Sydney, Australia

tina.holmes@unsw.edu.au
teresa.crea@unsw.edu.au
g.hannah@unsw.edu.au
xueqing.lu@unsw.edu.au

Abstract: Simulation-based education is a well-established approach for introducing medical students to real-world clinical scenarios within controlled learning environments; however, ensuring accessibility and inclusion for diverse learners remains an ongoing challenge in immersive e-learning contexts (Lukava et al., 2022). This article explores the unique challenges faced by neurodivergent medical students and emphasises the critical role of participatory design and lived experience in shaping VR environments that are inclusive and responsive to diverse needs. The paper advocates for a paradigm shift in medical education and VR design, one that not only supports neurodivergent medical students but also fosters inclusivity across the broader neurodivergent health professions, educational and workforce landscape. Furthermore, it highlights the intersection between Universal Design Principals (UDL), empirical evidence, and lived experience as essential for creating accessible and immersive VR learning environments.

Keywords: Lived Experience, Participatory Design, Virtual Reality, Neurodiversity, Universal Design for Learning

1. Introduction

Worldwide, mass-casualty incidents (MCI) are occurring with increasing frequency and complexity, driven by factors such as political conflict and natural disasters (Merson et al., 2025, Mundo et al., 2025). Medical responders are therefore required to make rapid, high- stakes decisions in unpredictable environments, often while exposed to significant psychosocial risks, including witnessing traumatic events and violence (Reale et al., 2023). Such exposures place medical responders at increased risk of psychological distress and physical harm (Utara, 2025, Mundo et al., 2025).

In response, simulation-based training for MCI management in senior medical education has progressively evolved through the adoption of advanced technologies, including virtual reality (VR) (Dzhumatovich and Salimrouhi, 2025). The increasing use of immersive platforms has enhanced experiential learning and clinical skills training; however, accommodations for neurodiverse learners and those with sensory accessibility needs remain under-represented in VR design and pedagogical research (Dahlstrom-Hakki et al., 2024, Lukava et al., 2022, Schmidt and Glaser, 2026, Schmidt et al., 2023, Wehrmann and Zender, 2024).

While VR offers a safe space for senior medical students to practise complex clinical competencies, its intensive cognitive and sensory demands, such as visual complexity, auditory load, and time-pressured interaction may create barriers to engagement for neurodivergent students (Giroux and Pélissier-Simard, 2021, Lukava et al., 2022, Creed et al., 2024). Effective instructional design therefore requires careful consideration of multisensory inputs, including visual, auditory, proprioceptive, and haptic feedback (Boyd et al., 2018). Many existing VR learning platforms insufficiently address cognitive diversity and accessibility, reinforcing the need for inclusive, learner-centred design approaches (Wehrmann and Zender, 2024, Lukava et al., 2022).

1.1 Project Background

1.1.1 Project Collaboration

This collaborative research project aims to address the need for inclusive VR Team Based Learning experiences for undergraduate medical students at the University of New South Wales (UNSW). Structured as a co-production initiative, the project brings together students, academic, and technology specialists to collaboratively refine instructional design and enhance learning outcomes (Strnadová et al., 2020). Using a

participatory design methodology, the research explores how immersive VR environments can support inclusive team-based medical education.

A key strength of the project is its interdisciplinary collaboration, integrating expertise from medical students, design students, academics, and immersive technology specialists. Supported by in-kind contributions from the School of Arts & Design, School of Clinical Medicine and Pro-Vice Chancellor Education (PVCE) Media & Immersive team, this collaborative partnership enhances both innovation and scalability. Delivered as a Work Integrated Learning (WIL) initiative, the project engages Master of Simulation and Immersive Technology (MSIT) students through successive capstone cohorts, each building on prior research and prototype development to support an iterative and cumulative design process. Undergraduate medical students contribute as clinical consultants, providing domain expertise and insights into the cognitive, sensory, and communication demands of clinical practice. As intended end users, their input ensures that VR scenarios are clinically relevant, pedagogically meaningful, and technically feasible.

Individuals with lived experience of neurodiversity and sensory accessibility needs are actively involved as co-designers across all stages of the project, including, ideation, prototyping, testing, and evaluation. The inclusion of team members with lived experience supports the identification of key challenges related to sensory processing and attentional challenges, and communication, including sources of overload, distraction, and breakdown. These insights are critical in designing for neurodivergent users and position lived experience act as a bridge between technical development and clinical application.

While the project was initially conceived as the development of a multiplayer VR platform grounded in Universal Design for Learning (UDL) principles, its scope has evolved through iterative design to incorporate lived experience, enabling a nuanced exploration of neurodivergent needs within immersive learning environments (CAST, 2024).

1.2 Multiplayer VR Trauma Simulation

1.2.1 Mass Casualty Incident: Trauma and Hypothermia onboard the Antarctic II Research Vessel

In this study, a mass casualty scenario is used in which students are deployed (via helicopter) to a remote maritime disaster site (onboard the AVR11) and tasked with triaging and stabilising patients. The simulation is designed to support the development of clinical management skills, as well as communication and teamwork in a high-pressure, resource- constrained setting.

1.2.2 Project Objective

To design and develop a virtual reality learning experience through a co-production framework that engages students, academics and technology specialists from ideation to prototype testing, ensuring meaningful inclusive design and collaborative innovation.

1.2.3 Infrastructure

The project employs simulation-based education delivered through a multiplayer VR platform developed in Unity game engine and implemented using the Meta Quest 3 head-mounted display and hand controllers. The system enables users to navigate avatars within a clearly defined operational environment, including a rescue helicopter and a ships deck, designed to replicate key spatial and workflow features of a real-world maritime emergency.

A core- decision support system, currently under development by MSIT and medical students, underpins the educational experience. The simulation is structured around a custom designed mass-casualty scenario developed by UNSW academics, representing a high-risk disaster onboard a research vessel. Consistent with simulation-based pedagogy, the environment prioritises psychological fidelity over sensory threat realism; no overt visible or auditory threat stimuli are included (Pedram et al., 2023). This approach allows users to focus on objectives such as team communication, triage decision-making, role coordination and patient management.

The scenario includes five simulated casualties with varying injuries and clinical presentations, enabling deliberate practice across multiple levels of clinical complexity. Users are orientated through an in-scenario pre-briefing tutorial., consistent with best practice simulation design. User actions including movement patterns, triage tagging decisions and patient management are captured to support structured debriefing and reflective

learning. An optional academic facilitator, represented by an in-simulation avatar, can be included to support facilitation, and reinforce educational objectives.

2. Study Design

Our project integrates a participatory co-design framework with UDL, foregrounding lived experience to inform flexible, accessible VR simulation design. UDL provides the pedagogical lens, while participatory co-design operationalises inclusion through learner-informed decision-making.

2.1 Universal Design for Learning

Universal Design for Learning offers a framework that informs the design of learning environments that promote inclusivity, accessibility, and meaningful cognitive challenge for every learner (CAST, 2024). This project is theoretically grounded in UDL, which frames the design and implementation of the VR simulation as an inclusive learning environment responsive to learner variability. UDL provides a lens through which learning is conceptualised not as a fixed outcome, but as emerging from the interplay between users, the environment and tasks (CAST, 2024).

Guided by UDL principals of multiple means of engagement (the 'why' of learning), representation (the 'what of learning), and action and expression (the 'how' of learning), the VR platform is intentionally designed to reduce barriers to participation and support neurodiverse learners, psychological safety and authentic performance (CAST, 2024). The prototype design has been conceptualised to incorporate multiple means for learners to engage with tasks, access information, and demonstrate understanding, enabling flexibility in how learning is experienced and enacted within the simulation. The principal of engagement focuses on fostering, affirming and sustaining learners' interests and identities. Representation highlights the importance of multiple ways of knowing and making meaning, while ensuring the inclusion of diverse identities, perspectives and narratives. Action and expression centres on the need to challenge exclusionary practices and to promote more accessible, inclusive, and user-centred learning environments (CAST, 2024).

2.1.1 UDL Methods in VR in the Study

Accessibility was central to the design of this multiplayer VR simulation. Accordingly, we operationalised key UDL principles within the VR environment to enhance flexibility in how students engage with content and demonstrate learning in immersive contexts. We explored design strategies to support diverse learners, including those with colour vision deficiency, auditory sensitivity, hearing impairment, and autism, with the aim of enhancing usability, comfort, engagement, and learning effectiveness.

Targeted interventions include high-contrast colour palettes and pattern overlays; adjustable Hz levels to accommodate sensory sensitivity; and multimodal feedback such as haptic cues, visual indicators, and in-environment virtual assistants. Additional design considerations include shape-based communication, multilingual support, and alternative interaction methods. These include simplified point and click modes to reduce cognitive load, verbal communication for team interaction, and gaze-based controls to minimise fine motor demands.

Importantly, these features were not implemented as a fixed or perfect solution but evolved through iterative co- design and ongoing refinement. Continuous feedback from individuals with lived experience and incremental improvements are central to the project.

3. Student Co-Design Projects

The following student co-design projects illustrate how UDL principals were operationalised through participatory design.

3.1 Auditory Sensitivity: Hz Level Controls to End Users

While there are VR systems that explicitly manipulate audio frequencies, most current systems address neurodivergence through broad sensory controls rather than precise frequency tuning. To address auditory sensitivity, this project explored selective, band-based frequency modulation rather than relying on broad sensory controls. Users can adjust specific audio frequency bands through selectable audio sources, presets, and real-time frequency visualisation. The approach enables personalise sensory regulation, supporting

neurodivergent users through greater control over audio input, and aligning with inclusive and user- centred design principals.

3.2 Colour Vision Deficiency

Colour blindness was explored by simulating visual impairments using Unity's channel mixer. Three environments were compared: non-optimised colours, colour friendly palettes, and palettes enhance with pattern overlays. In a pilot useability test (n=3), participants completed object selection tasks with randomised order; results showed increased reaction times with optimised colours, and further improvement when patterns were added, suggesting combined cues may enhance accessibility.

3.3 Hearing Impairment

This project explored how audio cues in VR can be supplemented or replaced with visual or haptic feedback. Two simulations were compared, one with multimodal feedback (visual, haptic, and audio) and another with limited cues. The prototype explores how different feedback designs influence attention and task performance, with future work needed to evaluate effectiveness and consistency.

3.4 Neurodivergence (Autism): Cognitive Overload

User interface design was examined to reduce sensory and cognitive overload for autistic users. Key features included simplified fonts, rounded buttons, low-stimulation colour palettes, multimodal feedback (including optional voice guidance), and adjustable audio-visual settings. A comparison with a poorly designed interface highlighted these improvements. The design acknowledges variability in user needs, and the limitations of the prototype in capturing real-world complexity.

Collectively, these student co-design projects highlighted the need for further usability testing with neurodivergent participants through first-person lived experience of the VR environments.

4. Neurodivergence and Lived Experience

While UDL is a significant framework bringing flexibility into the VR design and presentation of a learning environment it still is still subject to a degree of standardisation in its delivery through a technological framework. There remains a risk of mismatch between the bespoke processing needs of the individual user based on their specific neurodivergent needs and the technology platform used to deliver the learning. To understand this gap in more detail, we conducted an additional useability study based on lived experience feedback. We collaborated with neurodivergent participants, including neurodivergent members of the project team, to undertake a first-person reflective exploration of two distinct technology VR platforms utilising the phenomenological research methodologies of lived experience.

4.1 First Person Reflective Exploration: Method

Participant responses were tracked via a guided qualitative phenomenological first-person analysis paying close, minute by minute attention, to emotional reactions, sensory impressions, and cognitive processes within the virtual environment to capture the lived experience in real time. Responses we captured via a subjective think aloud narration and followed by an additional subjective reflection.

The experiment comprised two parts a) lived experience of the first-person subjective narrative b) personal reflective analysis. The exercise was repeated on two classes of VR headsets commercially available in 2025: the Apple Vision Pro and Meta Quest 3.

4.2 Preliminary Findings

In addition to findings related to UDL, this study includes a sample of subjective reflections derived from a lived experience exploration, presented as qualitative evidence. These reflections illustrate the types of unexpected insights that can emerge through detailed, first-person experiential inquiry. All reflections are reported anonymously to protect participant identity.

The following extract highlights how limited embodiment, poor spatial grounding, and interface complexity can lead to cognitive overload, disorientation, and delayed engagement in VR contents:

> The AVP presented only my hands, offering no view of the ground or my legs, which hindered my ability to ground myself in this new digital space. Instead, a vast digital sky and an array of menus filled my view, creating an immediate sense of being "a little overwhelming to understand where I start and where I finish". It took a significant amount of time to even process the options presented to me.

The participant described a tension between immersion and safety. The extract highlights a sense of being pulled between engagement in the virtual environment and the need for real world grounding. This loss of physical grounding contributed to a diminished awareness of their surroundings, resulting in feelings of disorientation and discomfort:

> Throughout my experience, the tension between immersion and a pervasive need for real-world awareness was palpable. While the initial creation of a virtual environment felt "beautiful," like a "movie", the sudden disappearance of my real surroundings "threw me off". The inability to see my legs or feet felt "off" and unsettling, triggering a desire for haptic feedback when interacting with virtual objects like a table, to better ground my physical presence.

The following extract highlights how gesture-based interactions in VR, when lacking tactile feedback and requiring sustained visual fixation, can feel unnatural, disrupt interaction flow and reduce user trust and confidence:

> My primary mode of interaction with the AVP involved looking at a dot and pinching my fingers to select. This interaction felt not like a normative experience [...] A pinch gesture feels intuitive in theory, but in practice it demands rigid fixation that breaks flow. A finger ray resembles pointing, but without muscular resistance it fails to build trust.

The participant describes how the design works against how they see and process information: *“my "dyslexic scanning tendencies," where the eyes tend to dart ahead to anticipate the next point, directly conflicted with the system's narrow tolerances”*

This extract describes how non-intuitive special set-up processes and inflexible interaction design can conflict with neurodivergent perceptual styles, increasing cognitive load and reducing usability:

> The process of spatial orientation and room setup was remarkably challenging and non-instinctual. While the system could auto-load a saved room preset, which beneficial for repeat users, I keenly felt the absence of an immediate option to declare a new environment for calibration. My "dyslexic nature," characterized by eyes shifting past targets to track the next movement, directly conflicted with the system's "rigid grammar of action”

The central theme of misalignment between VR design and neurodivergent cognitive, sensory, and perceptual needs emerged within the findings. The lived experienced highlighted reduced embodiment and spatial grounding, leading to disorientation and discomfort. Interaction methods, particularly those requiring sustained visual fixation and precise gestures, were perceived as cognitively demanding. Additionally, ridged interaction structures and limited flexibility conflicted with scanning behaviours, further increasing cognitive load. These findings highlight how current VR design can mistakably exclude neurodivergent users, underscoring the need for more flexible, and inclusive approaches informed by lived experience and UDL principals.

5. Insights and Recommendations

Normally, user interface design (UX) evaluates the extent to which a device conforms to user’s needs and enables intuitive interaction. However, in VR systems, interaction protocols often shift with each new product iteration. While UDL principals may be mandated, the interface shifts and technological developments across VR systems do not necessarily foster more fluid, inclusive learning or VR engagement experiences for neurodivergent users.

Some of the insights and recommendations to emerge from lived experience include:

- the need for more structured and prescriptive onboarding, ensuring that the user feel safe, grounded and in control when entering immersive environments;
- the importance of sensory channels and anchors to support orientation, predictability, and presence;
- the need to move beyond overly rigid or ‘forced narrowness’ in system design allowing, for diverse perceptual and sensory processing needs;
- the simplification of controls and action triggers that invertedly hinder neurodivergent users by increasing cognitive load or ambiguity in interaction.

- the need review all of the above with each new product iteration

Formative qualitative data generated through participatory design activities and lived experience recounts enables iterative design refinement throughout the development process, ensuring that accessibility considerations are embedded from the outset rather than applied retrospectively.

6. Limitations and Future Work

While our preliminary findings suggest that a multipronged approach combining UDL principals, empirical evidence, and lived experience insights through participatory co design may enhance accessibility in VR, several important limitations must be acknowledged.

First, heterogeneity within neurodivergent populations presents a significant challenge. Neurodivergence among medical students is not a uniformed experience, and sensory profiles vary widely both between individuals and within the same individual overtime. This variability limits the effectiveness of predefined solutions and reinforces the need for highly flexible and tailored systems.

Second, the proposed combined UDL and lived experience approach, including self-adjustment of audio parameters, may introduce usability barriers. Although user-adjustable EQ precision allows for precise control (e.g., reducing 4,000 Hz by -8 dB), it may also increase cognitive load or require technical understanding that the user may not have.

Third, the current work is based on preliminary, small scale, and potentially non-representative samples. Although, lived experience insights and participatory co-design provide valuable depth, findings cannot yet be generalised across broader neurodivergent populations or cultural contexts.

Finally, there are technical and implementation constraints. Shared VR environments present challenges related to processing performance and maintaining consistency of shared experiences. For example, disparities in audio filtering between users may affect team-based communication, shared attention, and collaborative task performance.

Future work will prioritise larger, more diverse, and longitudinal studies to better capture variability and changes in neurodivergent sensory experiences over time. More work is needed to explore collaborative implications of tailored audio in multiplayer VR environments, including how differing audio filters influence communication, presence and group dynamics. An important implication of this work is the need to further develop participatory co-design methodologies, enabling neurodivergent users to shape not only featured requirements but also VR interface design and system behaviour.

7. Conclusion

Mass-casualty incidents are increasing in frequency and complexity, placing greater demands on medical responders under intense cognitive and psychologically challenging conditions. As immersive VR platforms are increasingly utilised in medical education to prepare students for future clinical practice, the need for equitable and accessible design becomes critical. The literature underscores that despite their pedagogical potential, many VR learning environments fail to adequately account for neurodiversity and sensory accessibility, thereby risking exclusion of some learners from training opportunities (Shaw et al., 2023, Dahlstrom-Hakki et al., 2024, Giroux and Pélissier-Simard, 2021, Schmidt et al., 2023, Creed et al., 2024).

These limitations in VR learning environments may perpetuate a culture of non-disclosure, as unaddressed sensory or cognitive barriers can discourage learners from engaging or accessing appropriate supports (Green et al., 2025). Addressing these issues within VR-based training is particularly important given the sensory intensity and time-pressured nature of such simulations. Increasing awareness of neurodivergence, alongside the development of inclusive VR design, has the potential to create more accessible, supportive, and immersive learning environments.

Formative qualitative data generated through participatory design activities and lived experience informed iterative design refinement throughout the development process, ensuring that accessibility considerations are embedded from the outset rather than applied retrospectively. However, neurodivergent users have highly individual and context-dependent sensory needs, making universal design inherently complex. Approaches such as user-adjustable parametric EQ and neurodivergent-informed audio profiles could represent future directions for creating VR learning environments that are not only accessible but adapted to individual sensory preferences.

This study supports a paradigm shift in both medical education and VR design, moving beyond standardised approaches toward more inclusive, flexible, and neurodivergent-affirming practices. By integrating a multipronged approach that combines UDL principles with empirical evidence and lived-experience insights through participatory co-design, this work contributes to a more nuanced and responsive understanding of accessibility.

Ethics Declaration

This study is covered by the ethics approval obtained from the UNSW Human Research Ethics Committees (HC230349 - The Use of Media and Immersive Technologies in Higher Education).

AI Declaration

This manuscript was copyedited and proofread using AI-assisted (Microsoft Co-pilot) tools in accordance with IPEd guidelines. All intellectual content and argumentation are the authors own.

Acknowledgements

This research was undertaken in collaboration with undergraduate Medicine students and postgraduate Master of Simulation and Immersive Technology capstone students, University of New South Wales, Sydney.

References

BOYD, L. E., DAY, K., STEWART, N., ABDO, K., LAMKIN, K. & LINSTEAD, E. 2018. Leveling the Playing Field: Supporting Neurodiversity Via Virtual Realities. *Technology and Innovation,* 20**,** 105–116.

CAST. 2024. *Universal Design for Learning Guidelines version 3.0* [Online]. Available: https://udlguidelines.cast.org [Accessed 2024].

CREED, C., AL-KALBANI, M., THEIL, A., SARCAR, S. & WILLIAMS, I. 2024. Inclusive Augmented and Virtual Reality: A Research Agenda. *International Journal of Human–Computer Interaction,* 40**,** 6200–6219.

DAHLSTROM-HAKKI, I., ALSTAD, Z., ASBELL-CLARKE, J. & EDWARDS, T. 2024. The impact of visual and auditory distractions on the performance of neurodiverse students in virtual reality (VR) environments. *Virtual Reality,* 28**,** 29.

DZHUMATOVICH, A. S. & SALIMROUHI, E. 2025. Simulation-based education as a tool for enhancing the quality of medical education. *Journal of Preventive and Complementary Medicine,* 4**,** 45–59.

GIROUX, M. & PéLISSIER-SIMARD, L. 2021. Shedding light on autistic traits in struggling learners: A blind spot in medical education. *Perspectives on Medical Education,* 10**,** 180–186.

GREEN, E. J., BROWN, M. E. L., VANCE, G. H. S. & KEENAN, I. D. 2025. Undergraduate medical education for neurodivergent students: a scoping review. *BMC Medical Education,* 26**,** 103.

LUKAVA, T., MORGADO RAMIREZ, D. Z. & BARBARESCHI, G. 2022. Two sides of the same coin: accessibility practices and neurodivergent users' experience of extended reality. *Journal of Enabling Technologies,* 16**,** 75–90.

MERSON, J. R., MIGEON, M., MARN, J. & SHOWSTARK, M. 2025. Mass Casualty Incidents. *Physician Assistant Clinics,* 10**,** 771–791.

MUNDO, W., O'CONNOR, T. & MCGLADREY, L. 2025. Operational Stress Injuries in Disaster Responders: A Framework for Supporting Mental Health in Climate Crisis. *Wilderness & Environmental Medicine,* 36**,** 35–43.

PEDRAM, S., KENNEDY, G. & SANZONE, S. 2023. Toward the validation of VR-HMDs for medical education: a systematic literature review. *Virtual Reality,* 27**,** 2255–2280.

REALE, C., SALWEI, M. E., MILITELLO, L. G., WEINGER, M. B., BURDEN, A., SUSHEREBA, C., TORSHER, L. C., ANDREAE, M. H., GABA, D. M. & MCIVOR, W. R. 2023. Decision-making during high-risk events: a systematic literature review. *Journal of Cognitive Engineering and Decision Making,* 17**,** 188–212.

SCHMIDT, M. & GLASER, N. Rethinking the Extended Reality and Autism Research Agenda: Toward Neurodiversity-Affirming, Transdiagnostic, Inclusive Design. *In:* KRüGER, J. M., SCHMIDT, M., MIKROPOULOS, A., KOUTROMANOS, G., PEDROSA, D., BECK, D., MYSTAKIDIS, S., SMITH-NUNES, G., PEñA-RIOS, A. & RICHTER, J., eds. Immersive Learning Research Network, 2026// 2026 Cham. Springer Nature Switzerland, 165–181.

SCHMIDT, M. M., LEE, M., FRANCOIS, M. S., LU, J., HUANG, R., CHENG, L. & WENG, Y. 2023. Learning Experience Design of Project PHoENIX: Addressing the Lack of Autistic Representation in Extended Reality Design and Development. *J Form Des Learn,* 7**,** 1–19.

SHAW, S. C. K., FOSSI, A., CARRAVALLAH, L. A., RABENSTEIN, K., ROSS, W. & DOHERTY, M. 2023. The experiences of autistic doctors: a cross-sectional study. *Frontiers in Psychiatry,* Volume 14 - 2023.

STRNADOVá, I., DOWSE, L. & WATFERN, C. 2020. Doing research inclusively: Guidelines for co-producing research with people with disability.

UTARA, I. G. A. G. A. 2025. The Silent Aftermath: Addressing Mental Health in Emergency Healthcare Providers After Mass Casualty Events. *Bali Journal of Anesthesiology,* 9**,** 55–56.

WEHRMANN, F. & ZENDER, R. 2024. Inclusive Virtual Reality Learning: Review and 'Best-Fit' Framework for Universal Learning. *Electronic Journal of E-Learning,* 22**,** 74–89.

Student Ambassadors for Sustainable AI EdTech Knowledge Sharing in Higher Education

Hyein Cho, Min Soo Lee, Young-Woong Ko, Yong Geun Kim, Jion Kim, Minho Kim and Tae-Il Yoon
Hallym University, Chuncheon, Republic of Korea

chohyein@hallym.ac.kr
minsoolee@hallym.ac.kr
youngwoongko@hallym.ac.kr
yogkim@hallym.ac.kr
jion972@hallym.ac.kr
kimminho@hallym.ac.kr
icarus44@hallym.ac.kr

Abstract: This conceptual paper develops a framework for appraising institutional knowledge assets produced through AI EdTech dissemination. Knowledge mobilisation research emphasises dissemination, interpretation, usefulness, uptake and context, but does not typically isolate representational quality as a distinct asset-level appraisal dimension. This gap is especially consequential when verification asymmetry and promotional incentive co-occur: claims may exceed inspectable evidence, overstate attribution or omit material conditions and risks even when assets remain visible, accessible and reusable. The paper therefore conceptualises representational quality through four criteria: claim traceability, attribution restraint, condition disclosure and risk representation. It further specifies a student ambassador model in which case-linked direct-user knowledge is combined with evidence access, expert review and revision rights across a governed production-and-maintenance lifecycle. Production-stage verification constrains distortion before dissemination, while post-hoc fidelity audit supports correction after publication; engagement analytics substitute for neither. Five falsifiable propositions address construct distinctness, verification, direct-user informational advantage, conditional diagnostic value and maintenance after material change. The paper reports no empirical findings. Its contribution is to position representational quality as an explicit property of AI EdTech knowledge assets and to specify the conditions and governance mechanisms under which it should be appraised.

Keywords: Representational Quality; AI EdTech; Sustainable Knowledge Sharing; Student Ambassadors; Knowledge Mobilisation

1. Introduction

Higher education institutions are increasingly adopting AI-enabled educational technologies, including adaptive learning, automated feedback, learning analytics, AI assistants and intelligent tutoring systems. Reviews have mapped applications, outcomes, stakeholder perceptions, ethics and implementation challenges (Bond et al. 2024; Zawacki-Richter et al. 2019), but have paid less attention to how local implementation experience becomes durable institutional knowledge. Institutional value therefore depends not only on adoption but also on whether implementation knowledge can be examined, shared and reused. The transition from local experience to reusable knowledge constitutes a dissemination gap between AI-enabled innovation and sustainable institutional learning.

AI EdTech intensifies this gap because dissemination evidence is often difficult to interpret. AI systems may be difficult for users to inspect (Miao and Holmes 2023), while observed outcomes may be difficult to attribute to the technology rather than to novelty, selection, assessment design or concurrent teaching changes (Shadish, Cook and Campbell 2002). AI-enabled education also raises privacy, bias, transparency and learner-data concerns (Miao and Holmes 2023; Slade and Prinsloo 2013). Institutional communication may therefore produce claims that are stronger or clearer than the available evidence warrants.

Knowledge mobilisation (KMb), knowledge brokering and boundary-spanning research explain why evidence must be interpreted, adapted and connected to practice rather than merely transmitted (Akkerman and Bakker 2011; Carlile 2004; Davies, Nutley and Walter 2008; Levin 2013; Ward 2017). Recent KMb-product research emphasises user needs, relevance, usefulness, accessibility, stakeholder involvement and uptake (Elliott et al. 2024; Kastner et al. 2020; Sharp et al. 2022; Ziam et al. 2024). These concerns are important but do not by themselves establish whether circulating claims are proportionate to their evidential basis.

This paper develops a conceptual framework that distinguishes representational quality from visibility, accessibility and reusability, and specifies a student ambassador model for governing these asset-quality

dimensions. It further identifies verification asymmetry and promotional incentive as conditions under which explicit representational-quality appraisal is expected to add the greatest diagnostic value. Students are treated as trained, case-linked partners rather than as intrinsically more credible communicators. The model combines case-linked student participation with structural safeguards for evidence access, review and revision. The paper addresses three research questions:

RQ1. What asset-level criteria are required to appraise AI EdTech knowledge assets, and what does representational quality add to established criteria of visibility, accessibility and reusability?

RQ2. Which mechanisms create a need for representational-quality appraisal, and under what conditions does that dimension add the greatest diagnostic value?

RQ3. How should a student ambassador model be specified to strengthen representational quality during production and maintenance, and what evidence would count against its claims?

The paper addresses these questions analytically. It reports no empirical findings and does not present an effectiveness evaluation. Its claims are stated as propositions that subsequent empirical work could support, qualify or disconfirm.

2. Theoretical Framework

2.1 AI EdTech as Institutional Knowledge Assets

AI EdTech contributes to institutional learning when implementation experience is converted into knowledge that can be examined, interpreted and reused. KMb conceptualises this process as interpretation, adaptation, circulation and connection to local action rather than linear transfer (Davies, Nutley and Walter 2008; Levin 2013; Ward 2017). For the purposes of this framework, sustainable knowledge sharing is defined as the continued availability, interpretability, evidential warrant and revisability of knowledge assets as technologies and implementation conditions change. In this framework, institutional knowledge assets are digital outputs that preserve and organise implementation experience for subsequent interpretation and use beyond the setting in which it was produced.

Recent KMb research examines knowledge products in relation to their purpose, intended users, stakeholder involvement, adaptation and evaluation (Elliott et al. 2024; Kastner et al. 2020; Sharp et al. 2022; Ziam et al. 2024). Together, these strands provide a basis for treating AI EdTech dissemination outputs as knowledge assets rather than as transient communication products.

2.2 Student Ambassadors as Knowledge Mobilisers and Boundary Spanners

The students-as-partners (SaP) perspective provides a pedagogical and normative basis for student participation. SaP positions students as collaborators and emphasises reciprocity, shared responsibility and co-creation, thereby recognising experiential knowledge as a legitimate contribution (Bovill 2020; Healey, Flint and Harrington 2014). In AI EdTech initiatives, such experiential knowledge may capture aspects of use that formal implementation records overlook, including confusion, discontinuation, workarounds and support needs. Student participation is therefore relevant not because students are inherently more credible, but because direct experience may provide access to implementation conditions that are otherwise difficult to reconstruct.

Boundary-spanning theory explains how actors and artefacts connect communities that differ in vocabulary, evidence standards and priorities, while knowledge brokering emphasises the interpretation and adaptation of knowledge for use across settings (Akkerman and Bakker 2011; Carlile 2004; Rycroft-Smith 2022; Ward 2017). In the proposed framework, student ambassadors are conceptualised as performing these interpretive and connective functions, while digital knowledge assets serve as boundary objects through which implementation knowledge can circulate across learner, faculty, technical and institutional communities.

Building on these strands, the framework distinguishes direct-user status from the ambassador role. Direct-user status provides access to case-specific experiential evidence, whereas the ambassador role concerns translating, contextualising and mobilising that evidence across institutional boundaries. The proposed informational advantage is therefore bounded to case-linked direct users and should not be generalised to students as a category. Where ambassadors are not direct users, the framework does not assume equivalent access to case-specific experiential evidence or an equivalent informational advantage. This distinction locates the potential

contribution of student ambassadors in the combination of experiential access and boundary-spanning capacity rather than in student status itself.

2.3 Theoretical Synthesis and Design Assumptions

Drawing on the preceding literature, the framework adopts four design assumptions. First, knowledge assets require sufficient contextual information to support later interpretation and adaptation (Carlile 2004; Davies, Nutley and Walter 2008; Ward 2017). Second, stakeholder participation should be treated as substantive involvement in knowledge production rather than as a guarantee of asset quality; SaP scholarship emphasises reciprocity, shared responsibility and genuine partnership (Bovill 2020). Third, digital knowledge assets can function as boundary objects when they retain sufficient shared structure to connect communities with different vocabularies, evidence standards and priorities (Akkerman and Bakker 2011; Carlile 2004). Fourth, continued applicability requires maintenance because material changes in systems, policies or implementation conditions may alter the conditions under which a knowledge asset remains warranted for current use.

Together, these assumptions distinguish participation from the governance of knowledge assets. Experiential access can strengthen contextual reconstruction, but it does not by itself ensure the quality of the resulting knowledge asset. These assumptions nevertheless leave unresolved how the quality of circulating knowledge assets themselves should be assessed.

3. Asset Quality in AI EdTech Dissemination

3.1 Asset-Quality Dimensions from Knowledge Mobilisation

The KMb literature does not offer a single established taxonomy of knowledge-asset quality. Existing research spans diverse purposes, actors, processes and evaluation targets, including relevance, usefulness, stakeholder involvement, accessibility, uptake, effects and impacts (Elliott et al. 2024; Kastner et al. 2020; Sharp et al. 2022; Ward 2017; Ziam et al. 2024). These contributions inform KMb design and evaluation but do not establish a common set of properties for appraising the circulating asset itself.

For this framework, three baseline asset-level properties are distilled because they concern the asset's capacity for circulation, interpretation and reuse rather than its production process or downstream outcomes. Visibility concerns whether intended audiences can locate the asset (Davies, Nutley and Walter 2008; Levin 2013; Ward 2017). Accessibility concerns whether the asset can be understood by audiences with different knowledge and technical backgrounds (Akkerman and Bakker 2011; Carlile 2004; Elliott et al. 2024; Ward 2017). Reusability concerns whether sufficient contextual information is retained for later interpretation and adaptation (Carlile 2004; Davies, Nutley and Walter 2008; Ward 2017). These dimensions establish a baseline for whether an asset can be found, interpreted and reused, but do not establish whether its claims are proportionate to their evidential basis.

3.2 Verification Asymmetry and Promotional Incentive

Verification asymmetry arises when circulating claims exceed what producers or audiences can readily verify from the available evidence. In this framework, it has two distinct sources. Mechanism opacity concerns whether users can understand how an AI system generated an output (Miao and Holmes 2023), whereas attribution uncertainty concerns whether observed educational outcomes can reasonably be attributed to the technology rather than to plausible alternatives (Shadish, Cook and Campbell 2002; Weidlich, Hicks and Drachsler 2024). The two are analytically distinct: greater system transparency does not establish causal attribution, and stronger causal evidence does not make system operation transparent.

A separate condition is promotional incentive, which arises when institutions communicate their own outcomes in contexts shaped by reputational, recruitment or funding interests. AI education discourse is susceptible to hyperbole and value-laden claims (Selwyn 2022), although direct evidence of distortion in institutional AI EdTech communication remains limited. Evidence from university science communication nevertheless suggests that exaggerated institutional press releases can be associated with exaggerated subsequent coverage (Sumner et al. 2014), providing a relevant but bounded analogy. Within the framework, promotional incentive creates pressure toward favourable representation, while verification asymmetry makes resulting overstatement or omission more difficult to detect or challenge.

When verification asymmetry and promotional incentive are both substantial, an asset may therefore remain visible, accessible and reusable while overstating attribution, omitting material conditions or underrepresenting risks. Their conjunction defines a high-risk condition in which improved dissemination may amplify rather than correct distortion. These conditions do not define representational quality; rather, they identify when explicit appraisal of that dimension is expected to add the greatest diagnostic value. This limitation of the baseline dimensions motivates an additional asset-level appraisal dimension.

3.3 Representational Quality: Definition and Boundaries

Representational quality is defined here as the degree to which an asset's substantive claims are traceable to and proportionate to the available evidence, with sufficient context, limitations, uncertainty and ethical risks retained for responsible interpretation. This formulation aligns with broader concerns about transparency, uncertainty and ethical responsibility in educational AI and data use (Miao and Holmes 2023; Slade and Prinsloo 2013). Representational quality is therefore treated as a property of the circulating artefact rather than of the producer's intention or the audience's response.

Four criteria specify its appraisal. Claim traceability concerns whether substantive claims can be linked to specified sources or documented observations. Attribution restraint concerns whether outcomes are attributed to the technology only to the extent supported by the available design and evidence (Shadish, Cook and Campbell 2002). Condition disclosure concerns whether material features of the setting, cohort, duration, support arrangements and concurrent changes are reported. Risk representation concerns whether known limitations, uneven effects and ethical concerns are represented in the asset rather than omitted from the circulating account.

Representational quality is related to, but not reducible to, other forms of quality judgement. Factually accurate or persuasive statements may still misrepresent an outcome by omitting material conditions or implying unsupported causality, while ethical compliance concerns how evidence is obtained and used. Transparency concerns the visibility of processes or evidence, and credibility reflects audience judgement; representational quality instead concerns whether an asset's claims remain proportionate to the evidence and sufficiently contextualised. It may therefore correlate with visibility, accessibility or reusability without being reducible to any of them; empirical separability remains a proposition to be tested.

4. Student Ambassador Knowledge-Sharing Model

4.1 Integrated Framework

Figure 1 synthesises the preceding theoretical arguments into an integrated framework linking risk conditions, asset quality and the student ambassador lifecycle. Verification asymmetry and promotional incentive constitute the framework's principal risk conditions identified in Section 3. Student ambassadors contribute experiential interpretation and boundary-spanning functions grounded in the partnership and knowledge-mobilisation literature (Akkerman and Bakker 2011; Bovill 2020; Ward 2017). The framework pairs these functions with structural safeguards for evidence access, expert review and revision across production and maintenance.

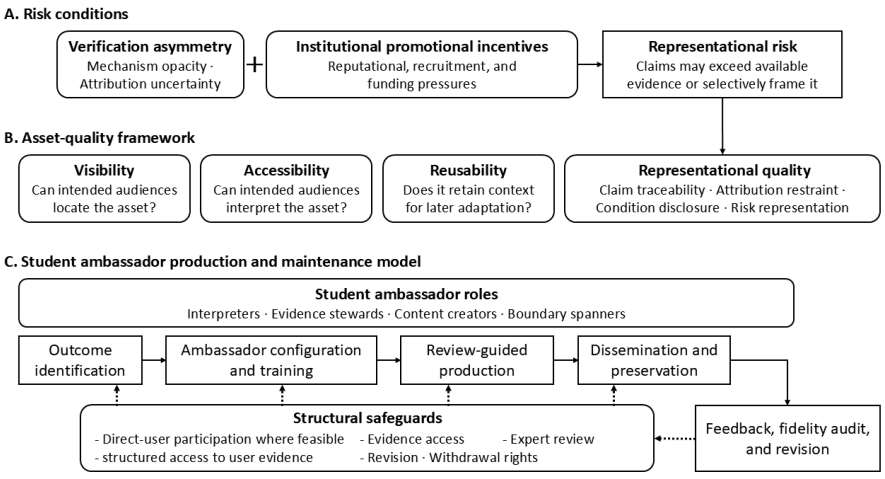


Figure 1: Student Ambassador AI EdTech Knowledge-Sharing Framework

4.2 Five-Stage Production-and-Maintenance Lifecycle

Drawing on the theoretical arguments developed in Sections 2 and 3, Table 1 specifies an author-derived five-stage production-and-maintenance lifecycle linking core activities, structural safeguards, outputs and asset-quality targets. Production-stage verification comprises claim–evidence mapping, expert review and revision before release, while post-publication fidelity audit reassesses whether an asset continues to represent its underlying evidence and implementation conditions faithfully. Together, these mechanisms distinguish pre-publication verification from post-publication maintenance and support evidential scrutiny and revision across the asset lifecycle.

Table 1: Five-Stage Student Ambassador Knowledge-Sharing Model

Stage	Core activity	Structural safeguards and outputs	Mechanism and asset-quality target
1 Outcome identification	Screen cases for relevance, evidence inspectability, ethical acceptability and reuse potential; map evidence before drafting.	Evidence access and case-eligibility screen; evidence map and review criteria.	Sets an evidential floor and excludes cases that cannot be responsibly qualified; targets representational quality.
2 Ambassador configuration and training	Prioritise case-linked direct users; otherwise provide structured access to user evidence. Train in AI literacy, interviewing, evidence stewardship and attribution restraint.	Direct-user participation or structured user-evidence access; role allocation, templates and revision/withdrawal rights.	Converts experiential access into interpretable evidence without presuming superior credibility; targets accessibility and representational quality.
3 Review guided production	Create assets using claim-evidence mapping, condition disclosure and risk prompts; revise through expert review before release.	Evidence access, expert review, claim-evidence matrix and pre-publication revision gate.	Tests evidential warrant before dissemination and retains material context; targets representational quality and accessibility.
4 Dissemination and preservation	Disseminate approved assets; preserve versions, provenance, metadata and retrieval information in durable repositories.	Version control and traceable provenance; published and archived assets.	Increases findability while preserving context for later adaptation; targets visibility and reusability.
5 Feedback, fidelity audit and revision	Use engagement analytics for uptake, fidelity audit for representational quality, and change monitoring to trigger revision or withdrawal.	Expert re-review, revision/withdrawal rights, audit record and archived versions.	Separates uptake from evidential warrant and supports quality maintenance; targets all four dimensions.

4.3 Ambassador Roles and Outputs

The model assigns student ambassadors four complementary roles: interpreters, content creators, evidence stewards and boundary spanners. They interpret user experience and context, translate knowledge into appropriate formats, steward claim–evidence links and connect learner perspectives with faculty, technical and institutional priorities, extending established interpretive and boundary-spanning functions (Akkerman and Bakker 2011; Carlile 2004). These roles position ambassadors across interpretation, production and revision rather than one-way dissemination.

The model yields reviewed digital artefacts and accumulated knowledge about their production, dissemination and revision. The former are appraised against the four asset-quality dimensions, while the latter informs subsequent production and revision.

5. Propositions and Boundary Conditions

5.1 Propositions

The framework's claims are stated as propositions rather than design intentions. Each is paired with evidence that would count against it.

5.1.1 Proposition 1

Representational quality is empirically distinguishable from visibility, accessibility and reusability.

Evidence against this proposition would be empirical inseparability from these dimensions or no incremental information beyond them.

5.1.2 Proposition 2

Production-stage verification will yield higher initial representational quality than comparable workflows without such verification.

Evidence against this proposition would be no difference in initial representational quality between the two workflows.

5.1.3 Proposition 3

Case-linked direct-user ambassadors will demonstrate greater condition disclosure than non-direct-user or institutional communication actors under comparable access to documentary evidence and review support.

Evidence against this proposition would be no such advantage, or a disadvantage, under matched production conditions.

5.1.4 Proposition 4

The incremental diagnostic value of representational-quality appraisal will be greatest when verification asymmetry and promotional incentive are both substantial.

Evidence against this proposition would be no greater diagnostic value under this joint high-risk condition than under other conditions.

5.1.5 Proposition 5

Post-publication maintenance will better preserve representational quality when material changes affect an asset's continued applicability.

Evidence against this proposition would be no advantage for maintained over unmaintained assets under such conditions.

5.2 Counter-Argument to Direct-User Advantage

Proposition 3 admits an important counter-argument. Direct experiential access does not eliminate institutional dependence, which may constrain students' willingness or capacity to represent unfavourable evidence. This concern is consistent with students-as-partners scholarship emphasising reciprocity, shared responsibility and genuine partnership as conditions of meaningful participation (Bovill 2020).

The framework therefore treats direct-user advantage as informational rather than motivational and limits it primarily to condition disclosure. Institutional dependence may therefore remain a constraint on attribution restraint and risk representation, reinforcing the framework's use of evidence access, expert review and revision rights as structural safeguards.

5.3 Boundary Conditions

The framework has two principal boundary conditions. First, the incremental diagnostic value of representational-quality appraisal is expected to be greatest when verification asymmetry and promotional incentive are both substantial; under lower-risk conditions, its added value may be more limited. Second, the proposed informational advantage applies specifically to case-linked direct users and should not be generalised to students as a category. For non-user ambassadors, structured access to direct-user evidence may mitigate this limitation but does not establish an equivalent informational advantage.

6. Discussion

6.1 Theoretical Contributions

The paper makes three theoretical contributions. First, it positions representational quality as an explicit asset-level property within institutional KMb. Existing KMb and product-oriented research addresses a broad range of concerns, including relevance, usefulness, stakeholder processes, uptake and impact (Elliott et al. 2024; Kastner et al. 2020; Sharp et al. 2022; Ziam et al. 2024). The framework adds an explicit appraisal of whether circulating claims remain proportionate to their evidential basis and sufficiently contextualised for responsible interpretation. Representational quality therefore complements, rather than replaces, visibility, accessibility and reusability.

Second, the framework differentiates mechanism opacity from attribution uncertainty as distinct sources of verification asymmetry. The former concerns whether system operations can be inspected, whereas the latter concerns whether observed educational outcomes can reasonably be attributed to the technology. This distinction separates technical transparency from causal warrant. Promotional incentive is theorised as a separate institutional risk condition, with the incremental value of representational-quality appraisal expected to be greatest when verification asymmetry and promotional incentive are both substantial.

Third, the framework provides a conditional account of student participation by separating direct-user informational access from the ambassador's boundary-spanning role. Its contribution lies in specifying the source of that advantage as case-linked experiential access combined with boundary-spanning capacity, rather than student status itself. The proposed advantage is bounded to case-linked direct users and primarily concerns condition disclosure rather than overall asset quality. The framework further formulates its central claims as falsifiable propositions, allowing the proposed construct distinctions, mechanisms and boundary conditions to be empirically supported, qualified or disconfirmed.

6.2 Practical Implications

For practice, universities should treat AI EdTech dissemination as a governed knowledge-management process rather than a sequence of publicity outputs. Knowledge assets should preserve sufficient evidence, context and provenance to support later interpretation and review. Visibility and engagement remain useful indicators of circulation and uptake, but they do not establish that claims remain evidentially warranted or current.

Governance should extend across the asset lifecycle. Before publication, verification should test whether claims are adequately supported and contextualised. After publication, engagement analytics should assess circulation and use, while fidelity audit should reassess representational quality. Material changes in system features, policies or implementation conditions should trigger review, revision or archival replacement where continued applicability would otherwise be misleading.

Student ambassador programmes should also differentiate roles according to evidential access. Case-linked direct users can contribute first-hand knowledge of implementation conditions, whereas non-user ambassadors may contribute interviewing, interpretation and production capabilities but should be supported by structured access to direct-user evidence. Evidence access, expert review and meaningful revision or withdrawal rights should therefore function as cross-cutting safeguards rather than as optional additions to student participation.

6.3 Limitations and Future Research

This study is limited by its conceptual design. The proposed construct, appraisal criteria, mechanisms and student ambassador model have not yet been empirically validated, including whether representational quality can be reliably distinguished from the other asset-quality dimensions and improved through the proposed governance mechanisms.

Future research should therefore begin by developing and testing a structured appraisal protocol for the four representational-quality criteria—claim traceability, attribution restraint, condition disclosure and risk representation—and applying it to AI EdTech knowledge assets produced in actual student ambassador programmes. Initial studies could compare condition disclosure between case-linked direct-user and non-user ambassadors, test whether structured evidence access improves representational quality among non-user ambassadors, and examine changes in representational quality before and after expert review. These designs would allow the framework's core mechanisms to be assessed under feasible implementation conditions while

examining which safeguards are associated with improvements in particular dimensions of representational quality. Subsequent matched comparisons with institutional communication staff could further test whether the proposed informational advantage of direct-user ambassadors is specific to condition disclosure rather than indicative of superior overall asset quality.

7. Conclusion

This paper reframes AI EdTech dissemination as a problem of knowledge-asset governance rather than circulation alone. Its central argument is that visibility, accessibility and reusability alone do not establish whether circulating accounts are proportionate to their evidential basis or sufficiently contextualised for responsible interpretation. Representational quality addresses this gap through claim traceability, attribution restraint, condition disclosure and risk representation, with its diagnostic value expected to be greatest when verification asymmetry and promotional incentive are both substantial.

The student ambassador framework translates this argument into a governed production-and-maintenance lifecycle. The framework does not assume that students are inherently more credible communicators; its rationale rests on the combination of case-linked experiential access, boundary-spanning capacity and governance safeguards. The broader implication is that sustainable institutional learning depends not simply on disseminating AI EdTech experience more widely, but on preserving the evidential and contextual conditions under which shared knowledge remains interpretable, warranted and revisable. Without such governance, wider dissemination may amplify poorly warranted claims rather than strengthen institutional learning.

Acknowledgements

This research was supported by the ANCHOR program (Glocal University30) through the Gangwon ANCHOR Center, funded by the Ministry of Education (MOE) and the Gangwon State (G.S.), Republic of Korea (2026-ANCHOR-10-009).

Ethics Declaration

Ethical approval was not required for this conceptual paper because the study did not involve the recruitment of human participants or the collection, analysis or reporting of personal or participant-derived data.

AI Declaration

Generative AI tools were used to support language editing and consistency checking. All arguments, citations and final wording were independently reviewed and verified by the authors, who take full responsibility for the manuscript. AI-generated material was not used as empirical evidence or research data.

References

Akkerman, S.F. and Bakker, A. (2011) “Boundary crossing and boundary objects”, *Review of Educational Research*, Vol 81, No. 2, pp 132–169, doi: 10.3102/0034654311404435.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S.W. and Siemens, G. (2024) “A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour”, *International Journal of Educational Technology in Higher Education*, Vol 21, Article 4, doi: 10.1186/s41239-023-00436-z.

Bovill, C. (2020) “Co-creation in learning and teaching: The case for a whole-class approach in higher education”, *Higher Education*, Vol 79, No. 6, pp 1023–1037, doi: 10.1007/s10734-019-00453-w.

Carlile, P.R. (2004) “Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries”, *Organization Science*, Vol 15, No. 5, pp 555–568, doi: 10.1287/orsc.1040.0094.

Davies, H., Nutley, S. and Walter, I. (2008) “Why ‘knowledge transfer’ is misconceived for applied social research”, *Journal of Health Services Research & Policy*, Vol 13, No. 3, pp 188–190, doi: 10.1258/jhsrp.2008.008055.

Elliott, S.A., Bialy, L., Scott, S.D. and Hartling, L. (2024) “Exploring methods for creating or adapting knowledge mobilization products for culturally and linguistically diverse audiences: A scoping review”, *Archives of Public Health*, Vol 82, Article 111, doi: 10.1186/s13690-024-01334-0.

Healey, M., Flint, A. and Harrington, K. (2014) *Engagement Through Partnership: Students as Partners in Learning and Teaching in Higher Education*, Higher Education Academy, York, [online], available at: https://www.advance-he.ac.uk/knowledge-hub/engagement-through-partnership-students-partners-learning-and-teaching-higher (accessed 1 September 2026).

Kastner, M., Makarski, J., Hayden, L., Lai, Y., Chan, J.Y.T., Treister, V., Harris, K., Munce, S.E.P., Holroyd-Leduc, J., Graham, I.D. and Straus, S.E. (2020) "Improving KT tools and products: Development and evaluation of a framework for creating optimized, Knowledge-activated Tools (KaT)", *Implementation Science Communications*, Vol 1, Article 47, doi: 10.1186/s43058-020-00031-7.

Levin, B. (2013) "To know is not enough: Research knowledge and its use", *Review of Education*, Vol 1, No. 1, pp 2–31, doi: 10.1002/rev3.3001.

Miao, F. and Holmes, W. (2023) *Guidance for Generative AI in Education and Research*, UNESCO, Paris, doi: 10.54675/EWZM9535.

Rycroft-Smith, L. (2022) "Knowledge brokering to bridge the research-practice gap in education: Where are we now?", *Review of Education*, Vol 10, No. 1, Article e3341, doi: 10.1002/rev3.3341.

Selwyn, N. (2022) "The future of AI and education: Some cautionary notes", *European Journal of Education*, Vol 57, No. 4, pp 620–631, doi: 10.1111/ejed.12532.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston.

Sharp, C.A., Boaden, R.J., Dixon, W.G. and Sanders, C. (2022) "Does the process of developing products for knowledge mobilisation from healthcare research influence their uptake? A comparative case study", *Implementation Science Communications*, Vol 3, Article 132, doi: 10.1186/s43058-022-00360-9.

Slade, S. and Prinsloo, P. (2013) "Learning analytics: Ethical issues and dilemmas", *American Behavioral Scientist*, Vol 57, No. 10, pp 1510–1529, doi: 10.1177/0002764213479366.

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C.A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F. and Chambers, C.D. (2014) "The association between exaggeration in health related science news and academic press releases: Retrospective observational study", *BMJ*, Vol 349, Article g7015, doi: 10.1136/bmj.g7015.

Ward, V. (2017) "Why, whose, what and how? A framework for knowledge mobilisers", *Evidence & Policy*, Vol 13, No. 3, pp 477–497, doi: 10.1332/174426416X14634763278725.

Weidlich, J., Hicks, B. and Drachsler, H. (2024) "Causal reasoning with causal graphs in educational technology research", *Educational Technology Research and Development*, Vol 72, No. 5, pp 2499–2517, doi: 10.1007/s11423-023-10241-0.

Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F. (2019) "Systematic review of research on artificial intelligence applications in higher education—where are the educators?", *International Journal of Educational Technology in Higher Education*, Vol 16, Article 39, doi: 10.1186/s41239-019-0171-0.

Ziam, S., Lanoue, S., McSween-Cadieux, E., Gervais, M.-J., Lane, J., Gaid, D., Chouinard, L.J., Dagenais, C., Ridde, V., Jean, E., Fleury, F.C., Hong, Q.N. and Prigent, O. (2024) "A scoping review of theories, models and frameworks used or proposed to evaluate knowledge mobilization strategies", *Health Research Policy and Systems*, Vol 22, Article 8, doi: 10.1186/s12961-023-01090-7.

Pedagogical Innovation through Generative AI: Benefits, Challenges, and Responsible Use in Higher Education

Muhammad Imran[1,2] and Norah Almusharraf[1]
[1]Prince Sultan University, Riyadh, Saudi Arabia,
[2]Khazar University, Baku, Azerbaijan.

mimran@psu.edu.sa

Abstract: The rapid integration of Generative Artificial Intelligence (GAI) into higher education has transformed how students engage in learning and educational activities, including personalized learning, creativity, and knowledge production. This study aims to explore the adoption patterns, perceived benefits, and challenges of using GAI tools such as ChatGPT, DeepSeek, and Google Gemini in higher education. A quantitative survey was conducted with 102 students from higher education institutions across various disciplines and degree levels. Findings revealed that 77.5% of participants reported using GAI daily, master's students reported the highest benefit perception scores, while ethical and accuracy concerns were identified as the most pressing challenges; PhD students adopted a more critical stance. Findings further suggest that GAI has a positive influence on academic engagement and creativity; however, concerns about the ethical implications, responsible use, and originality remain critical challenges. This study, by highlighting both opportunities and challenges, provides actionable insights for teachers, learners, policymakers, and technology developers. It emphasizes the need for artificial intelligence AI literacy and responsible integration to ensure that GAI fosters sustainable and meaningful pedagogical innovation in higher education.

Keywords: Generative AI, Academic Engagement, Pedagogical Innovations, Digital Learning Outcomes, AI Ethics

1. Introduction

GAI has rapidly emerged as a transformative technology development across multiple sectors, including education (Younas et al., 2025; Bragg et al., 2021). In an academic context, students and professionals from various fields are utilizing GAI tools to tackle complex tasks effectively, such as information synthesis, academic writing, and problem-solving, often achieving substantial time-efficiency gains. So, the number of students who are integrating these tools into their educational activities has been accelerated markedly, as GAI tools are perceived to facilitate skill development through adaptive and personalized support mechanisms and are crucial for students to learn new skills in a unique and efficient manner (Mubarak & Selimin, 2023; Maqbool et al., 2024; Alghazo et al., 2025).

Recent studies by Elanwar (2025), Lang et al. (2025), and Mittal et al. (2024) revealed that GAI, including ChatGPT, Large Language Models (LLMs), and DALL-E, enhances students' learning and facilitates the creation of automated content while developing their creative problem-solving capabilities. Henceforth, to develop critical thinking skills among students, educational institutions are increasingly utilizing these tools on a broader scale, according to Luckin et al. (2022). Nevertheless, current research lacks an understanding of how generative AI affects students at the university level; this powerful technology may, however, enhance engagement and academic outcomes (Shamim et al., 2026). Thus, the primary objective of this research paper is to examine the impact of GAI on higher education, specifically in terms of academic management, creativity, and self-reported learning gains, as well as the associated challenges and ethical concerns.

This current study aimed to fill these gaps and provide a remarkable, empirically supported insight into the impact of Generative Artificial Intelligence tools on shaping students' academic engagement, creativity, and self-reported learning gains at the university level in Pakistan. In this research, we studied the pedagogical impacts of GAI at the university level, whereas the existing literature has examined its impact at a general level. By analyzing adoption patterns, perceived benefits, and challenges, this research will not only advance theoretical understanding but also yield practical implications for educators, policymakers, and technology developers. The findings demonstrated that recurrent use of generative AI correlates positively with perceived academic benefits, highlighting the importance of AI literacy and structured integration strategies. This highlights the transformative role of generative AI in enhancing learning efficiency, while also emphasizing the ethical and accuracy challenges that must be addressed for its sustainable adoption.

1.1 Research Objectives

- To quantify and characterize the adoption patterns of generative AI tools among university students across disciplines, degree levels, and institution types.
- To evaluate the perceived academic benefits of generative AI tools on learning outcomes.
- To identify and categorize the primary challenges (technical integration, accuracy concerns, ethical implications) associated with generative AI adoption in academic contexts.

1.2 Research Questions

- What are the patterns of GAI adoption among Pakistani university students?
- How does GAI usage influence academic engagement, creativity, and learning outcomes?
- What challenges are associated with GAI adoption in higher education?

2. Literature Review

In recent years, the inclusion of generative Artificial Intelligence in education, particularly in higher education, has dramatically increased interest among educators, researchers, students, and policymakers. This literature review synthesizes findings from recent studies published in renowned journals, such as those by Springer, to explore the impact of generative AI on academic engagement, creativity, and self-reported learning gains among university students. The review is structured around three key themes: (1) the role of generative AI in enhancing academic engagement, (2) its influence on fostering creativity, (3) its effects on self-reported learning gains, challenges while using AI tools, and (4) ethical considerations for university students. In the following subsections, we will highlight recent literature on these targeted research areas.

Recent studies (Mohamed et al., 2025; Owoseni et al., 2024; Sharma et al., 2025) suggest that generative AI tools, such as ChatGPT and other large language models, can significantly enhance student engagement by providing personalized learning experiences. For instance, Yu and Yao (2024) found that AI-driven platforms can adapt to individual learning styles, offering tailored feedback and resources that keep students motivated and engaged. Similarly, studies by Liu et al. (2025) and Song et al. (2025) found that generative AI can facilitate interactive learning environments, enabling students to participate more actively in discussions and problem-solving. However, concerns have been raised about the potential for over-reliance on AI tools, which may lead to reduced critical thinking and self-directed learning. As noted by George et al. (2024), while AI can enhance engagement, it is crucial to balance its use with traditional pedagogical methods to ensure students remain intellectually challenged.

Students at the university level are involved in technology-related academic activities to enhance their engagement, and GAI is playing a vital role in these activities. For instance, the research by Khan and Ullah (2024) and Meng et al. (2025) found that ChatGPT-supported classrooms in higher education significantly improved participation rates and reduced academic disengagement. Similarly, Du (2025), Imran and Almusharraf (2023), and Jabeen (2023) reported that AI-driven personalized tutoring systems enhanced time-on-task behaviors and improved intrinsic motivation among undergraduate learners. However, researchers caution that constant reliance on AI may lead to "engagement inflation," in which students appear engaged but demonstrate shallow processing of content (Klingbeil et al., 2024). These findings align with the argument that AI should be integrated as a supplementary tool rather than a replacement for critical human-led instruction.

According to Feng et al. (2025), AI-based writing assistants improved argumentation quality and coherence in undergraduate essays, while Srivastava and Agarwal (2024) demonstrated improved retention rates in STEM subjects through multimodal AI-generated simulations. Nevertheless, differential outcomes have been reported across disciplines; for instance, humanities students often question AI's interpretative validity compared to STEM learners. After careful analysis of these studies, we conclude that ethical issues, such as bias in AI outputs and data privacy risks, remain pressing challenges that can influence both the reliability and fairness of self-reported learning gains when relying entirely on GAI technologies during educational activities.

Although past studies have examined AI in education broadly, few focus specifically on GAI's role in enhancing engagement, creativity, and measurable self-reported learning gains among university students in developing nations. Existing studies often focus on the broader applications of AI or its ethical implications, leaving the pedagogical and psychological impacts of generative AI underexplored. This study addresses this

gap by providing empirical evidence on how generative AI tools influence university students' academic experiences, offering a foundation for future research and practice in this emerging field. This research addresses this gap by providing empirical evidence from Pakistan, thereby contributing to both the local and global discourse on the integration of educational AI. The literature reviewed underscores the transformative potential of generative AI in higher education, particularly in enhancing academic engagement, fostering creativity, and improving learning outcomes. However, it also highlights the need for a balanced approach that integrates AI tools with traditional teaching methods while addressing ethical and pedagogical challenges.

3. Theoretical Framework and Methodology

For this study, we applied an integrated conceptual framework combining the Technology Acceptance Model (TAM) (Davis, 1989) and the Self-Determined Learning (Heutagogy) (Blaschke, 2012) framework. TAM posits that perceived usefulness and perceived ease of use are primary determinants of technology adoption. To measure the impact of GAI, we conceptualize self-rated proficiency as a substitute for ease of use and 'Perceived Benefits' as a measure of usefulness. Therefore, a quantitative, cross-sectional survey design was employed to enable statistical examination of patterns, correlations, and group differences in GAI usage and perceptions. For data analysis and visualization, we used Python libraries such as Matplotlib, NumPy, and Pandas in Google Colab.

This study investigates the impact of generative artificial intelligence on university-level education, with a specific focus on students' academic engagement, creativity, and learning outcomes. A stratified sampling method ensures representation across disciplines and degree levels. We employed a structured questionnaire, developed through a rigorous process using Microsoft Forms, to collect data from participants. The initial item pool was generated based on a comprehensive literature review of constructs related to technology adoption (Lai et al., 2022; Lai, 2017), academic engagement (Alrashidi et al., 2016; Perkmann et al., 2021), and creativity (Egan et al., 2017). The questionnaire was then reviewed by a panel of three experts in educational technology and psychometrics to establish content and face validity. The questionnaire was distributed digitally to ensure accessibility and broad participation. The study aims to provide empirical insights into how generative AI technologies are influencing higher education settings and to identify potential benefits and challenges associated with their integration into academic practices. Descriptive statistics, Spearman correlations, with Tukey post-hoc tests, and confirmatory factor analysis were performed using SPSS v28 and AMOS v26. Significance level was set at $\alpha = 0.05$.

4. Findings and Discussion

This section presents the results on the frequency and purpose of GAI tool use in higher education. Figure 1 shows a word cloud visualization of the distribution of academic fields reported by participants.



Figure 1: Major Field of Studies of Participants

The following Figure 2 illustrates the distribution of generative AI tool utilization frequency among 102 participants. The data reveals a predominance of daily usage, reported by 79 individuals, corresponding to 77.5% of the sample. Weekly use was reported by 16 participants (15.7%), while monthly use was reported by 5 participants (4.9%). Infrequent exertion patterns, including yearly use and non-use, were minimal, each reported by only 1 participant (1.0%). The clear concentration on the "Daily" category indicates greater integration of generative AI tools into participants' routines.

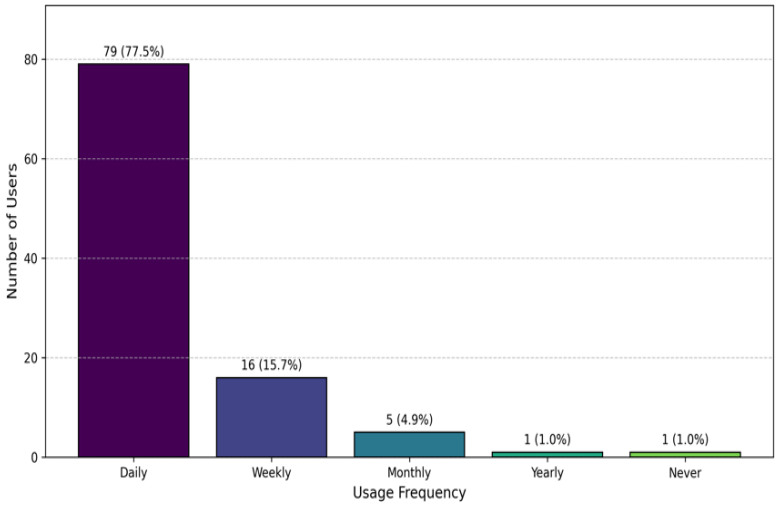


Figure 2: Usage Frequency Distribution of GAI Tools

From an interpretative standpoint, the prevalence of daily practice implies that generative AI tools have likely become essential operational aids for most users in this cohort. Such patterns could be driven by academic, professional, or personal productivity needs, reflecting both the perceived utility and accessibility of these technologies. The minimal representation of infrequent or non-users suggests that barriers to adoption in this population are either low or outweighed by perceived benefits, warranting further exploration of the specific drivers of sustained high-frequency use.

4.1 Influence of GAI on Academic Engagement, Creativity, and Self-reported Learning Gains

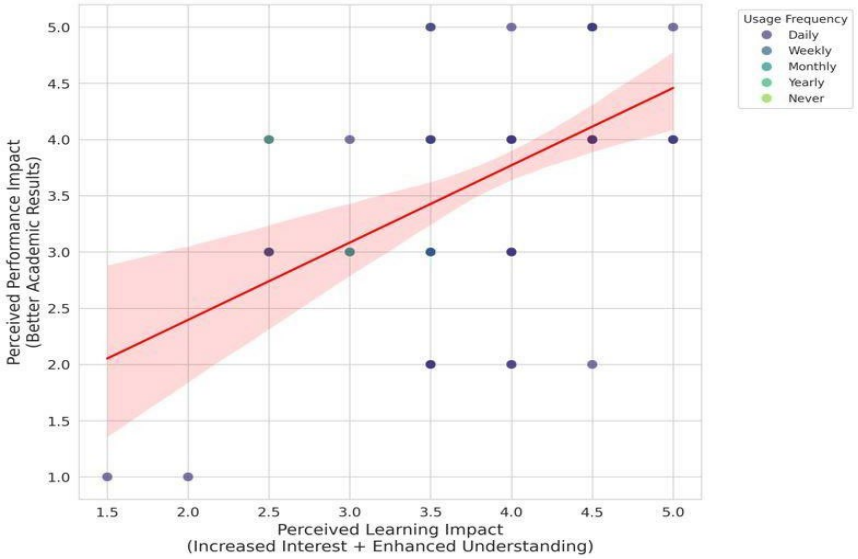


Figure 3: Relationship between perceived learning impact and perceived performance impact, with regression line (95% CI) indicating a weak but positive association across different usage frequencies.

Figure 3 displays a scatter plot with an overlaid linear regression line to visually explore the relationship between Perceived Learning Impact (increased interest and enhanced understanding) and Perceived Performance Impact (better academic results), categorized by GAI usage frequency. Visually, the regression line indicates a potential positive trend, suggesting that higher self-reported learning impacts may align with stronger perceived performance outcomes. The shaded region represents the 95% confidence interval, which appears narrower at higher response tiers (between 4.0 and 5.0). However, it is important to note that these observations are purely descriptive; no formal statistical testing (such as correlation coefficients or significance tests) was performed to confirm the strength, direction, or statistical significance of this association. Similarly, descriptive clustering shows that data points in the upper-right quadrant are predominantly associated with self-reported Daily users, while lower-impact scores appear more scattered among Weekly or Monthly users. While this descriptive pattern suggests a possible link between usage

frequency and perceived utility, these differences are not validated by formal comparative or inferential analyses. Consequently, any apparent variations between usage-frequency groups represent exploratory observations rather than definitive statistical conclusions and must be interpreted with caution.

4.2 Challenges Associated with GAI Adoption

Figure 4 highlights a violin plot visualizing the distribution of concern levels (measured on a Likert scale from 1 to 5) across three primary categories of AI-related challenges: Integration, Accuracy, and Ethical considerations. This type of plot integrates the features of a boxplot with a kernel density estimate, offering both summary statistics and the distribution shape for each challenge type. The horizontal axis categorizes the challenge type, while the vertical axis quantifies the level of concern.

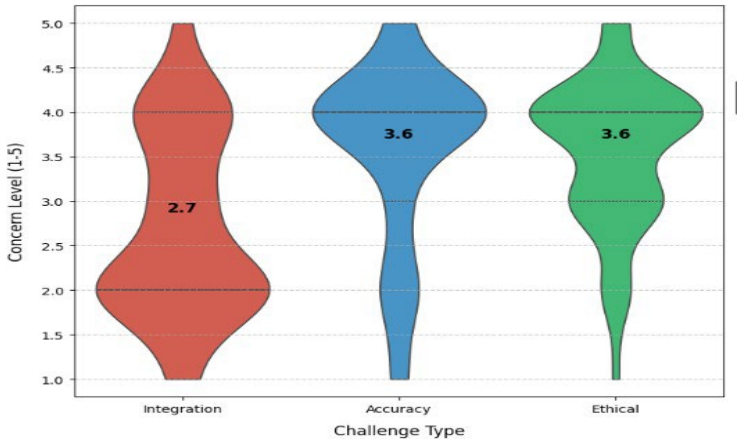


Figure 4: Perceived Challenges in AI adoption

Superimposed on each violin is a bold numerical label denoting the mean concern level for that category. Specifically, integration registers a lower mean concern of 2.7, whereas both Accuracy and Ethical challenges exhibit elevated concern levels of 3.6, suggesting heightened awareness or uncertainty in those domains. The differences in central tendency indicate that respondents are relatively less apprehensive about Integration issues, possibly due to increased familiarity or perceived manageability. At the same time, concerns about Accuracy and Ethics remain pronounced, likely reflecting persistent technical limitations and unresolved normative questions.

The findings of this research paper indicate that GAI is deeply embedded in students' academic activities, with periodic use far exceeding occasional use. Interestingly, master's students reported higher perceived benefits than both undergraduates and PhD candidates, perhaps reflecting optimal alignment between academic workload and AI support. Ethical and accuracy concerns remain key barriers to uninhibited adoption, a significant concern in past research (Chan & Tsi, 2024; Afzaal et al., 2020). Contrary to stronger claims in earlier drafts, the association between self-rated proficiency and self-reported academic improvement was statistically significant but weak (Spearman's $\rho = 0.09$, $p < .001$). This indicates that while higher proficiency is associated with slightly greater perceived academic improvement, proficiency alone explains very little variance in perceived outcomes.

This research work demonstrates its obviousness by highlighting perceptual differences at various academic levels within the university in Pakistan. The results of this study show that postgraduate students reported the highest perceived advantages, as they primarily used GAI tools for their research work, to complete assignments, and various research activities. In contrast, undergraduates benefited less, possibly due to their limited academic maturity, while PhD students expressed greater scepticism, reflecting a critical stance on AI's limitations.

Students reported that some responses were not accurate. Similarly, Weekes and Eskridge (2022) noted that AI can simultaneously act as a productivity enhancer and a cognitive risk. Ethical concerns remain salient, with students expressing unease about plagiarism, academic honesty, and bias in AI outputs. These issues highlight the urgency of implementing AI literacy programs that prioritise responsible use, critical appraisal of AI-generated content, and adherence to academic integrity standards. In Pakistan, where regulatory frameworks for GAI in higher education are still in the developing phase, these interventions would be critical for the smooth integration of GAI at the university level.

The findings further highlight intradisciplinary differences, as STEM students reported higher benefits than those in the humanities. This trend supports the findings of Yan et al. (2025), who found significant applicability of AI in technical and quantitative domains. Without discipline-specific frameworks, there is a risk of deepening existing educational disparities, particularly in resource-constrained contexts where institutional and socioeconomic barriers already shape access to digital tools (Munir et al., 2025). Universities should therefore adopt tailored strategies to ensure equitable integration across disciplines, recognizing the varying pedagogical needs of STEM and humanities learners.

Finally, the findings would advance the discourse on generative AI by reinforcing its pedagogical value while critically addressing its inherent limitations. AI tools can foster engagement, creativity, and efficiency. However, their transformative potential depends on balanced integration with traditional teaching, the cultivation of higher-order thinking skills, and the establishment of robust ethical safeguards. Also, longitudinal research is necessary to assess the enduring effects of AI adoption on knowledge retention, independent problem-solving, and critical reasoning. By situating the Pakistani selected universities case within the global discourse, this study may underscore the imperative for evidence-based policies and context-sensitive interventions that maximize the benefits of generative AI while safeguarding educational equity and intellectual integrity.

5. Conclusion

This study provides preliminary empirical observations regarding the integration of GAI tools among the selected Pakistani university students. Results revealed that 77.5% of participants use GAI regularly, with usage frequency and self-rated proficiency correlating with perceived benefits. Master's students disclosed the highest benefit perception scores, while ethical and accurate concerns were pinpointed as the most pressing challenges. The determinations further highlighted that, within this exploratory sample, GAI use is weakly associated with self-reported engagement and creativity, while also underscoring the importance of addressing ethical implications and ensuring responsible use. This preliminary analysis also offers insightful suggestions and recommendations for educators, policymakers, and learners, paving the way for future researchers to investigate GAI at a broader level and explore its impact in higher education. It further concludes that the higher prevalence of day-to-day GAI utilization and the tools that support it play a fundamental role as essential academic aids. At the same time, the correlation between proficiency and perceived benefits suggests that targeted AI literacy programs could amplify their educational value. Master's students' higher benefit perception points to a stage in academic progression where GAI tools optimally support research, writing, and problem-solving demands. Future research should explore longitudinal effects and the role of educators in mediating AI-driven learning experiences to maximize benefits for students.

6. Recommendations

This study proposes pivotal recommendations for educators, policymakers, and researchers. For educational practice, there is a need to develop and integrate mandatory AI literacy modules into the university curriculum to address the identified challenges and ethical considerations, along with prompt engineering, to save academic integrity and bias. Current analysis also advises policymakers to define acceptable applications of GAI that minimize plagiarism concerns during the evaluation of students' assignments, etc. Finally, this work would pave the way for future researchers, suggesting that to track the long-term impacts of GAI, longitudinal studies with mixed-methods approaches are needed, combining large-scale surveys and in-depth interviews.

Acknowledgment

The authors would like to thank Prince Sultan University for financial and technical support to complete this study.

Conflict Statement

The authors declare no conflict of interest.

AI Declaration

We used Grammarly Premium for proofreading and language improvement in this study.

Ethical Declaration

During this study, all ethical principles were strictly followed; no human or animal cells or personal data were collected or used.

References

Afzaal, M., Ahmad, S., Imran, M., & Xiangtaoc, D. (2020). Artificial intelligence, context, and meaning making in language: A rationalization approach. *International Journal of Future Generation Communication and Networking*, *13*(3), 115-122.

Alghazo, S., Rababah, G., El-Dakhs, D. A. S., & Mustafa, A. (2025). Engagement strategies in human-written and AI-generated academic essays: A corpus-based study. *Ampersand*, 100237.

Alrashidi, O., Phan, H. P., & Ngu, B. H. (2016). Academic engagement: an overview of its definitions, dimensions, and major conceptualisations. *International Education Studies*, *9*(12), 41-52.

Bizami, N. A., Tasir, Z., & Kew, S. N. (2023). Innovative pedagogical principles and technological tools capabilities for immersive blended learning: a systematic literature review. *Education and Information Technologies*, *28*(2), 1373-1425.

Blaschke, L. M. (2012). Heutagogy and lifelong learning: A review of heutagogical practice and self-determined learning. *The International Review of Research in Open and Distributed Learning*, *13*(1), 56-71.

Bragg, L. A., Walsh, C., & Heyeres, M. (2021). Successful design and delivery of online professional development for teachers: A systematic review of the literature. *Computers & education*, 166, 104158.

Chan, C. K. Y., & Tsi, L. H. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, 83, 101395.

Davis, F. D. (1989). Technology acceptance model: TAM. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, *205*(219), 5.

Du, Q. (2025). How artificially intelligent conversational agents influence EFL learners' self- regulated learning and retention. *Education and Information Technologies*, 1-67.

Egan, A., Maguire, R., Christophers, L., & Rooney, B. (2017). Developing creativity in higher education for 21st century learners: A protocol for a scoping review. *International Journal of educational research*, *82*, 21-27.

Elanwar, R. (2025). Creative Learning Using Generative AI: Empowering Problem-Solving and Programming Skills. In *Human-Computer Creativity: Generative AI in Education, Art, and Healthcare* (pp. 115-133). Springer.

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology*, *29*(1).

George, A. S., Baskar, T., & Srikaanth, P. B. (2024). The erosion of cognitive skills in the technological age: How reliance on technology impacts critical thinking, problem- solving, and creativity. *Partners Universal Innovative Research Publication*, *2*(3), 147- 163.

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464.

Imran, M., Almusharraf, N., Abdellatif, M. S., & Abbasova, M. Y. (2024). Artificial Intelligence in Higher Education: Enhancing Learning Systems and Transforming Educational Paradigms. *International Journal of Interactive Mobile Technologies*, *18*(18).

Jabeen, M. (2023). Exploring ChatGPT's Role in Creative Writing: A Short Review. *International Review of Literary Studies*, *5*(2), 32-34.

Khan, O., & Ullah, F. (2024). Impact of ChatGPT on Academic Engagement and Personalized Learning among Postgraduate Students- A Literature Review. *Research Journal of Psychology*, *2*(3), 153-179.

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, *160*, 108352.

Lai, J. W., De Nobile, J., Bower, M., & Breyer, Y. (2022). Comprehensive evaluation of the use of technology in education–validation with a cohort of global open online learners.
Education and Information Technologies, *27*(7), 9877-9911.

Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. *JISTEM-Journal of Information Systems and Technology Management*, *14*(1), 21-38.

Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025). Transforming Education with Generative AI (GAI): Key Insights and Future Prospects. *IEEE Transactions on Learning Technologies*, *18*, 230-242.

Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence*, *9*, 100438.

Luckin, R., Cukurova, M., Kent, C., & Du Boulay, B. (2022). Empowering educators to be AI- ready. *Computers and Education: Artificial Intelligence*, *3*, 100076.

Maqbool, M. A., Asif, M., Imran, M., Bibi, S., & Almusharraf, N. (2024). Emerging e-learning trends: a study of faculty perceptions and impact of collaborative techniques using fuzzy interface system. *Social Sciences & Humanities Open*, *10*, 101035.

Meng, X., Yang, B., Yang, L., Zhang, J., & Liu, Y. (2025). A Novel AI-Empowered, Student- Centered Teaching Strategy for Large Classes in Higher Education. *International Journal of Science and Mathematics Education*, 1-29.

Mittal, U., Sai, S., & Chamola, V. (2024). A comprehensive review on generative AI for education. *IEEE Access*. *12*, 142733-142759.

Mohamed, F. N., Azhar, J., Yasmeen, S., Hussain, I., & Khawar, M. (2025). Generative Artificial Intelligence and Personalized Learning Environment: Challenges and Opportunities. *Southern Journal of Computer Science*, *1*(01), 1-36.

Mubarak, Z. K., & Selimin, M. A. (2023). Significance of innovative learning skills in the era of education 4.0. *International Journal of Sustainable Construction Engineering and Technology*, *14*(3), 339-352.

Munir, M. T., Li, B., Carter, S., & Hussain, S. (2025). Engineering education's Odyssey with ChatGPT: Opportunities, challenges, and theoretical foundations. *International Journal of Mechanical Engineering Education*, (online first). https://doi.org/10.1177/03064190251337477

Owoseni, A., Kolade, O., & Egbetokun, A. (2024). Enhancing personalised learning and student engagement using generative AI. In *Generative AI in Higher Education: Innovation Strategies for Teaching and Learning* (pp. 123-150). Springer.

Perkmann, M., Salandra, R., Tartari, V., McKelvey, M., & Hughes, A. (2021). Academic engagement: A review of the literature 2011-2019. *Research policy*, *50*(1), 104114.

Sharmin, L., Kalima, R., Imran, M., Chauca, M., & Balladares, J. (2026). Teachers' perceptions of AI in supporting students' learning within a globally diverse digital settings. *Discover Education*, *5*(1), 75.

Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: a systematic review of educational impact. *Discover Sustainability*, *6*(1), 1-24.

Song, Y., Huang, L., Zheng, L., Fan, M., & Liu, Z. (2025). Interactions with generative AI chatbots: unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. *International Journal of Educational Technology in Higher Education*, *22*(1), 12.

Srivastava, A. P., & Agarwal, S. (Eds.). (2024). *Utilizing AI tools in academic research writing*. Hershey, PA: IGI Global.

Weekes, T. R., & Eskridge, T. C. (2022, June). Responsible human-centered artificial intelligence for the cognitive enhancement of knowledge workers. In *International Conference on Human-Computer Interaction* (pp. 568-582). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-21707-4_41

Yan, W., Li, B., & Lowell, V. L. (2025). Integrating Artificial Intelligence and Extended Reality in Language Education: A Systematic Literature Review (2017–2024). *Education Sciences*, *15*(8), 1066.

Younas, M., El-Dakhs, D. A. S., & Jiang, Y. (2025). A comprehensive systematic review of AI- driven approaches to self-directed learning. *IEEE Access, 13*, 38387–38403. https://doi.org/10.1109/ACCESS.2025.3546319

Yu, C., & Yao, G. (2024). Enhancing student engagement with ai-driven personalized learning systems. *International Transactions on Education Technology (ITEE)*, *3*(1), 1-8.

Designing Meaningful AI Learning Activities in Pre-Service Teacher Education

Antonín Jančařík[1], Kateřina Jančaříková[2,3] and Martina Šmejkalová[1]
[1]Faculty of Education, Charles University, Prague, Czech Republic
[2]Faculty of Science, Jan Evangelista Purkyně University in Ústí nad Labem, Czech Republic
[3]Department of Biology, Faculty of Science, Humanities and Education, Technical University of Liberec, Studentská Czech Republic

antonin.jancarik@pedf.cuni.cz
martina.smejkalova@pedf.cuni.cz
katerina.jancarikova@ujep.cz
katerina.jancarikova@tul.cz

Abstract: This article examines how artificial intelligence can be meaningfully integrated into pre-service teacher education through a case study approach. Rather than treating AI as a general technological innovation or a universal tool for increasing efficiency, the article focuses on three concrete teaching interventions in the subject didactics of Czech language, mathematics, and biology. Each case was designed and implemented in university courses for future teachers and illustrates a different way in which AI can support students in professionally demanding areas where they often experience uncertainty, anxiety, or lower self-efficacy. The study is grounded in the assumption that the educational value of AI does not lie in the mere use of a digital tool, but in its integration into subject-specific and pedagogically meaningful activities. In Czech language education, AI supports work with authentic language data, particularly when formulating corpus queries, interpreting corpus results, and didactically transforming language material for school use. In mathematics, AI functions as a consulting partner in the design of empirical investigations and the statistical processing of data in students' final theses, helping students understand the relationship between research questions, data, methods of analysis, and interpretation. In biology, AI-based applications facilitate the identification of plants and animals using visual and audio inputs, thereby allowing students to focus not only on correct species naming but also on observation, comparison, verification, and argumentation. Across the three cases, the analysis shows that AI can serve as didactic scaffolding. It lowers the entry barrier to complex activities that student teachers might otherwise avoid, while still requiring them to check, refine, interpret, and critically evaluate AI outputs. The aim is therefore not to replace expert judgement, but to make it more visible and to support its development. The cases also show that AI-supported activities become educationally meaningful only when they are clearly connected to future classroom practice and when students are guided to reflect on the limits and responsible use of AI. The article formulates common principles for designing AI-supported activities in teacher education: subject-didactic anchoring, orientation toward professionally demanding areas, critical work with AI outputs, explicit links to future work with pupils, development of AI literacy, and support for students' professional confidence. In this sense, AI can contribute not only to technological competence but also to subject reasoning, pedagogical judgement, and the didactic transformation of knowledge.

Keywords: Artificial Intelligence, Teacher Education, Self-Efficacy, Subject Didactics, AI literacy, Pedagogical Design

1. Introduction

The growing availability of artificial intelligence tools opens up new questions for pre-service teacher education (Bae et al., 2024). In both academic and public debate, AI has often been presented as a technological means to make information retrieval, the creation of teaching materials, or data analysis more efficient (Xiaoyu et al., 2025; Yadav et al., 2026). Attention is increasingly shifting, however, to how AI-based learning activities should be designed to genuinely support pedagogical goals and the professional competences of future teachers (Blonder et al., 2025). One of the approaches discussed is AI-supported scaffolding in lesson planning, that is, the use of AI as a supportive tool for designing, refining and reflecting on learning activities (Cai et al., 2025), as well as for developing critical thinking (Oliveira et al., 2025).

This article assumes that AI is meaningful in teacher education if it is understood not merely as an isolated technological innovation, but as part of a concrete subject-didactic activity. This assumption follows the TPACK framework, according to which meaningful integration of technology into teaching emerges only through the interconnection of teachers' technological, pedagogical and content knowledge (Mishra & Koehler, 2006). The article focuses primarily on technological content knowledge, which is subsequently linked to pedagogical content knowledge, thereby creating space for the development of TPACK. A future teacher thus learns not only to use a new tool within a subject, but also to interpret its outputs and transform them into work with pupils.

The article presents three practical examples from the preparation of Czech language, mathematics and biology teachers. All three cases share the fact that AI enters areas where students often experience uncertainty, anxiety, or low confidence in their own abilities. AI is understood here as a means of helping to bridge the gap between theoretical knowledge and its practical use in teaching.

2. Theoretical Background

The article's starting point is a conception of teacher education as a process in which content, pedagogical, and technological knowledge are mutually interconnected. This approach aligns with the TPACK framework, which builds on Shulman's concept of pedagogical content knowledge and emphasises that meaningful use of technology in teaching cannot be separated from knowledge of content and pedagogical intention (Shulman, 1986; Mishra & Koehler, 2006). For this article, the connection between technological content knowledge and pedagogical content knowledge is especially important: AI first supports students in working with subject-specific procedures, such as formulating a corpus query, designing statistical processing or recognising biological species, and then creates space for their didactic transformation into work with pupils. This emphasis aligns with current research on generative AI in teacher education, which shows that working with AI to design learning activities can reveal and develop pre-service teachers' pedagogical content knowledge (Blonder et al., 2025; Wen & Wen, 2024).

In the case of AI, this connection is particularly important. The mere ability to generate a text, propose a statistical procedure or recognise the species of an organism does not in itself mean that the output has educational value. Such value emerges only when the student can assess the quality of the output, relate it to subject knowledge, adapt it to pupils' needs, and reflect on its limitations. This ability to handle AI critically and knowledgeably corresponds to a broader understanding of AI literacy, which includes not only technical understanding of AI tools but also the ability to interpret their outputs, understand their limitations and use them responsibly in specific situations (Long & Magerko, 2020). AI is therefore not understood as a substitute for expert or pedagogical judgement, but as a tool that can activate and make such judgement visible (Blonder et al., 2025; Celik et al., 2026).

The second starting point of the article is the concept of self-efficacy. Bandura defines self-efficacy as an individual's belief in their own ability to perform a particular activity successfully; this belief affects whether a person begins the activity, how much effort they invest in it and whether they persist when obstacles arise (Bandura, 1977). In teacher education, this concept is developed as teacher self-efficacy, that is, teachers' belief in their own ability to support pupils' learning and manage the demands of teaching (Tschannen-Moran et al., 1998). In teacher preparation, students often encounter areas where they are unsure of their competence. These may include working with the formal language of corpus queries, statistical evaluation of data or identifying natural objects in the field. Such areas are not marginal; on the contrary, they are important components of a teacher's professional equipment. Low self-efficacy, however, may lead students to avoid them, replace them with simpler procedures, or omit them from their future practice (Bandura, 1977; Tschannen-Moran et al., 1998).

In such situations, AI can serve as a supportive tool or a form of scaffolding. It does not take over responsibility for expert decisions, but it lowers the entry barrier, provides working drafts, enables iteration and supports the student's active reflection. Research on AI-supported scaffolding in teacher education shows that ChatGPT can support the development of lesson-planning skills, especially when structured through targeted questions, proposal critique, and lesson-plan revisions (Cai et al., 2025). From the perspective of teacher education, it is therefore essential that work with AI does not end with accepting or rejecting its output, but continues through its interpretation, subject-based justification and didactic transformation into a teaching situation (Celik et al., 2026; Wen & Wen, 2024).

3. Research Design and Methodology

The study is designed as a practice-oriented multiple-case study. It presents three teaching interventions implemented in university-based teacher preparation. Each intervention comes from different subject didactics and was designed by a different university teacher. The aim is to identify common principles that enable the meaningful use of AI in teacher education.

The analysed cases are:

1. The use of AI when working with authentic language data in Czech language didactics.

2. The use of AI when designing and statistically processing empirical investigations in mathematics didactics.
3. The use of AI when identifying plants and animals in biology didactics.

Attention is paid to three questions:

1. What area of insufficient knowledge, skills or low self-efficacy does AI help students overcome?
2. How does AI support subject-specific thinking?
3. How is the AI output transformed into subsequent work with pupils?

4. Case Studies

4.1 Czech Language: AI, Corpus Data and Text-Centred Teaching

In language didactics, including Czech, one important challenge is how to work with authentic language data and transform them into material suitable for school teaching. The current communicative conception of Czech language teaching emphasises text-centredness and situationality: language phenomena should be explored in relation to texts and communicative situations, not only through isolated model sentences (Šebesta, 2005; Štěpáník, 2020).

Language corpora support this approach because they provide authentic data, frequency counts, collocations, and real contexts of use (Boulton & Cobb, 2017; Curry & McEnery, 2025). This is usually discussed as data-driven learning (DDL), in which learners investigate language data themselves (Lusta et al., 2023). In the Czech context, the educational use of corpora has been addressed, for example, by Šmejkalová and Kopřivová (2019) and Lukešová and Lukeš (2021).

For student teachers, however, corpus work is demanding. The Czech National Corpus (Český národní korpus, n.d.) offers extensive language material, but effective use requires the ability to formulate queries, navigate the corpus interface and interpret results. AI can therefore serve as a mediator between a naturally formulated teaching intention and the formal language of a corpus query. Its outputs must still be checked, refined and didactically interpreted.

The intervention included several AI-supported activities with language data, authentic or AI-generated communicative texts, and syntactic reasoning. Three of them are presented here to show how AI can support lesson preparation without turning student teachers into corpus linguists. The common aim was to help students use language data and AI outputs critically when preparing Czech language lessons.

The first activity focused on Czech spelling, specifically on the so-called “listed words” (*vyjmenovaná slova*). These are traditional sets of Czech words in which pupils memorise the use of the letter *y* after certain consonants, although contemporary pronunciation no longer distinguishes it from *i*. Student teachers used AI to translate a natural-language teaching question into a workable corpus search strategy. They first asked AI how to search for the occurrences and contexts of selected words in the Czech National Corpus, then checked the suggested procedure or query in the corpus interface and interpreted the results didactically.

For example, they examined the pair *dmýchat* and *rozdmýchat*. Both words are relatively archaic: *dmýchat* means “to blow air into a fire”, while *rozdmýchat* means “to fan into flame” and is now more often used figuratively, for example, to “stir up” a conflict or emotion. The aim was not merely to count occurrences, but to decide whether such words are suitable for younger pupils, in which contexts they appear, and whether they are meaningful for teaching. Corpus data thus support professional decision-making by showing which words are common in contemporary Czech, which are peripheral, and which may be didactically difficult or poorly motivating.

The second activity examined the possibilities and limits of generative AI in creating short communicative texts for school use. With frequent words, AI can produce plausible texts appropriate to pupils’ age; with infrequent or archaic words, it often creates unnatural contexts, historicising situations or linguistically problematic formulations. This failure is didactically useful because it shows student teachers that AI outputs must be evaluated against authentic language use, pupils’ age and teaching aims.

The third activity focused on syntactic competence and metalinguistic reasoning. In work with attributes, AI or an image generator can visualise the difference between a text without attributes and one supplemented with them. Pupils compare the texts and visual outputs, identify the attributes and discuss how they refine meaning.

Other examples involve false syntactic pairings or attachment ambiguity, such as *Filé pro děti bez kostí* ("fillet for children without bones") or *Šampon na vlasy s chmelovým extraktem* ("shampoo for hair with hop extract"). Such cases show that AI may misinterpret Czech syntactic structures, often under the influence of English or general syntactic categories. The learner therefore becomes an active corrector of the model: they identify the inaccuracy, reformulate the prompt and explain why the construction sounds comic or ambiguous.

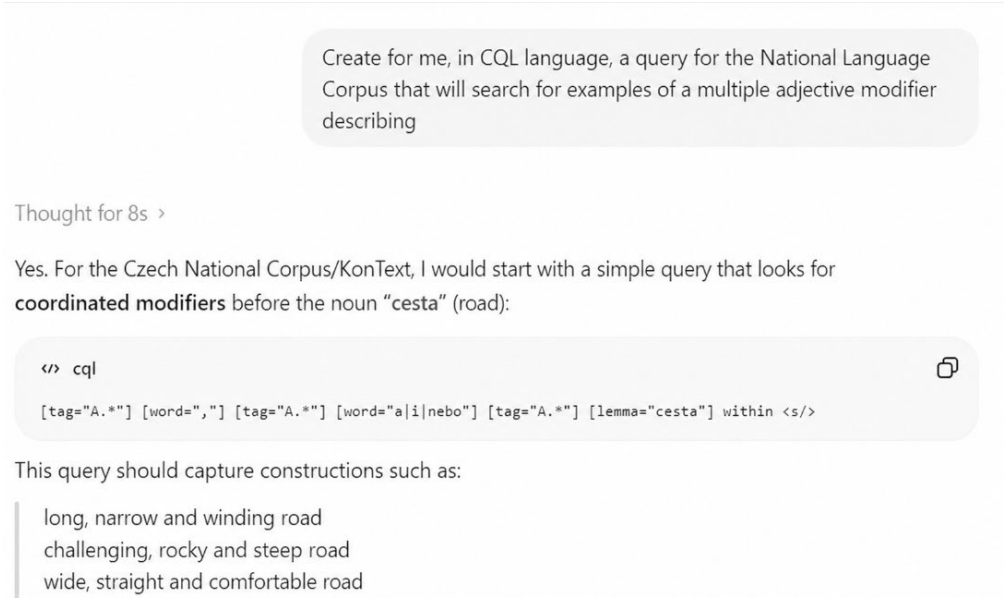


Figure 1: Example of generative AI being used to create a CQL query for the Czech National Corpus / KonText (English translation of the original Czech screenshot).

For advanced students or exceptionally gifted pupils, AI can also support the formulation of corpus queries for different types of attributes. A natural-language request is transformed into CQL queries that must be refined: overly broad queries yield unmanageable results, while syntactically incorrect queries return irrelevant examples. Students must therefore use their own syntactic knowledge, evaluate the retrieved occurrences and decide which examples are didactically useful.

Overall, the Czech language case shows that AI can support work with language data, metalinguistic thinking and the didactic transformation of subject content. Its value lies not in producing ready-made teaching materials, but in helping student teachers enter demanding work with language data while requiring expert checking, interpretation and pedagogical decision-making.

4.2 Mathematics: AI and Statistical Reasoning in Final Theses

In mathematics, the intervention focused on statistical data analysis in student teachers' final theses. Although students complete a theoretically oriented Statistics course, they often use statistical tools only rarely and inefficiently in their own empirical investigations. An analysis of final theses showed that students frequently avoid quantitative analysis, often citing small research samples, even when simple quantitative analysis would be possible.

This problem is not only a matter of statistical knowledge. It is also linked to statistical self-efficacy and statistics anxiety: students may doubt their ability to choose, interpret, and defend statistical procedures correctly (Finney & Schraw, 2003; Perepiczka et al., 2011). For this reason, AI-supported activities were included in the Final Thesis Seminar. They focused not only on evaluating collected data, but also on designing the research investigation itself. Students learned to ask, for example, what data they need, which variables will be observed, whether the expected result can be evaluated quantitatively, what a realistic sample size is, and what conclusions can legitimately be drawn.

In this activity, AI acts as a consulting partner. It can help refine a research question, propose the structure of a questionnaire, distinguish between data types, recommend possible methods of analysis, and formulate results in an appropriate professional style. Recent research suggests that tools such as ChatGPT can support quantitative data analysis, generate code for statistical software and help interpret results, although their outputs require expert checking (Prandner et al., 2025; Xing, 2024).

The student must therefore verify whether the proposed procedure is appropriate to the nature of the data, the sample size and the aims of the thesis. The main benefit is not that AI identifies a statistical test, but that it lowers the entry barrier and helps students understand the purpose of statistical processing and its connection to research design. Teaching thus moves from mechanical test selection to the development of statistical reasoning.

This experience is also relevant for future teaching practice. Student teachers learn to work with data as a means of reflecting on teaching, for example, when evaluating pupils' results, assessing the effectiveness of teaching methods or planning pedagogical interventions. AI does not take over analytical responsibility; rather, it enables students to enter an area they might otherwise avoid.

4.3 Biology: AI, Species Recognition and Field-Oriented Learning

The case study in biology didactics focuses on the use of applications for recognising plants and animals based on visual and audio inputs. In the preparation of future biology teachers, correct species identification is usually associated with uncertainty. Students and beginning teachers may worry that, in the field, they will not be able to identify all plants, animals or their vocalisations correctly. This uncertainty may lead them to avoid field teaching, reduce it to familiar and safe examples, or shift attention away from direct observation of nature toward work with a textbook or pre-prepared materials.

In this area, AI tools for image and sound recognition can lower the entry barrier. In the preparation of future teachers, the main applications used were Merlin Bird ID, Seek by iNaturalist and PlantNet/Pl@ntNet. Merlin Bird ID supports identification from photographs or sound recordings; Seek by iNaturalist enables identification of various groups of organisms; and PlantNet/Pl@ntNet is primarily used for identifying plants from image data. These tools enable rapid species identification and allow students or teachers to name species correctly.

The educational value of these applications, however, does not consist only in the fact that they provide the name of an organism. More importantly, they make it possible to shift attention from encyclopaedic identification to the observation of features, comparison of similarities and differences, and formulation of arguments. AI functions here as a supportive means, not as an authoritative identification key. The application's suggestion becomes a prompt for further verification: students and pupils examine which features of the organism the application probably evaluated, which features are actually relevant and how the result can be confirmed or challenged using additional sources.

Teaching of future teachers, therefore, includes an activity in which students work with several similar plants, photographs or sound recordings. First, they themselves describe the observed features and formulate a hypothesis about whether they are dealing with the same species or different species. Only then do they use the application to suggest an identification. They then compare the AI output with their own observation, identification literature or other expert sources. The aim of the activity is not only to arrive at the correct species name, but to learn to recognise similar and different features, work with uncertainty and justify one's own conclusion.

This activity directly overlaps with future teaching practice. The future teacher learns not only to operate an application, but above all to organise a learning situation in which AI supports pupils' inquiry-based and observational activities. Pupils may, for example, compare two similar plants, observe differences in leaf shape, flowers, or habitat, or work with bird sound recordings to identify recurring acoustic features. The application provides a suggestion, but teaching is organised so that pupils must interpret, verify, and justify the result.

Situations in which the application offers an uncertain or incorrect result have special didactic value, for example, when it identifies an animal species that does not occur in the Czech Republic (Figure 2). Such an error is not merely a technical deficiency, but can become a learning opportunity. The teacher can discuss with pupils why the incorrect identification may have occurred: whether the sound recording was of poor quality, whether an important diagnostic feature was missing, whether similar species were involved, or whether the application was working with limited probability, possibly even suggesting a first occurrence in the Czech Republic. Precisely such situations lead pupils to engage in critical thinking, verify results in other sources, develop scientific literacy, work with species distribution ranges, and understand that biological identification is not a mere assignment of a name but an argumentative process. Student teachers thus learn that the AI output is not a final answer, but part of the process of species identification.

For future teachers, this experience is also significant for self-efficacy. AI helps them overcome the fear that they will not know the names of all organisms in the field. At the same time, however, it leads them to a more general

professional competence: guiding pupils toward observation, classification, comparison and argumentation. Species-level identification is therefore not a self-serving goal, but a means of developing biological thinking. The teacher does not have to act as an infallible expert on every organism found around the school; on the contrary, they can model a professionally responsible procedure in which hypotheses are verified, results are compared, and uncertainty becomes part of learning.

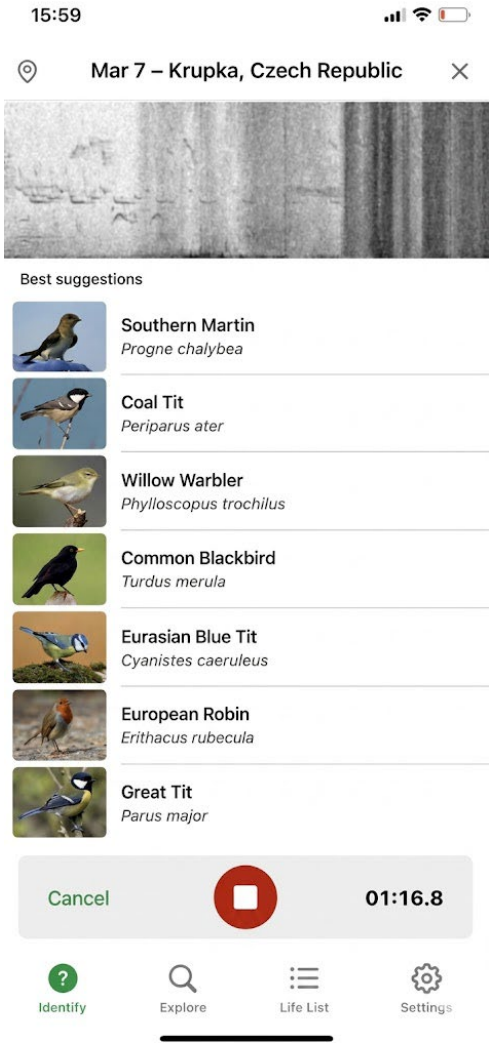


Figure 2: AI misidentification example: a South American bird species, the southern martin, suggested for a recording from the Czech Republic.

The contribution of AI in this activity does not lie in the automation of expert decision-making, but in lowering the entry barrier to an activity that may be associated with uncertainty for future teachers. Applications for organism recognition are therefore not used merely to name species, but as support for teaching based on observation, comparison, verification and justification.

5. Discussion

The three examples differ in subject area, tools used and teaching situation, but they share a common structure. In each case, AI enters an area that is professionally demanding for student teachers and may be associated with uncertainty or lower self-efficacy. In Czech language education, this concerns work with authentic language data and formal corpus queries; in mathematics, the design of empirical investigations and statistical data analysis; and in biology, the identification of organisms and work with natural material.

Across the cases, AI lowers the entry barrier to activities that students might otherwise avoid. It offers initial scaffolding, a working draft, an interpretative guide or the possibility of iteratively refining a procedure. Its educational value, however, does not lie in the output itself, but in what students do with it: checking, refining,

interpreting and transforming it for teaching. Students therefore learn not only to use AI, but also to question it, correct it and decide when expert knowledge, additional sources or teacher judgement are needed.

From the perspective of the TPACK framework, all three cases show that AI first supports work with subject-specific content, but becomes pedagogically meaningful only when this content is transformed for work with pupils. In Czech language education, corpus data and AI-supported queries are used to prepare authentic or didactically adapted language material. In mathematics, AI supports understanding the relationship between a research question, data, an analytical method, and the interpretation of results. In biology, AI-based identification tools support observation, comparison, verification and biological argumentation rather than merely naming species.

Several design principles for integrating AI into teacher education can be derived from these cases:

- The AI activity should be anchored in a specific subject didactics.
- The task should target an area in which students experience uncertainty or lower self-efficacy.
- AI should serve as support for entering a demanding activity, not as a substitute for expert decision-making.
- The AI output should be subject to critical checking, interpretation and possible revision.
- The activity should have a clear overlap with future work with pupils.
- The activity should include reflection on the limits of AI and responsible use of the tool.

These principles suggest that meaningful AI integration in teacher education is not isolated technology training. Its value lies in connecting technological support, subject content and pedagogical decision-making. AI can become a means of professional learning when embedded in carefully designed activities that foster subject understanding, critical reflection and didactic transformation.

6. Conclusion

The case studies show that AI can serve as didactic support in pre-service teacher education, especially in areas that are professionally demanding and often associated with low self-efficacy. Its main benefit is not the automation of teachers' work or the replacement of expert judgement, but the opportunity for students to engage in complex activities safely and learn to evaluate, interpret, and use AI outputs didactically.

Meaningful use of AI, therefore, requires a clear pedagogical intention, subject anchoring, critical work with outputs and a connection to future classroom practice. Under these conditions, AI can support not only technological skills, but also subject reasoning, pedagogical judgement and the professional confidence of future teachers.

Acknowledgements

This paper was supported by the project TAČR N. TQ23000334 – AI-based Adaptive Learning Platform.

Ethics Declaration

The research is based on an analysis of learning activities implemented as part of regular university teaching. The text does not include students' personal data or individually identifiable statements.

AI Declaration

During the preparation of the working version of the text, ChatGPT was used as support for structuring the argumentation, formulating the language, and compressing the source materials' style. All content was subsequently checked by the authors, who are responsible for the factual accuracy, data interpretation, and the final form of the article.

References

Bae, H., Hur, J., Park, J., Choi, G. W. and Moon, J. (2024). Pre-Service Teachers' Dual Perspectives on Generative AI: Benefits, Challenges, and Integration into their Teaching and Learning. *Online Learning*, 28(3), 131–156. http://files.eric.ed.gov/fulltext/EJ1446841.pdf.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.

Blonder, R., Feldman-Maggor, Y. and Rap, S. (2025). Are they ready to teach? Generative AI as a means to uncover pre-service science teachers' PCK and enhance their preparation program. *Journal of Science Education and Technology*, 34(6), 1301–1310. https://doi.org/10.1007/s10956-024-10180-2.

Boulton, A. and Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. https://doi.org/10.1111/lang.12224.

Cai, H., Han, B., Sun, J., Li, X. and Wong, L. H. (2025). Harnessing AI for teacher education to promote inclusive education: Investigating the effects of ChatGPT-supported lesson plan critiques on the development of pre-service teachers' lesson planning skills. *The Internet and Higher Education*, 67, 101022. https://doi.org/10.1016/j.iheduc.2025.101022.

Celik, I., Kontkanen, S., Laru, J. and Dalyanci, A. A. (2026). Co-constructing adaptive lesson plans with GenAI: Pre-service teachers' Intelligent-TPACK and prompt engineering strategies. *Computers & Education*, 105485. https://doi.org/10.1016/j.compedu.2025.105485.

Český národní korpus. (n.d.). *KonText – corpus interface of the Czech National Corpus*. https://www.korpus.cz/kontext/.

Curry, N. and McEnery, T. (2025). Corpus linguistics for language teaching and learning: A research agenda. *Language Teaching*, 1–20. https://doi.org/10.1017/S0261444824000430.

Finney, S. J. and Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28(2), 161–186. https://doi.org/10.1016/S0361-476X(02)00015-2.

Lukešová, L. and Lukeš, D. (2021). Frekvence vyjmenovaných slov v současné češtině: korpusová studie. *Didaktické studie*, 13(1), 125–156.

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://dl.acm.org/doi/abs/10.1145/3313831.3376727.

Lusta, A., Demirel, Ö. and Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9(12). https://doi.org/10.1016/j.heliyon.2023.e22731.

Mishra, P. and Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x.

Oliveira, L., Tavares, C., Strzelecki, A. and Silva, M. (2025). Prompting minds: Evaluating how students perceive generative AI's critical thinking dispositions. *Electronic Journal of E-learning*, 23(1), 1–18. https://eric.ed.gov/?id=EJ1471471.

Perepiczka, M., Chandler, N. and Becerra, M. (2011). Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. *Professional Counselor*, 1(2), 99–108. https://eric.ed.gov/?id=EJ1063046.

Prandner, D., Wetzelhütter, D. and Hese, S. (2025, January). ChatGPT as a data analyst: an exploratory study on AI-supported quantitative data analysis in empirical research. *Frontiers in Education* (Vol. 9, p. 1417900). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1417900.

Šebesta, K. (2005). *Od jazyka ke komunikaci: didaktika českého jazyka a komunikační výchova*. Univerzita Karlova v Praze, Nakladatelství Karolinum.

Šmejkalová, M. and Kopřivová, M. (2019). *Jazykové korpusy a elektronické zdroje ve výuce českého jazyka*. Univerzita Karlova, Pedagogická fakulta, Prague.

Štěpáník, S. (2020). Konstruktivistické a kognitivně-komunikační paradigma jako východisko koncepce výuky českého jazyka. *Pedagogika*, 70(1), 5–28.

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. https://doi.org/10.3102/0013189X015002004.

Tschannen-Moran, M., Hoy, A. W. and Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202–248. https://doi.org/10.3102/00346543068002.

Xiaoyu, W., Zainuddin, Z. and Hai Leng, C. (2025). Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024). *Cogent Education*, 12(1), 2485499. https://doi.org/10.1080/2331186x.2025.2485499.

Xing, Y. (2024). Exploring the use of ChatGPT in learning and instructing statistics and data analytics. *Teaching Statistics*, 46(2), 95–104. https://doi.org/10.1111/test.12367.

Yadav, A., Lachney, M., Hu, A. and Tavernier, L. (2026). Integrating Generative AI in teacher education: Pre-service teachers' perspectives, attitudes, and design challenges. *Journal of Research in Childhood Education*, 40(1), 24–45. https://doi.org/10.1080/02568543.2025.2581712.

Wen, W. and Wen, H. (2024). Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. *Contemporary Issues in Technology and Teacher Education*, 24(4), 582–611. https://www.learntechlib.org/primary/p/223935/.

Pedagogically Asymmetric Elo Models for Multidimensional Adaptive Learning Systems

David Janda, Antonín Jančařík, Jakub Michal and Tomáš Jeřábek
Charles University, Faculty of Education, Prague, Czech Republic

david.janda@pedf.cuni.cz
antonin.jancarik@pedf.cuni.cz
jakub.michal@pedf.cuni.cz
tomas.jerabek@pedf.cuni.cz

Abstract: Adaptive learning systems commonly use psychometric models to estimate learner ability and task difficulty. Among the most widely used approaches are Item Response Theory (Harvey & Hammer, 1999) and modifications of the Elo rating system (Pelánek, 2016). These models typically assume a symmetric learner–task interaction, with maximum diagnostic information occurring at approximately a 50% probability of success. While appropriate for testing, this equilibrium may not correspond to pedagogically desirable conditions for sustained learning. This article presents a pedagogically asymmetric parameterisation of a multidimensional Elo model for adaptive mathematics education. The model combines a multidimensional representation of mathematical competencies based on the Global Proficiency Framework for Mathematics (UNESCO et al., 2020), scaffold-aware partial-credit scoring, and separate equilibrium parameters for learner and task updates. This enables learner-profile development and task-difficulty dynamics to be controlled differently rather than being governed by the same symmetric equilibrium. The model was implemented in the Edu-AI system for preparing Czech elementary school learners for standardised mathematics entrance examinations (Jančařík et al., 2023). The pilot dataset comprised 63,020 learner–task interactions involving 628 learners and 233 tasks. A replay-based simulation preserved the chronological interaction histories, achieved scores, and scaffold use while recalculating learner and task parameters under iteratively adjusted equilibrium settings. The analysis focused on model stability, learner-profile dynamics, task-difficulty convergence, and the interaction between equilibrium settings and scaffold-aware scoring. The replay results showed that relatively small changes in equilibrium parameters substantially affect learner and task dynamics. A configuration with an equilibrium of 0.80 for learner updates and 0.85 for task updates produced stable and pedagogically interpretable behaviour under the scoring policy used in the pilot. These values are treated as promising initial settings rather than as optimal parameters for learning effectiveness. The study therefore demonstrates the feasibility of introducing pedagogical asymmetry into a multidimensional Elo model and provides a basis for prospective testing of the model's educational effects.

Keywords: Adaptive Learning, Elo Rating System, Multidimensional Learner Modelling, Mathematics Education, Educational AI, Learning Analytics

1. Introduction

Adaptive learning systems represent a significant area of current research in educational technology. Many current systems utilise psychometric models based on Item Response Theory (IRT) (Harvey & Hammer, 1999; Cai et al., 2016; Chen et al., 2025), modifications of the Elo rating system (Pelánek, 2016; Vermeiren et al., 2025)—originally developed for evaluating chess players—or combinations thereof (Antal, 2013). These approaches enable the continuous modelling of a student's abilities and the dynamic adjustment of task difficulty.

However, most adaptive systems are primarily focused on the accuracy of measuring student performance. Optimal diagnostic value occurs when the probability of correctly solving a task is approximately 50%, which, among other things, determines the selection of appropriate subsequent tasks for a specific student. This principle corresponds to maximising the informational value of the learner-task interaction, but not necessarily to optimising the conditions for the student's learning, which is generally not considered in testing.

When the primary goal of the learner-task interaction is to support the student's learning through problem solving, the situation is different. From a pedagogical perspective, it can be assumed that effective learning occurs more in situations of manageable challenge, where the student solves tasks with a relatively high probability of success while remaining exposed to appropriate cognitive demands (de Bruin et al., 2023). From this perspective, it is important for an adaptive system to keep tasks within the realm of manageable challenges, rather than prolonged failure.

This paper builds upon the previous development of an AI assistant for mathematics education presented at this conference (Jančařík et al., 2022). The original system focused primarily on chatbot-supported student problem-

solving. It used pre-prepared sequences of problems and supporting materials, along with simple adaptive mechanisms based on difficulty selection and navigation between thematic units.

The current version of the system represents a further step toward a fully adaptive multidimensional learner model that utilises a pedagogically modified Elo rating system, scaffold-aware scoring, and a curriculum-oriented representation of mathematical competencies.

The proposed model shifts the emphasis from maximising diagnostic accuracy toward supporting pedagogically appropriate learning conditions. The main objective of this article is to present a pedagogically asymmetric parameterisation of the Elo system and examine its behaviour using empirical learner–task interaction data and replay-based simulation. The study addresses two research questions:

RQ1: Can a pedagogically asymmetric Elo model, using separate equilibrium parameters for learner and task updates, produce dynamics consistent with the intended pedagogical behaviour?

RQ2: Which learner and task equilibrium parameter settings appear to provide a stable and pedagogically interpretable balance between learner-profile development and task-difficulty dynamics under scaffold-aware scoring? The article presents four key innovations of the proposed model:

1. pedagogically asymmetric parameterisation of the Elo system,
2. multidimensional representation of mathematical competencies,
3. scaffold-aware scoring that reflects the use of support mechanisms,
4. a stabilised mechanism for updating task difficulty.

2. Theoretical Background

2.1 Elo Models in Education

The Elo rating system was originally developed to evaluate the performance of chess players (Elo, 1978). It began to be used in educational settings primarily because of its simplicity, continuous parameter updates, and low computational requirements. In educational applications, the Elo model is typically interpreted similarly to IRT models: a student's ability and the difficulty of a task are represented on a single scale, and the probability of success is derived from the difference between these values (Pelánek, 2016).

Compared to classical IRT models, the Elo model allows for simpler continuous parameter updates and better supports dynamic learner modelling in real time (Pelánek, 2016; Abdi et al., 2019). This is particularly important in adaptive learning environments, where the system continuously responds to the student's current performance.

Traditional Elo models implicitly assume a symmetrical system. If the difficulty of a task matches the student's level, the expected probability of success is approximately 50%. This principle is suitable for rating subjects that interact directly with one another (e.g., in performance evaluation, such as chess) and for diagnostic and testing environments where, through interaction with tasks, direct comparisons among students are possible. However, the target expected score of 50% may not align with the pedagogical requirements for effective learning.

2.2 Adaptive Learning and Desirable Difficulty

Current research in adaptive learning emphasises the importance of personalisation, scaffolding, and long-term learner support. Adaptive systems no longer serve merely to diagnose performance but are increasingly focused on supporting learning, motivation, and the long-term development of competencies (VanLehn, 2011; Walkington & Bernacki, 2018).

Educational theories of desirable difficulty assume that effective learning through problem-solving occurs in situations involving manageable challenges. A task should be challenging enough to support the development of competencies, yet it must not lead to the learner's long-term failure (Bjork & Bjork, 2020; de Bruin et al., 2023).

From a learning perspective, the explainability and interpretability of learner models also play a significant role in the interaction between the student and the task (Bull & Kay, 2016; Jiang, 2025). It is important for both teachers and students that the system not only provides a final score but also enables students to reflect on the outcome of their interaction with the task and interpret their strengths and weaknesses in specific areas.

Alternative approaches to adaptive learner modelling pursue related but distinct objectives. Knowledge-tracing models use previous learner interactions to estimate evolving knowledge states and predict future performance (Liu et al., 2025). Adaptive task-difficulty models may instead regulate task difficulty around a predefined target performance; for example, Kassenkhan et al. (2026) use a target mastery rate of 0.75. The present approach differs in that it introduces separate equilibrium parameters for learner and task updates, allowing their dynamics to be adjusted asymmetrically.

2.3 Previous Development of the System

The proposed model builds on the AI Assistant for Students and Teachers project, which focuses on supporting mathematics and the Czech language through AI technologies. The original version of the system combined a chatbot interface, thematic exercise sequences, instructional videos, and simple adaptive mechanisms. The system was implemented as both a web-based application and a Telegram app.

The pilot implementation revealed a need for more personalised learning, more detailed learner modelling, and more precise task adaptation based on the student's specific mathematical competencies. These requirements led to the design of the multidimensional Elo model presented in this article. At the same time, gamification elements are being integrated into the system, and a chat system is being developed to facilitate direct discussions about the task being solved.

3. System Design

3.1 Introduction to the Edu-AI System

The Edu-AI system (Figure 1) is an adaptive learning environment designed to prepare Czech elementary school students for the standardised mathematics entrance exams. The system combines pre-designed learning pathways with a continuously updated learner model based on a multidimensional Elo vector.

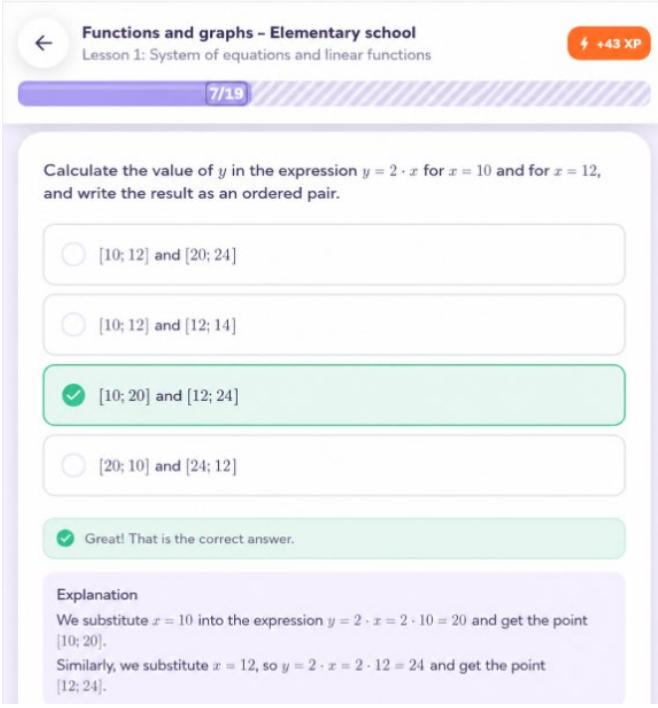


Figure 1: System environment

The pilot implementation of the system was designed as a guided, adaptive environment that combines curriculum-structured instruction and personalised practice. All students first take an initial diagnostic test to establish the baseline learner profile. The result of this test is an initial multidimensional Elo vector representing the student's current level of mathematical competence across the components of the GPF framework. Initial task parameters were estimated using AI-assisted matching to analogous examples and competency descriptions in the Global Proficiency Framework and subsequently reviewed by members of the research team with expertise in mathematics education.

Following an initial assessment, students proceed through eight thematically organised lessons focused on areas of mathematics aligned with the requirements of the standardised entrance exams in the Czech Republic. At this stage, instruction is guided by pre-prepared content that combines:

- instructional materials,
- illustrative examples,
- scaffolded tasks,
- practice activities.

The aim of this section is not only to assess the student's performance but also to gradually modify the learner profile through interactions with tasks of varying relevance and difficulty. Each completed task updates the individual components of both the task's and the student's Elo vector based on:

- correctness of the solution,
- use of hints,
- relevance of the task to specific competencies,
- estimated difficulty of the task.

After completing the guided portion of the lesson, the system proceeds to a short practice session related to the lesson and, if necessary, to an adaptive practice phase based on a database of exercises.

In the current version, all tasks are multiple-choice and include the question, answer options (with feedback for incorrect answers), corresponding scores, complete solutions, two levels of hints (a brief hint and a more detailed hint), and the metadata required for implementation in the application. The selection of subsequent tasks for any additional practice beyond the scope of individual lessons is determined dynamically by the current distance between the learner's profile and the task profile.

At the same time, the entire system strives to maintain a pedagogically defined balance between student success and the cognitive difficulty of the tasks. The resulting adaptive learning path combines guided instruction, personalised practice, scaffolded support, and real-time learner-model updates.

3.2 Multidimensional Representation of Competencies

The system uses a multidimensional model based on the Global Proficiency Framework for Mathematics (GPF). The pilot implementation uses a 35-dimensional representation of mathematical competencies corresponding to the individual subconstructs of the GPF. An Elo vector, therefore, represents each student's profile:

$$L = (l_1, l_2, \dots, l_{35})$$

In accordance with this approach, each task is described by a 35-dimensional difficulty vector D, which expresses the difficulty of the task in the context of individual competencies, consisting of 35 components with values on a scale of 1 to 10, as well as a 35-dimensional relevance vector expressing the relevance of the task to individual competencies with values ranging from 0 to 2 (0 = no relationship, 1 = marginal relevance, 2 = significant relevance).

Task classification was performed using AI-assisted classification followed by human validation.

3.3 Pedagogically Asymmetric Parameterisation of the Elo Model

3.3.1 The Classic Elo Model

As a starting point, we used the basic Elo rating formula to calculate each category's Elo rating:

$$E = \frac{1}{1+e^{-\lambda(L-D)}}.$$

It works in such a way that when $L = D$ the value of E is 0.5. Thus, the student is expected to have a success rate of approximately 50%. The parameter λ influences the resulting value of E when $L \neq D$ and determines the impact of a specific interaction on the subsequent classification of the task.

The new value of L' (or D') is then determined using the following equation:

$$L' = L + K_n \cdot (S - E),$$

$$D' = D - K_m \cdot (S - E),$$

where S is the actual score, E is the expected score, and K_n is a function that limits the magnitude of the change depending on the number of student interactions n (or task m), defined as follows:

$$K_n = \frac{K_0}{1 + \alpha n}.$$

Given the 0–10 scale used to describe the student's profile, $K_o = 1$ and $\alpha = 0.1$ were selected.

3.3.2 Scaffold-Aware Scoring

The current version of the app contains only multiple-choice questions, intended to demonstrate and explore the effects of the scaffold-aware scoring system, one of the pillars of the pedagogically oriented Elo model. Unlike binary scoring or a three-point chess rating system, this scoring system considers the use of hints:

- – correct solution without assistance,
- 0.75 – correct solution with basic hints,
- 0.5 – correct solution with additional hints,
- 0 – incorrect solution.

The original model design assumed a pedagogical balance approaching 0.9, which corresponds to the intuitive notion of high student performance on multiple-choice tasks in adaptive learning.

In a scaffold-aware scoring environment, this configuration led to an excessive decline in learner profiles. The reason was that a significant portion of interactions did not yield a binary score of 1.0 but rather values of 0.75 or 0.5, reflecting the use of hints. The effective distribution of learner scores thus shifted toward lower values, suggesting the model needed adjustment.

3.4 Modifications to the System for Replay-Based Simulation

To perform replay-based simulations, a system modification was designed to reflect the scaffold-aware scoring described above while also meeting the required pedagogical asymmetry (from the student's perspective, it is desirable to present challenges with a theoretically lower expected score than from the task's perspective). To adjust the student's Elo rating, we therefore use:

$$E_L = \frac{1}{1 + e^{-(\lambda(L-D)+b_L)}}.$$

Similarly, to calculate the Elo rating of a task, we use:

$$E_D = \frac{1}{1 + e^{-(\lambda(L-D)+b_D)}}.$$

The purpose of the replay-based simulation is to test appropriate settings for the b_L and b_D parameters to meet the requirements described above when applied to the obtained dataset.

3.5 Testing Process and Testing Dataset

An introductory diagnostic test was designed for the app, followed by a total of eight courses covering basic algebraic expressions, ratios and proportionality, linear functions, equations and inequalities, measurements in geometry, the basics of statistics, descriptive geometry, and fractions. The courses were released sequentially in the order listed, one week apart. All participating teachers were informed of each course's release and sent an access code for their students.

Testing of the application took place during March and April 2026. Students were contacted through their teachers. The teachers were tasked with administering an initial diagnostic test to the students and regularly engaging them in active use of the application at least once a week for at least 15 minutes.

After basic data pruning, anonymous access and inactive users with insufficient recorded interaction data were excluded. The resulting dataset included 63,020 interactions among 628 students across 43 classes and 233 tasks. During actual testing, the model's results were not made available to the students at this stage to avoid any potential undue influence on the testing process.

4. Results of the Replay-Based Simulation

The replay-based simulation preserved the original chronological sequence of learner–task interactions, including achieved scores and scaffold use, while recalculating learner and task Elo parameters under iteratively adjusted equilibrium settings. Each replay informed the subsequent parameter adjustment. The aim was not exhaustive optimisation, but identification of a stable and pedagogically interpretable configuration. The goal of the simulation was to:

- analyse the stability of learner and task profiles,
- monitor the impact of scaffold-aware scoring,
- analyse learner growth,
- stabilise task dynamics.

Configurations were evaluated primarily for the absence of divergent behaviour, excessive short-term fluctuations, and unstable task-difficulty dynamics.

4.1 Stable Task Dynamics

The replay-based simulation showed that more stable task dynamics were achieved at target equilibrium values lower than those originally proposed, specifically:

- $target_student_at_equal$ at value 0.8 ($L = D$ theoretically occurs from the student's perspective at a score of 0.8; the student's Elo rating remains unchanged at this value, increases at a score of 1, and decreases at a score < 0.8 according to the parameters of the logistic function,
- $target_task_at_equal$ at value 0.85 ($L = D$ theoretically occurs from the task's perspective at a score of 0.85, when the task's Elo rating remains unchanged; at a score of 1, it decreases slightly, and at a score < 0.85, it increases according to the parameters of the logistic function).

The parameters of the logistics function correspond to these target values:

$$b_L = \ln\frac{0.8}{0.2} \approx 1.386,$$

and

$$b_D = \ln\frac{0.85}{0.15} \approx 1.735.$$

4.2 Relationship Between the Target Equilibrium and the Scoring Policy System

A key finding of the replay-based simulation was the strong link between the target equilibrium parameters and the system's scoring policy. The results also show that the design of pedagogically asymmetric Elo systems cannot be separated from the scoring policy and the method of interpreting the scaffolded solution.

The classical Elo model assumes an equilibrium state at an expected success rate of 50%. This setting is suitable for diagnostic purposes because it maximises the informational value of individual interactions. From a pedagogical perspective, however, it represents a situation in which the learner fails in approximately half of cases, which may not be suitable for long-term learning or maintaining motivation.

In the proposed model, therefore, the learner equilibrium is shifted to 0.80. Given the same learner level and task, the system assumes an approximately 80% probability of a successful solution. The goal of this setting is to create an environment where the learner predominantly experiences success while remaining sufficiently cognitively challenged.

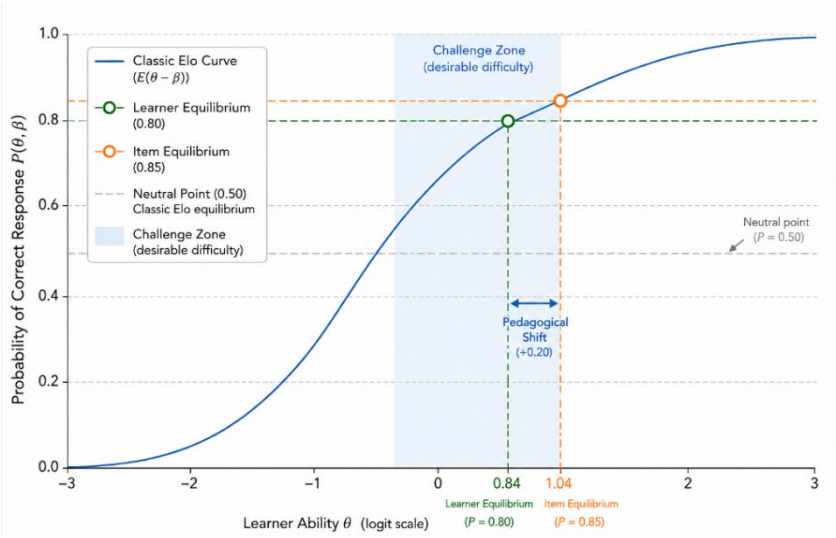


Figure 2: A Shift of Pedagogical Equilibrium

The balance for updating task difficulty is set even more conservatively. Based on simulations, the task equilibrium was set to 0.85. Task difficulty, therefore, stabilises only when students' long-term success rate reaches approximately 85%. This setting limits overly rapid increases in task difficulty while simultaneously supporting the maintenance of the so-called "challenge zone"—an area of productive challenge where tasks are sufficiently demanding to foster skill development but do not lead to systematic failure.

Replay-based simulations have shown that even small changes to these parameters significantly influence the system's long-term dynamics. Shifting the learner equilibrium increases the likelihood of growth in learner profiles, while a higher task equilibrium contributes to the stability of task difficulty. Experimentally, the most balanced configuration was achieved with the settings $target_student_at_equal = 0.80$ and $target_task_at_equal = 0.85$ (Figure 2).

Replay-based simulations revealed a strong interdependence between learner and task dynamics. Even relatively small changes in pedagogical balance parameters significantly influenced the distribution of increases and decreases in learner profiles (Figure 3).

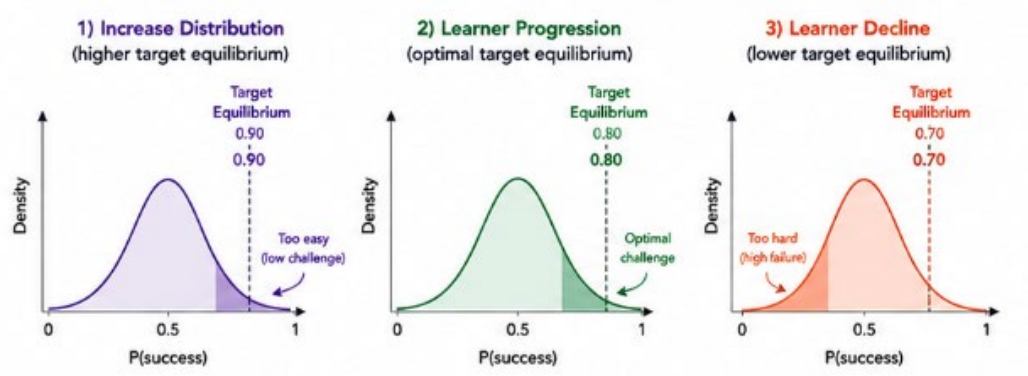


Figure 3: The effect of target equilibrium parameters on learner growth

The most stable configuration was achieved experimentally with:

- $target_student_at_equal = 0.8$,
- $target_task_at_equal = 0.85$.

4.3 Task Convergence Behaviour

To illustrate the model's behaviour, the trajectories of the Elo parameters for individual tasks were analysed within specific components of the competence vector.

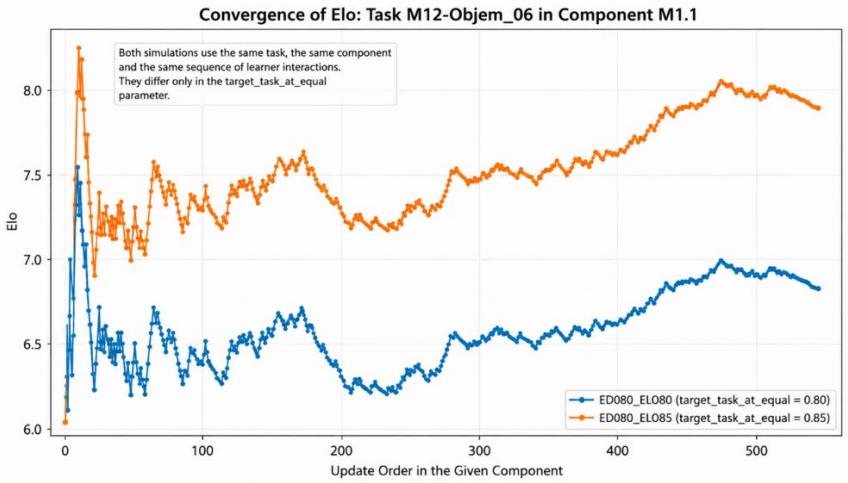


Figure 4: Convergence of task difficulty for task M12-Objem_06 in Component M1.1

Figure 4 shows the evolution of the difficulty parameter for task *M12-Objem_06* in component M1.1 across two model configurations. This task was selected as an illustrative example to demonstrate how the same interaction history produces different difficulty trajectories under different task-equilibrium settings; it was not selected as a statistically representative case.

Both simulations are based on the same task, the same component of the competence vector, and the same historical sequence of learner–task interactions. They differ only in the setting of the *target_task_at_equal* parameter. In both cases, the estimate shows high sensitivity in the initial phase, when only a small number of responses are available. As the number of interactions increases, the oscillations gradually dampen, and the difficulty parameter stabilises.

Importantly, however, the two configurations do not converge to the same value. The ED080_ELO80 configuration stabilises at a lower estimated difficulty, while the ED080_ELO85 configuration converges to a significantly higher value. This shows that the target equilibrium setting not only affects short-term fluctuations in the model but also systematically shapes the resulting estimate of the task's difficulty.

It is also significant that the system does not exhibit long-term uncontrolled drift in any configuration. After enough interactions, changes in difficulty gradually diminish, and the task parameter oscillates within a limited range around a stable value. At the same time, the result confirms that the choice of parameters must be evaluated not only in terms of task stability but also in relation to the dynamics of learner profiles, since task calibration and the growth of learner parameters are recursively linked in the model.

4.4 Learner Convergence Behaviour

In addition to the stability of task parameters, the dynamics of the learner model were also monitored. Figure 5 shows the development of Student 213's Elo parameter in component A3.3 across two model configurations.

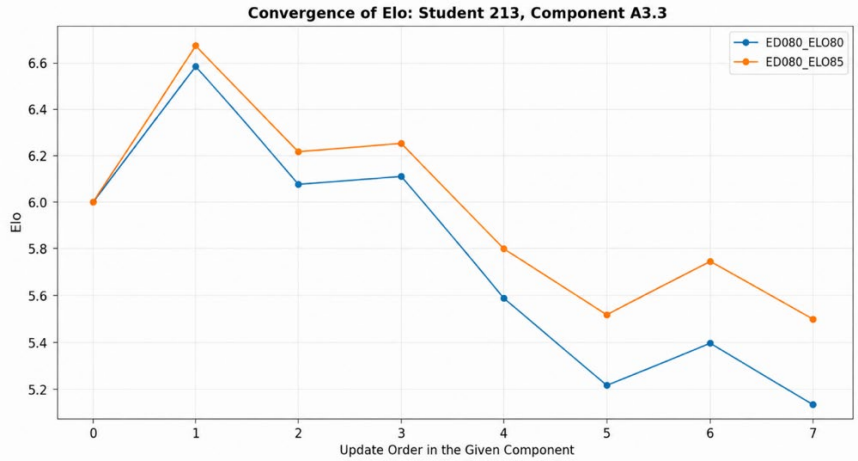


Figure 5: Convergence of learner parameter for student 213 in component A3.3

Student 213 was selected because this learner had a comparatively high number of updates in component A3.3, allowing the trajectory to be illustrated over multiple iterations. The figure is therefore intended as an illustration of model dynamics rather than as evidence based on a representative individual case.

Unlike task parameters, learner parameters exhibit greater short-term variability. This is to be expected, as the number of interactions within the individual components of the competence vector is significantly lower than the number used to estimate task difficulty.

The graph shows that the ED080_ELO85 configuration yields higher learner parameter values in the observed component than the ED080_ELO80 configuration. This difference suggests that the choice of target equilibrium parameters significantly influences not only task stability but also the learner profile's trajectory. In this case, the ED080_ELO85 configuration better aligns with the pedagogical expectation that an active and regularly working student should not be systematically shifted downward by the model.

5. Discussion

The proposed model represents a shift from psychometrically oriented adaptive systems toward pedagogically oriented adaptive learning. The fundamental difference lies in the redefinition of the adaptation goal. While traditional systems primarily optimise the accuracy of ability estimates, the proposed system is designed to support pedagogically appropriate learning conditions through asymmetric learner and task dynamics. A significant contribution is the multidimensional representation of knowledge, which enables more detailed modelling of mathematical competencies. The integration with the Global Proficiency Framework simultaneously increases the model's interpretability and enables diagnostic feedback to both teachers and students.

Replay-based simulations demonstrated the importance of stabilised updates to task difficulty values. Without limiting the update rate, faster growth of learner profiles led to unstable increases in task difficulty. At the same time, however, it turned out that the stability of task parameters alone is not a sufficient criterion for model quality. Configurations that yield more stable estimates of task difficulty may also result in undesirable declines in learner profiles.

A key finding was the strong interdependence between learner and task dynamics. It turned out that even small changes in target equilibrium parameters significantly influence the distribution of learner growth and learner decline. This result confirms the nonlinear nature of recursively defined adaptive systems and shows that parameter settings cannot be evaluated in isolation.

The system’s scoring policy plays a crucial role in this process. The original assumption pointed toward higher target success rates; however, the inclusion of hints and the use of support-adjusted scoring altered the effective distribution of achieved scores. A successful solution with a hint does not carry the same weight as an independent solution, so it was necessary to adjust the target values for learner and task equilibrium to 0.80 and 0.85, respectively.

The results also suggest that adjusting task difficulty should not be the system’s sole response to students’ repeated difficulties. If a particular task yields lower-than-expected success rates, this does not necessarily mean it is poorly calibrated. It may also mean that students need better preparation, more appropriate scaffolding, additional explanations, or a different sequence of preceding activities. Future development of the system should therefore focus on identifying pedagogically relevant interactions and designing supportive interventions that increase the likelihood of successfully mastering a given competency.

The pilot implementation also revealed several limitations of the system. The competency taxonomy is still being refined, and AI-assisted task classification may contain minor inconsistencies. Another limitation is the relatively low number of interactions in some components of the learner vector, which may increase the short-term variability of learner parameters. At the same time, the study did not directly measure learning gains, but focused primarily on the stability and pedagogical interpretability of the proposed model. The identified equilibrium settings should therefore be regarded as promising initial configurations rather than as optimal values for learning effectiveness.

6. Conclusion

This article presented a pedagogically asymmetric parameterisation of a multidimensional Elo system designed for adaptive mathematics education. The proposed approach shifts the use of the Elo model from a purely

diagnostic assessment of abilities toward supporting the learning process through pedagogically differentiated learner and task dynamics. A pilot implementation in the Edu-AI system and replay-based simulations utilising more than 63,000 historical interactions demonstrated the potential of this approach and showed that target-equilibrium settings substantially influence the long-term dynamics of learner growth and task stability. The scoring policy, which reflects scaffolded task-solving, also plays a significant role here. Therefore, the setting of target values cannot be separated from the way the system evaluates independent solutions, solutions with hints, and unsuccessful attempts.

The most balanced configuration observed in the replay-based simulation used the equilibrium settings 0.80 for learner updates and 0.85 for task updates. These values should be regarded as promising initial settings rather than as optimal parameters for learning effectiveness. Future research will focus on further refining the competency taxonomy, prospectively testing and adjusting equilibrium parameters, identifying pedagogically relevant interactions, and analysing the system's impact on students' educational outcomes.

Acknowledgements

This paper was supported by the project TAČR N. TQ23000334 – AI-based Adaptive Learning Platform.

Ethics Statement

Participation occurred only in classes enrolled in the pilot through their teachers, and the required informed consent was obtained before participation. The analytical dataset contained no personally identifiable information and was processed in anonymised form.

AI Declaration

Generative AI tools were used in preparing this article for graphic design, language support, and text organisation. The authors of this article developed all technical concepts, methodologies, and interpretations.

References

Antal, M. (2013). On the use of elo rating for adaptive assessment. *Studia Universitatis Babes-Bolyai, Informatica*, *58*(1), 29–41.

Abdi, S., Khosravi, H., Sadiq, S. and Gasevic, D. (2019). A multivariate Elo-based learner model for adaptive educational systems. *arXiv preprint arXiv:1910.12581*.

Bjork, R. A. and Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475. https://psycnet.apa.org/doi/10.1016/j.jarmac.2020.09.003.

Bull, S. and Kay, J. (2016). SMILI☺: A framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, 26(1), 293-331. https://doi.org/10.1007/s40593-015-0090-8.

Cai, L., Choi, K., Hansen, M. and Harrell, L. (2016). Item response theory. *Annual Review of Statistics and Its Application*, *3*(1), 297–321. https://doi.org/10.1146/annurev-statistics-041715-033702.

Chen, Y., Li, X., Liu, J. and Ying, Z. (2025). Item response theory – A statistical framework for educational and psychological measurement. *Statistical Science*, 40(2), 167–194. https://doi.org/10.1214/23-STS896.

de Bruin, A. B., Biwer, F., Hui, L., Onan, E., David, L. and Wiradhany, W. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review*, 35(2), 41. https://doi.org/10.1007/s10648-023-09766-w.

Elo, A.E. (1978). *The Rating of Chessplayers, Past and Present*. Arco Pub.

Harvey, R. J. and Hammer, A. L. (1999). Item response theory. *The Counseling Psychologist*, *27*(3), 353–383. https://doi.org/10.1177/0011000099273004.

Jančařík, A., Novotná, J. and Michal, J. (2022). Artificial intelligence assistant for mathematics education. *ECEL*, *21*, 143–148.

Jančařík, A., Michal, J. and Novotná, J. (2023). Using AI chatbot for math tutoring. *Journal of Education Culture and Society*, *14*(2), 285–296. https://doi.org/10.15503/jecs2023.2.285.296.

Jiang, B. (2025). *Toward Trustworthy Adaptive Learning: Explainable Learner Models*. Routledge.

Kassenkhan, A. M., Mendes, M. and Bekarystankyzy, A. (2026). An Adaptive Task Difficulty Model for Personalized Reading Comprehension in AI-Based Learning Systems. *Algorithms*, 19, 100. https://doi.org/10.3390/a19020100

Liu, Z., Guo, T., Liang, Q., Hou, M., Zhan, B., Tang, J., Luo, W. and Weng, J. (2025). Deep Learning Based Knowledge Tracing: A Review, a Tool and Empirical Studies. IEEE Transactions on Knowledge and Data Engineering, 37(8), 4512–4536. https://doi.org/10.1109/TKDE.2025.3552759

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, *98*, 169–179. https://doi.org/10.1016/j.compedu.2016.03.017.

UNESCO, USAID, FCDO, Bill & Melinda Gates Foundation and World Bank. (2020). *Global proficiency framework for mathematics: Grades 1 to 9*. https://www.uis.unesco.org/sites/default/files/medias/fichiers/2025/02/Global-Proficiency-Framework-Math.pdf.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369.

Vermeiren, H., Kruis, J., Bolsinova, M., van der Maas, H. L. and Hofman, A. D. (2025). Psychometrics of an Elo-based large-scale online learning system. *Computers and Education: Artificial Intelligence*, 8, 100376. https://doi.org/10.1016/j.caeai.2025.100376.

Walkington, C. and Bernacki, M. L. (2018). Personalization of instruction: Design dimensions and implications for cognition. *The Journal of Experimental Education*, 86(1), 50–68. https://doi.org/10.1080/00220973.2017.1380590.

The Impact of Content Design on the Effectiveness of an Educational Application

Tomáš Jeřábek, Jakub Michal and David Janda
Charles University, Faculty of Education, Prague, Czech Republic

tomas.jerabek@pedf.cuni.cz
jakub.michal@pedf.cuni.cz
david.janda@pedf.cuni.cz

Abstract: Interactive applications play a significant role in digital educational resources. Their integration has a notable impact, especially in non-formal and informal education. The success of applications depends not only on appropriate content but also on factors that increase users' motivation to use them. These aspects include gamification, activation, personalisation of content and environment, immediate feedback, multimediality or flexibility of application accessibility. The paper focuses on the optimal design of lesson content within an interactive learning application for ISCED 2 students in Mathematics. The paper aims to find relevant aspects and parameters of the application in this context and to identify their appropriate settings for more efficient learning. The study uses quantitative analysis of user behaviour within the application environment, drawing on pedagogical principles such as microlearning, content adaptivity, immediate feedback, and gamification. Lessons are divided into explanation blocks and related practice tasks. As part of the methodology, we evaluate the relationship between the design parameters of these blocks in individual lessons and users' progress through lessons. To assess the optimal setting, we examine the dependence between the overarching mathematical topic, the total amount of content, the number and type of lesson blocks, and tasks. This analysis is based on the theoretical assumption that interactive solving of educational tasks represents the so-called active screen time, which requires the user's direct cognitive involvement. Empirical data for analysis are based on system records of each student's course progress. The key indicators of interaction are analysed, particularly the time spent in individual lessons and blocks of lessons, with their structure taken into account. The synthesis of this objective data allows us to formulate recommendations to optimise educational content that maximise both students' cognitive engagement and the overall effectiveness of the learning process. Also, it helps set an appropriate methodology for further, deeper data analysis.

Keywords: Online Tutoring System, Mathematics Tutoring, Gaming, Content Design

1. Introduction

Tutoring, by which we mean private supplementary teaching conducted in parallel with schooling, is quite common in the Czech Republic. Although exact numbers are unknown, mathematics tutoring is among the most requested services (Šťastný, 2016). This learning support is usually sought to improve grades or to prepare for the high-stakes examinations (e.g. Průcha, 2015).

A free web and mobile app (Edu-AI) for primary and lower-secondary school mathematics tutoring was developed as part of an *AI-based Adaptive Learning Platform* project. It is a loose continuation of the *AI Assistant for Pupils and Teachers* project (see Jančařík et al., 2023). The current goal is to implement generative AI systems beyond chatbot use (e.g., to provide adequate feedback, hints, or personalised tasks). As we strive to make the app motivating for students and effective for learning (which does not depend solely on the quality and quantity of content), we also incorporate design principles that support long-term user motivation and cognitive engagement. Among the most frequently discussed aspects for achieving this goal are gamification elements, activation strategies, personalisation of both content and the learning environment, immediate feedback, multimedia integration, and flexible accessibility across devices and contexts. These features contribute not only to learner satisfaction but also to increased persistence and active participation in the learning process.

In its current piloting phase, the developed tutoring application comprises several teaching courses focused on lower-secondary school mathematics, specifically for 8th- and 9th-grade students. These courses stem from the *Global Proficiency Framework* (GPF, 2020), which distinguishes proficiency levels in greater depth than the *Czech National Curriculum for primary schools* (MŠMT, 2023).

In the pilot testing of the in-development application, students first completed an introductory diagnostic test, followed by eight courses focused on primary school mathematics, released weekly. This paper presents the collected dataset and provides an initial analysis of students' behaviour in the application, focusing on lesson and block characteristics that may influence engagement, particularly the time spent on different activities and possible signs of the so-called "gaming" phenomenon. The findings are intended to inform the further development of an engaging intelligent tutoring system that maintains students' motivation, provides an appropriate level of difficulty, and better adapts to students' interactions with the content.

2. Theoretical Background

Many studies have focused on tutoring or online education in mathematics (e.g., Son, 2024). Although there is no clear consensus on the effects of tutoring, and studies have yielded mixed results (Safarzyńska, 2013), several researchers have reported varying degrees of improvement in students' outcomes (Mabhamba et al., 2024; Zhang et al., 2022). Even though some studies did not observe better results of tutored students, they still observed improvements in students' well-being (Guill et al., 2020). This opportunity, however, is not available to all students, as tutoring is usually a paid service, which might widen the socio-economic gap (Safarzyńska, 2013). Not only for this reason, but also to enrich opportunities for all students, intelligent tutoring systems offered for free could improve this situation.

Even though mobile learning apps or similar environments might motivate students, improve their learning or positively influence their attitude towards mathematics (e.g., Canilao & Gurat, 2023; Taleb et al., 2014), such apps often lack the feedback, immediate assistance and communication facilitated by a real-life tutor, who can tailor the lessons to one's individual needs (Safarzyńska, 2013). There are, however, various approaches to address these restrictions (for instance, incorporating AI language models as chatbots or using animated agents [Jančařík et al., 2023; Baker et al., 2006]). Other issues tutoring systems need to take into account include students' disengagement or lack of motivation, the way curricular materials are presented or the scaffolding methods used.

2.1 Microlearning

Microlearning builds on principles from psychology and pedagogy and is often associated with e-learning (Kapp & DeFelice, 2019). Microlearning is usually explained in terms of either the short *time* during which learning occurs or the small *size* of the targeted teaching units (e.g., Cates et al., 2017). In both instances, learning is organised into small, manageable units of time or information, which is believed to help prevent mental fatigue by reducing demands on working memory. On small mobile screens, such an approach is welcomed (sometimes called *mobile microlearning*), as displaying too much information can lead to cognitive overload (Lee, 2023). Further advantages are also reported in the literature – greater flexibility, a focus on the knowledge to be learned, longer and easier information retention, and stronger engagement (Kadhem, 2017; Kapp & DeFelice, 2019; Lee, 2023; Şahin & Kirmizigül, 2026; Taylor & Hung, 2022).

The effectiveness of microlearning was, for instance, observed in research conducted by Závodná et al. (2024). Students achieved better results with microlearning approaches and were more motivated towards their work than those in the control group. Students further appreciated the visual features, gamification, and interactive quizzes. The effectiveness of microlearning was also observed in mathematics education. Şahin and Kirmizigül (2026) reported that students found microlearning mathematics courses engaging and flexible, and that they appreciated the opportunity to review the content. They also recommend integrating microlearning with face-to-face teaching.

The so-called micro-lectures usually employ different strategies when being made. Based on a literature review, Tang et al. (2022) identify several strategies, including assignments, problem-based learning, feedback interactions, connecting learning to objects from students' daily lives, knowledge strengthening, and experimental demonstrations.

2.2 Gaming the System Issue

One difficulty with tutoring systems is students' disengagement over time (Arroyo et al., 2007; Baker et al., 2006). This can manifest as skimming through the app, immediate use of hints without trying to solve or even read anything offered, or attempts to exploit the system rather than learn – a phenomenon known as „gaming" (Arroyo et al., 2007; Baker et al., 2006). Even though there are instances when gaming might not be problematic (in tasks students know well), it is unwanted behaviour in tasks students struggle with (Baker, 2006).

To address this difficulty, different approaches might be taken – from user-friendly application design and gamification features to self-reflection or self-monitoring systems. For instance, Arroyo et al. (2007) found that non-invasive interventions (such as providing study tips or a performance chart) reduced quick-guessing behaviour and increased students' motivation and re-engagement with a geometry tutoring system. As such behaviour might stem, for instance, from frustration at being unable to solve a challenging task, it is important that extra hints or help be offered (Baker et al., 2006). On the other hand, Arroyo et al. (2007) note that some intrusive attempts to remedy such behaviour (e.g., pop-up windows with information) might be

counterproductive, increasing students' frustration. Thus, they recommend making students aware of the consequences of such actions and their effects on learning progression, without strictly blocking this approach. For instance, Baker et al. (2006) were successful with such a positive approach, using an animated agent (a dog with a range of emotions) designed to improve gaming students' learning outcomes without significantly restricting the system for students who did not participate in gaming. It was done by praising non-gaming students, by challenging students to work carefully when gaming was identified, or by giving additional exercises designed to help gaming students cover the material they had skipped. This led to fewer gaming students (due to the dog's sole presence, regardless of the emotions displayed) and to students who struggled to catch up with their peers on the content.

2.3 Application Content Design

Based on the review above, several principles were adopted in the design of the application Edu-AI. The app is intended for tutoring mathematics and covers all lower secondary school topics. Among others, the microlearning principle was implemented. To increase engagement and prevent gaming, gamification features (collecting experience points that can be used to „buy" in-app avatars or visual modifications) and a visually appealing design were applied. Flexibility is also enhanced by the option to save blocks or quiz questions that a student finds problematic for later review. To accommodate students' diverse personal learning needs (e.g., additional personalised hints, feedback, or questions), an AI chatbot will be integrated in the next step of app development.

The content is based on a thematic course structure. Each course is divided into one or more teaching sections (lessons) intended for 10-minute study sessions. Lessons consist of building units (blocks) of various types–pre-prepared explanatory text blocks, blocks with illustrations, example blocks, or task blocks (see Figure 1). The explanatory blocks focus on helping students understand the concepts they are expected to know as they leave primary school and on connecting the material to what students know from day-to-day experience. In task blocks, students have access to two-step hints and immediate feedback, usually explaining the anticipated mistake (see Figure 1). The correct solution is also displayed before continuing to the following block.

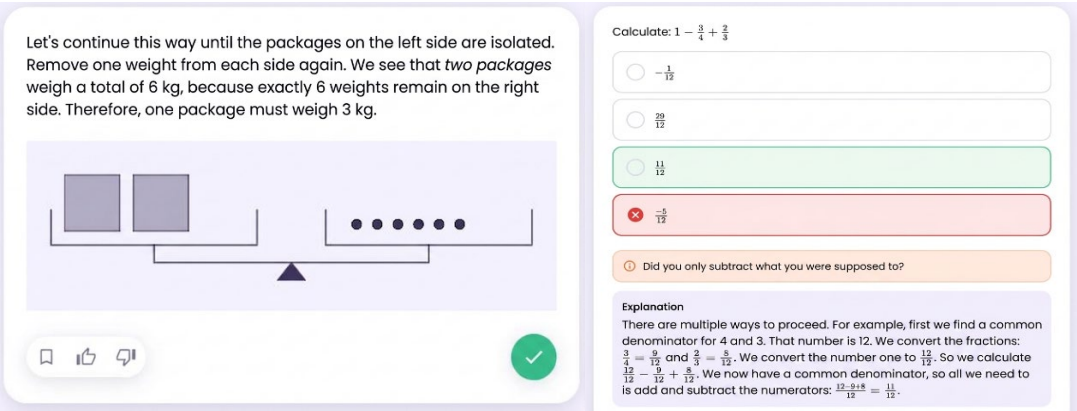


Figure 1: Example of the content block (left) and the task type block (right) (translated)

3. Materials and Methods

Testing of the application took place in March and April 2026 and was conducted in two ways. A total of 17 teachers participated with their classes in the testing.

The data obtained are primarily suitable for analysing specific usage patterns of the app among students from classes that demonstrably worked with the app. An important step before the analysis itself was determining an appropriate method to narrow the observed sample to a set of reliable users.

3.1 Data Selection

For the analysis, four courses (excluding summary quizzes) were selected: *Statistics* (C1); *Measurement in Geometry* (C2); *Fractions* (C3); *Equations and inequations* (C4). It was decided to analyse topics that differ not only from a mathematical point of view (statistics, geometry, arithmetic and algebra) but also in their design (different numbers of blocks, block lengths, and types).

To better analyse students' behaviour across lessons, the lesson blocks were coded. The coding was done by two researchers and compared. Based on the coding, six broader categories (block types) emerged (see Table 1). These were further refined and discussed until a consensus among coders emerged. The *Brief non-explanatory instruction* (B1) block provides short information or text, such as "Let's try it!" and does not offer any additional learning information. A *Short text* (B2) is used as motivation, a lesson introduction, or a recapitulation of the lesson or its content part. *Explanatory text* is the main matter of the lessons in which students are expected to learn or revise the matter. The explanatory text blocks varied greatly in length, so it was decided to distinguish between shorter (B3) and longer (B4) ones using two categories. The other block type in which learning is expected is the block that includes an interactive task. We distinguish between the *Basic task* (B5) and the *Advanced task* (B6).

Table 1: The number of block types in courses

Course / Block Type	B1	B2	B3	B4	B5	B6	Total
C1	0	2	8	7	1	1	19
C2	0	4	4	2	3	4	17
C3	6	4	9	18	7	3	47
C4	12	7	11	20	7	6	63
Total	18	17	32	47	18	14	

The *Statistics* course (C1) was primarily lecture-based and consisted of two lessons. Courses C3 and C4 were similar in structure, containing 3 and 5 lessons, respectively. In both courses, the percentage of exercises was around 20%; overall, *Longer explanatory blocks* (B4) predominated, but there was also a significant proportion of B1 blocks, i.e., blocks not expected to require high cognitive or time demands. Course C2, on the other hand, was more task-oriented (blocks with tasks accounted for 41.2% of the course), and the content blocks tended to be concise.

3.1.1 Narrowing the Sample

For this paper, it was necessary to narrow the sample. Data reduction involved selecting classes with a certain minimum number of students participating in the selected courses. The goal was to include entire groups of students (classes) rather than just individuals, to avoid bias arising from selecting only active participants. In this context, the following criteria were established to identify the selected classes of students: (1) a minimum average of 13 students per class across all 4 courses, (2) a minimum of 9 students in each course, and (3) a minimum average of 33 interactions per student in the given class.

Based on these criteria, 9 classes (G1–G9) were selected, comprising 201 unique students. A more detailed breakdown of students by class and participation in individual courses is shown in Table 2. The total number of students in each course does not add up to 201, because some students in each class did not complete all the courses.

Table 2: Number of students and interactions in individual classes across courses

Class	C1		C2		C3		C4		All Courses	
	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (Mean)	Student inter. (Mean)
G1	19	348	19	358	19	864	18	1175	18.8	36.6
G2	21	385	22	409	20	953	23	1454	21.5	37.2
G3	12	228	19	339	13	611	22	1372	16.5	38.6
G4	13	235	10	180	13	611	17	985	13.3	37.9
G5	15	268	14	235	14	613	21	1258	16.0	37.1
G6	11	184	18	318	17	690	15	875	15.3	33.9
G7	16	297	17	306	15	667	20	1205	17.0	36.4
G8	18	342	21	378	20	907	22	1386	20.3	37.2

Class	C1		C2		C3		C4		All Courses	
	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (n)	Student inter. (n)	Students (Mean)	Student inter. (Mean)
G9	9	171	20	359	21	987	22	1393	18.0	40.4
Total	134	2458	160	2882	152	6903	180	11103		

3.2 Data Analysis Procedure

The analysis evaluated the course design and its individual blocks using interaction data collected by the Edu-AI application. We decided not to automatically remove outliers from the time data captured in the application logs. Both extremely short and extremely long times may indicate the nature of user interaction with a given block, particularly regarding gaming behaviour, inactivity, or technical interruptions. At the same time, the way response-time outliers are handled can significantly influence the interpretation of results (Kovanović et al., 2015; Vankov, 2023). Therefore, user interaction times were classified rather than simply removed.

Interactions were classified separately for each block into four duration-based categories: *short_irrelevant*, *short_relevant*, *long_relevant*, and *long_irrelevant* as described below.

The classification combined relative and absolute thresholds, drawing on the MAD-based approach described by Leys et al. (2013) and the time-on-task issues discussed by Kovanović et al. (2015). Because the observed interaction-time distributions were strongly right-skewed, times were log-transformed before the median absolute deviation (MAD) was calculated. The MAD was multiplied by 1.4826, the conventional consistency factor that makes it comparable to the standard deviation under normality. The main threshold was set at 2.5 scaled MAD from the block-specific median. Although more conservative thresholds, such as 3 MAD, have also been proposed (Leys et al., 2019), the more sensitive threshold was retained because the aim was to classify potentially informative interactions rather than exclude them.

For each block, the lower and upper thresholds were calculated as the block-specific median plus and minus 2.5× scaled MAD, respectively, on the log-transformed scale, and then transformed back into seconds. If the lower threshold was too small (less than 1 s), it was set to 1 s. Similarly, when the upper threshold was below 600 s, it was treated as 600 s. Interactions at or below the lower threshold were classified as *short_irrelevant*; those above the lower threshold and at or below the median as *short_relevant*; those above the median and at or below the upper threshold as *long_relevant*; and those above the upper threshold as *long_irrelevant*.

The categories were retained as indicators of potentially different interaction patterns, not as direct measures of students' attention, engagement, or understanding.

4. Results

Overall response-time statistics for individual blocks per course are presented in Figure 2. It presents the mean of the medians of interaction time for each block type in the specific course. The number of interactions per block ranged from 124 to 180, with an average of 153.

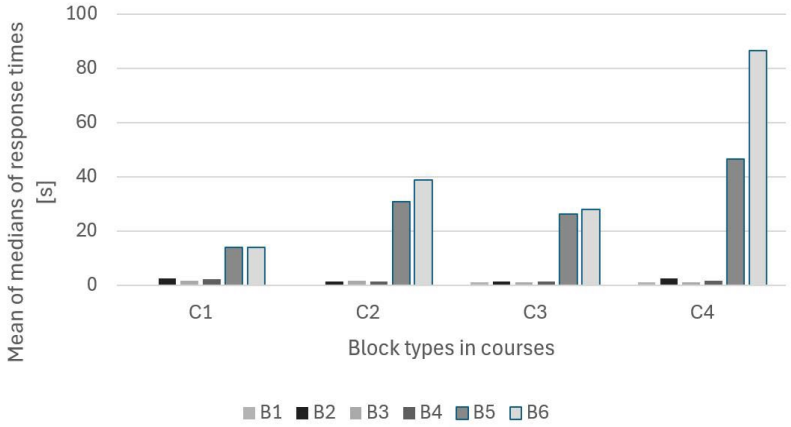


Figure 2: Mean of medians of response times (MMRT) for all interactions in courses and blocks by their type.

Response times were extremely low for blocks B1–B4. The block-level medians, summarised as the mean of medians, indicate that a typical block of this type was passed very quickly, often within one or two seconds. This suggests that many students did not spend enough time on explanatory or introductory blocks for substantial engagement with the content to be likely. For example, in the *Statistics* course, the median interaction time in blocks classified as longer explanatory text (B4) was very low, with block-level medians ranging only from approximately 1 to 4 seconds. This suggests that many student-block interactions with explanatory content were very brief.

If these blocks are compared with the task blocks (B5, B6), a surge in response times can be observed. For example, in the C3 course *Fractions*, 7 blocks were classified as *Basic tasks* (B5). In these, roughly half of the students spent more than 26 seconds on average. While this can be partially explained by technical differences, it is not necessarily the case. Figure 2 also shows that the medians distinguished between basic and advanced tasks (as classified by the authors). In the case of the C4 course, *Equations and inequations*, this difference is striking, indicating the actual attention paid to the tasks.

However, given the very steeply skewed log-normal distribution, this overview has its visible limitations. Response time medians provide a good estimate of whether gaming occurs in the majority of the sample and also help identify significant systemic differences (such as those between tasks and other blocks). On the other hand, they do not provide further information, for example, on the specific numbers of interactions and their characteristics. We attempted to capture these using the analysis described in section 3.2.

4.1 Evaluation of Block Types Based on the Student-Block Interactions

The MAD-based classification provided only limited additional information as a four-category classifier, especially for explanatory blocks. This limitation follows from the extremely short interaction times observed, mostly in non-task blocks (B1-B4). When a substantial part of the sample spends only one or two seconds on a block, the distinction between *short_irrelevant* and *short_relevant* becomes difficult to interpret substantively. In this context, the block-specific median appears to be the most practically meaningful threshold. Interactions below or at the median indicate only minimal engagement with the block in the learning context. This phenomenon is evident for all information blocks (B1 - B4) regardless of their scope and purpose (explanation of the material, motivation, recapitulation, etc.). The results show us that for these block types in most cases the interactions were so short that the students could not even read the content or cognitively grasp it, especially for explanatory blocks (B3, B4).

In the case of blocks B5 and B6, our method of response-time classification provides more space for reasonable interpretations. Within block B5, we identified a not-negligible number of relevant interactions (see Table 3), The representation of relevant interactions across all courses was 96%, with the lowest value of 95% in C4. It can be observed that the relative proportion of individual interaction types (short/long relevant/irrelevant) does not significantly correlate with the mean of medians (MMRT) in individual courses (see Figure 2), and is therefore essentially independent of the length of solving tasks (even without counting C1, where there was only one B5 block).

Table 3: Overview of interactions for blocks B5

Course	Number of Blocks (n)	Number of Interactions (n)				
		Total	short_irrelevant	short_relevant	long_relevant	long_irrelevant
C1	1	132	0	71	60	1
C2	3	476	0	240	232	4
C3	7	1038	40	465	532	1
C4	7	1233	51	550	624	8

5. Discussion

All recorded interaction times were directly related to the phenomena under investigation, including rapid progression through the app, possible gaming behaviour, inactivity, and technical interruption. Therefore, instead of trimming short and long outliers, we used a method (Leys et al., 2013) to classify student-block

interactions by their duration. Other methods for defining response-time thresholds (Stipek et al., 2025) could also be applied and may capture different aspects of student behaviour. Further work should therefore compare alternative classification approaches and examine their suitability for this type of application data.

At this stage, the data allow only a tentative comparison between courses, but they reveal interesting findings regarding block types. In longer explanatory blocks (B4), the share of *long_relevant* interactions was higher in the shorter courses C1 and C2 than in C3, and C4 occupied an intermediate position. This pattern suggests a working hypothesis that course length and the number of explanatory blocks may influence students' willingness to engage with explanatory content. We also observe a significant difference in interaction times between content-type blocks (B1-B4) and task blocks (B5, B6). For blocks B1-B4, we observe very short interactions, indicating little interest in the content presented. This lack of interest may be due to the fact that students in this type of application are more interested in solving tasks (e.g. similar to Duolingo). From a gamification perspective, the cause can also be identified at the motivational level, where a higher cognitive grasp of the content does not bring any benefit compared to quick skimming.

If any blocks captured students' attention, they were the tasks (B5 or B6). The sample selection might have influenced this, as students were just a few weeks before the entrance examination, and those who had been preparing themselves preferred to revise tasks similar to the examination format. In these blocks, we can observe more time-meaningful interactions, with a lower level of gaming and a more expected effective impact on student learning. However, it is necessary to take into account secondary factors that may influence the interaction time and that we are not able to detect using the method employed (e.g., viewing the solution or previous blocks). The four-item structuring of interactions also appears to be beneficial for subsequent, more detailed data analyses. For instance, in course C3, we can see 40 *short_irrelevant* interactions within B5, which can be used to identify ill-behaved tasks (which are, for example, gamed more often than others) or to model student behaviour (for example, identifying those that have not one but several irrelevant interactions).

In terms of block classification, the method used also more or less confirmed the correct classification of Basic and Advanced task blocks, with the difference in time spent on these blocks being obvious, especially in course C4 (see Figure 2).

6. Conclusion

The present analysis suggests two follow-up tasks. First, it is necessary to distinguish students whose interactions fall into irrelevant categories but who may nevertheless have completed the course responsibly, for example, because they processed some blocks quickly but consistently. Second, it is necessary to identify students whose overall interaction time exceeds the median yet still exhibit local gaming behaviour in specific parts of the course. These two groups are important because they cannot be reliably classified by overall course time alone. Together, they form a suitable basis for a secondary analysis aimed at evaluating individual blocks more precisely: not only whether a block was passed quickly or slowly, but whether its position, format, and function in the course contributed to meaningful engagement or encouraged superficial progression through the app.

In relation to the course or lesson structure, based on the findings we shall consider including more task type blocks that seem more engaging for students, or to transform explanatory blocks into another type that requires more interaction or connects more with the task type block.

In the future, it is also important to incorporate more features that limit the gaming phenomenon in the app, including testing new, more engaging block types. Additional qualitative data (such as interviews with students or teachers' feedback) are needed to help clarify some questions about blocks, students' behaviour in the app, and other observed phenomena.

Acknowledgements

This paper was supported by the project TAČR N. TQ23000334 – AI-based Adaptive Learning Platform.

Ethics Declaration

The research did not involve any personally identifiable information about the students. The data was processed in an anonymised form.

AI Declaration

Generative AI tools were used in the preparation of this article for graphic design, language support, and text organisation. The authors of this article developed all technical concepts, methodologies, and interpretations.

References

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (pp. 195–202). IOS Press.

Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent tutoring systems: ITS 2006* (Lecture Notes in Computer Science, Vol. 4053). Springer. https://doi.org/10.1007/11774303_39

Canilao, C. M., & Gurat, M. G. (2023). Effect of Mobile App on Students' Mathematics and Technology Attitude. *American Journal of Educational Research*, 11(10), 722–728.

Cates, S., Barron, D., & Ruddiman, P. (2017). MobiLearn Go: Mobile microlearning as an active, location-aware game. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Article 103, pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3098279.3122146

Guill, K., Lüdtke, O., & Köller, O. (2020). Assessing the instructional quality of private tutoring and its effects on student outcomes: Analyses from the German National Educational Panel Study. *British Journal of Educational Psychology*, 90(2), 282–300. https://doi.org/10.1111/bjep.12281

Jančařík, A., Michal, J., & Novotná, J. (2023). Using AI Chatbot for Math Tutoring. *Journal of Education Culture and Society*, 14(2), 285–296.https://doi.org/10.15503/jecs2023.2.285.296

Kadhem, H. (2017). Using mobile-based micro-learning to enhance students' retention of IT concepts and skills. 2017 International Conference on Knowledge Engineering and Applications (ICKEA), 128–132. https://doi.org/10.1109/ICKEA.2017.8169915

Kapp, K. M., & DeFelice, R. A. (2019). *Microlearning: short and sweet*. ATD Press.

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, 2(3), 81–110.https://doi.org/10.18608/jla.2015.23.6

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Mabhamba, M., Mukuka, A., & Alex, J. K. (2024). Evaluating the effectiveness of private supplementary tutoring on grade 12 learners' mathematics achievement. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(9), em2507. https://doi.org/10.29333/ejmste/15168

MŠMT. (2023). *Rámcový vzdělávací program pro základní vzdělávání*. Výzkumný ústav pedagogický. https://www.edu.cz/wp-content/uploads/2023/07/RVP_ZV_2023_cista_verze.pdf

Průcha, J. (2015). *Česká vzdělanost: Multidisciplinární pohled na fenomén národní kultury*. Praha: Wolters Kluwer.

Safarzyńska, K. (2013). Socio-economic Determinants of Demand for Private Tutoring. *European Sociological Review*, 29(2), 139–154.

Şahin, Z. G., & Kirmizigül, H. (2026). Reflections from the mathematics lesson with micro learning: A different educational experience in the digital age. J*ournal of Learning and Teaching in Digital Age*. Advance online publication. https://doi.org/10.53850/joltida.1702683

Son, T. (2024). Intelligent Tutoring Systems in Mathematics Education: A Systematic Literature Review Using the Substitution, Augmentation, Modification, Redefinition Model. *Computers*, 13(10), 270. https://doi.org/10.3390/computers13100270

Štípek, J., Voňková, H., & Králová, K. (2025). Vztah rychlého odpovídání v testu z anglického jazyka a výsledného skóre u různých skupin žáků. *Orbis Scholae*, 19 (1), 41-66. https://doi.org/10.14712/23363177.2025.7

Šťastný, V. (2016). *Fenomén soukromého doučování jako stínový vzdělávací systém v České republice* [dizertační práce]. Univerzita Karlova, Pedagogická fakulta, Oddělení pro vědeckou činnost.

Taleb, Z., Ahmadi, A., & Musavi, M. (2015). The effect of M-learning on mathematics learning. *Procedia - Social and Behavioral Sciences*, 171, 83–89.https://doi.org/10.1016/j.sbspro.2015.01.092

Tang, J., Wijaya, T. T., Weinhandl, R., Houghton, T., Lavicza, Z., & Habibi, A. (2022). Effects of Micro-Lectures on Junior High School Students' Achievements and Learning Satisfaction in Mathematics Lessons. *Mathematics*, 10(16), 2973. https://doi.org/10.3390/math10162973

Vankov, I. I. (2023). The hazards of dealing with response time outliers. *Frontiers in Psychology*, 14, 1220281.

Zhang, Y., Cui, C., He, Y., & Wang, L. (2022). Does private supplementary tutoring matter in Chinese students' learning of mathematics: A longitudinal study. *ZDM – Mathematics Education*, 54(3), 737–747. https://doi.org/10.1007/s11858-022-01346-6

Generative Artificial Intelligence Use Among Students: A Cross-National Comparison Between Türkiye and England

Merve Kara[1], Zeynep Cömert[1] and Kübra Kaymakçı Üstüner[2]
[1]Bahcesehir University, Istanbul, Türkiye
[2]Durham University, Durham, United Kingdom

karamerve0@gmail.com
zeynep.comert@bau.edu.tr
kubra.kaymakci-ustuner@durham.ac.uk

Abstract: Generative artificial intelligence (GenAI) tools are increasingly present in children's everyday digital lives, yet research on primary school students remains limited. This study examines how students in Türkiye and England conceptualise and use GenAI, and whether family AI use is associated with students' own use. A quantitative cross-national survey was conducted with 359 students in Türkiye and 98 students in England. The questionnaire included open-ended and closed-ended items on AI conceptualisations, frequency of use, purposes of use, and family AI use. Open-ended responses were coded thematically, and chi-square tests were used to examine group differences and associations. Findings showed that students in Türkiye more often described AI in technological and functional terms, whereas unclear understanding was more common among students in England. GenAI use frequency did not differ significantly between countries, with most students reporting occasional use. However, purposes of use varied: students in Türkiye reported more diverse and creative uses, while students in England more often used GenAI for information seeking or reported unclear/non-use. In the England sample, family AI use was strongly associated with students' own use frequency. The study highlights the need for developmentally appropriate AI literacy initiatives involving both schools and families.

Keywords: Generative Artificial Intelligence, AI Literacy, Primary School, Digital Technology Use, Cross-Cultural Comparison

1. Introduction

Artificial intelligence (AI), particularly generative AI (GenAI), is becoming increasingly embedded in children's everyday digital environments. Conversational agents, content generators, and AI-powered learning tools are no longer used only by adults or older students; children are also encountering these technologies at younger ages. Recent evidence suggests that many children aged 7–12 already interact with GenAI tools, indicating a rapid diffusion of these technologies into childhood digital practices (Mansfield et al., 2025). As AI becomes more accessible in both domestic and educational settings, understanding how children perceive, interpret, and use these tools has become an important research priority.

Existing research shows that children often conceptualize AI through familiar and human-centred ideas. Rather than explaining AI in technical terms, they frequently describe it as a robot, helper, intelligent machine, or human-like entity capable of thinking, responding, or assisting (Andries & Robertson, 2023; Samuelsson, 2023). Such descriptions suggest that children's early understandings of AI are shaped by everyday experiences and social interaction rather than by formal technical knowledge. At the same time, children's awareness of how AI systems process data, generate responses, and raise ethical or privacy-related concerns remains limited. This gap between use and understanding highlights the need to examine children's conceptualizations of AI alongside their actual patterns of engagement.

Children's engagement with GenAI is also shaped by social and familial contexts. Previous research on children's technology use shows that parents' and family members' practices influence when, how, and for what purposes children use digital tools (Beneteau et al., 2020; Druga et al., 2022; Nikken & Jansz, 2014). Children may encounter AI not only through school-based platforms, but also through siblings, parents, peers, household devices, games, and informal online activities. Therefore, children's early experiences with GenAI may be mediated as much by family and home environments as by formal educational policy or classroom practice.

This study focuses on Türkiye and England because these two contexts provide a meaningful comparison in relation to AI in education. Both have recently developed policy-level discussions and guidance concerning AI use, yet their emphases differ. In England, policy discourse has more strongly highlighted the pedagogical potential of AI and its possible role in supporting teaching and learning, whereas in Türkiye, recent policy developments have placed greater emphasis on ethical principles, safety, transparency, and responsible use. However, it remains unclear whether these national policy orientations are reflected in children's everyday encounters with GenAI, especially when much of their AI exposure may occur informally outside school. In response to this gap, the present study examines how primary school students in Türkiye and England

conceptualize and use GenAI, how frequently they use it, for what purposes they use it, and whether family members' AI use is associated with students' own AI use behaviors. By adopting a cross-national perspective, the study aims to identify both shared patterns and contextual differences that can inform age-appropriate AI literacy initiatives involving schools, families, and wider digital environments.

1.1 Research Questions

RQ1. How do students in Türkiye and England conceptualize artificial intelligence?

RQ2. How frequently do students use generative artificial intelligence?

RQ3. How do students' purposes of using generative artificial intelligence (e.g., homework, research, design, entertainment) differ between Türkiye and England?

RQ4. Is the presence of family members who use AI related to students' own AI use behaviours?

2. Background

2.1 Generative AI in Education

Generative artificial intelligence (GenAI), particularly large language model-based tools, has become increasingly visible in education by enabling natural language interaction, generating contextually relevant responses, and providing on-demand learning support (Bauer et al., 2025; OECD, 2026). With tools such as ChatGPT becoming widely accessible, AI is no longer limited to institutionally provided technologies but can also be used independently by students in their everyday learning practices.

Research suggests that students use GenAI for a range of purposes, including accessing information, understanding concepts, completing homework, and generating content (Zhu et al., 2025). These patterns indicate that GenAI can function both as a cognitive support tool and as a production-oriented assistant. However, students may also use these tools to speed up or simplify tasks rather than to deepen learning, raising questions about the quality of engagement and the development of AI literacy (Caffaratti et al., 2025; OECD, 2026).

A further issue is that students' GenAI use often occurs outside formal school settings. Many students access these tools at home and use them according to individual needs, often without systematic teacher guidance (Zhao et al., 2025; OECD, 2026). Although teachers are beginning to use GenAI for tasks such as lesson planning, content creation, and material preparation, systematic classroom integration remains limited (Diliberti et al., 2024; OECD, 2026). This suggests that students' early encounters with GenAI may be shaped not only by formal education but also by informal and domestic contexts.

2.2 Use of Generative Artificial Intelligence in Türkiye and England

Türkiye and England provide a meaningful basis for comparison because both contexts have recently addressed generative artificial intelligence in education, yet they differ in how AI is framed at the policy level. In England, official guidance from the Department for Education emphasises the potential of AI to support teaching and learning, reduce teacher workload, and contribute to pedagogical processes, while also recognising risks related to safety, accuracy, and data protection (Department for Education, 2023). Thus, AI is positioned both as an educational opportunity and as a technology requiring responsible implementation.

In Türkiye, recent policy developments have placed stronger emphasis on the ethical, safe, transparent, and responsible use of AI in educational settings. The Ethical Guidelines for Artificial Intelligence Applications in Education, published by the Ministry of National Education, identifies principles intended to guide educational stakeholders and manage potential risks associated with AI use (Ministry of National Education, 2026). Within this context, AI in education is framed primarily through ethical governance, awareness raising, data security, and responsible use.

Overall, while England more explicitly emphasises the pedagogical integration of AI, Türkiye foregrounds ethical principles and responsible implementation. However, in both contexts, children's actual encounters with GenAI may occur outside formal school structures, particularly through family members, peers, home technologies, and informal digital practices. This creates an important tension between policy-level expectations and

children's everyday experiences, which the present study addresses through a cross-national comparison of students' GenAI perceptions and use.

2.3 Theoretical Perspective

This study is informed by social constructivist and developmental perspectives on children's encounters with generative artificial intelligence. GenAI systems are increasingly relevant to children's everyday experiences because they provide dialogue-based interaction and respond in ways that may resemble human communication. Research in child development has long emphasised the role of conversational exchanges with parents, siblings, teachers, and peers in supporting language development, cognitive growth, and social understanding (Golinkoff et al., 2019; Gelman, 1990). As conversational AI systems become part of children's digital environments, it is important to consider how children interpret these systems within broader social and communicative contexts (Xu, 2023).

Social constructivism provides a useful lens for examining children's perceptions of AI because children's understandings of new concepts are shaped through interaction with their social and cultural environments rather than through isolated individual experience (Vygotsky, 1978). From this perspective, children's ideas about GenAI may develop through everyday interactions with family members, peers, digital media, and household technologies. Previous research similarly suggests that children often learn about emerging technologies through informal encounters with family members and peers rather than through structured instruction (Druga et al., 2022; Zhao et al., 2025). This perspective is directly relevant to the present study, which examines not only students' own AI use but also the role of family members' AI use in shaping children's engagement with these tools.

Vygotsky's concept of the "more knowledgeable other" further supports the focus on family influence. Family members who use AI may introduce children to these tools, model possible uses, and provide informal guidance about when and how AI can be used. In this sense, the family can function as an informal learning environment in which children's early AI experiences are socially mediated. This theoretical perspective is especially important because children's encounters with GenAI may occur at home as much as, or even more than, in formal school settings.

Developmental perspectives also help explain why children may conceptualize AI in anthropomorphic or human-centred ways. Children in middle childhood are still developing the ability to understand complex abstract systems and may rely on familiar concepts when interpreting unfamiliar technologies. Therefore, descriptions of AI as a robot, helper, or human-like entity may reflect age-appropriate attempts to make sense of computational systems through existing mental models (Andries & Robertson, 2023; Samuelsson, 2023).

Finally, the study is informed by contemporary discussions of AI literacy. AI literacy is often conceptualized as a multidimensional construct involving awareness, practical use, critical evaluation, ethical understanding, and knowledge of how AI systems operate (Zhou et al., 2025). However, given the age of the participants, this study focuses primarily on children's awareness, conceptual understandings, and usage experiences rather than advanced technical or ethical competencies. Combining social constructivist, developmental, and AI literacy perspectives therefore provides a framework for interpreting how children in Türkiye and England encounter, understand, and use GenAI within different social and national contexts.

3. Methodology

This study used a quantitative descriptive survey design to examine primary school students' use of, and awareness of, generative artificial intelligence in Türkiye and England. During the data collection process, a questionnaire was administered to students in both countries. Within the scope of the study, data were collected from 98 students in England and 359 students in Türkiye; valid sample sizes varied slightly by item because of unclear or missing responses. The participants were aged between 9 and 11.

The questionnaire was developed specifically for this study and included both open-ended and closed-ended items addressing students' conceptualisations of AI, frequency and purposes of GenAI use, awareness of these technologies, and family members' AI use.

All data were collected anonymously, and the research process was conducted in accordance with ethical guidelines. Ethical approval for the study was obtained from the relevant institution. Prior to data collection,

informed consent was obtained from school administrations, teachers, parents, and students. Participation was voluntary, and no compensation was provided to the participants.

3.1 Data Analysis

The data were analysed in accordance with the four research questions. For the first research question, students' open-ended responses defining artificial intelligence were analysed through content-based coding. The responses were classified into five conceptual categories: anthropomorphic/robotic, functional/help-oriented, technological/tool-based, evaluative/affective, and unclear/no clear understanding. Frequency and percentage distributions were calculated separately for the Türkiye and England samples. A chi-square test of independence was then conducted to determine whether the distribution of conceptual categories differed significantly by country. Cramer's V was calculated to evaluate the strength of the association.

For the second research question, responses concerning frequency of GenAI use were coded as never, sometimes, or very often. Descriptive statistics were calculated for each country, and a chi-square test of independence was performed to examine whether reported frequency of use differed between the Türkiye and England samples. Unclear or missing responses were excluded from the inferential analysis.

For the third research question, students' open-ended responses regarding their purposes of GenAI use were analysed using a multiple-response coding approach, as some participants reported more than one purpose of use. Responses were coded into four thematic categories: education and academic support, information and research, games and entertainment, and creative production. Responses indicating non-use or unclear purposes were coded separately. Frequencies and percentages were calculated on the basis of the total number of cases in each country sample.

For the fourth research question, students' responses regarding family members' AI use were coded as yes, no, or unclear/missing. Frequency distributions were first calculated to describe family AI use in each country. Cross-tabulation and chi-square analysis were then used to examine the association between family AI use and students' own reported frequency of GenAI use. Since the Türkiye sample showed almost no variation in the family-use variable, inferential analysis for this relationship was conducted only for the England sample. Cramer's V was calculated to assess effect size.

3.2 Limitations

While AI literacy is often conceptualized as a multidimensional construct involving technical understanding, critical evaluation, ethical awareness, and practical use, the present study focuses primarily on children's awareness and usage-related dimensions. Therefore, the findings should not be interpreted as a comprehensive assessment of AI literacy.

4. Findings

4.1 RQ1. How do Students in Türkiye and England Conceptualize Artificial Intelligence?

To examine children's conceptual understandings of artificial intelligence, the open-ended responses were coded into five categories.

Table 1: Conceptual Categories of Artificial Intelligence by Country

Conceptual Category	Türkiye n	Türkiye %	England n	England %
Anthropomorphic/robotic	77	21.4	16	16.3
Functional/help-oriented	96	26.7	19	19.4
Technological/tool-based	105	29.2	13	13.3
Evaluative/affective	52	14.5	13	13.3
Unclear / no clear understanding	29	8.1	37	37.8
Total	359	100.0	98	100.0

As shown in Table 1, children in Türkiye and England differed in how they conceptualized artificial intelligence. In the Türkiye sample, technological/tool-based and functional/help-oriented descriptions were the most common, whereas in the England sample the largest category was unclear/no clear understanding. While both groups also produced anthropomorphic/robotic, technological/tool-based, and evaluative/affective descriptions, the relative distribution of these categories varied across countries. The chi-square analysis showed that this difference was statistically significant, $\chi^2(4, N = 458) = 36.72$, $p < .001$, with a Cramer's V of .28, indicating a meaningful association between country and children's conceptualizations of AI.

4.2 RQ2. How Frequently do Students use Generative Artificial Intelligence?

To examine students' reported frequency of generative AI use, frequency and percentage distributions were calculated by country, followed by a chi-square test of independence.

Table 2: Frequency of Generative AI Use by Country

Frequency of GenAI Use	Türkiye n	Türkiye %	England n	England %
Never	92	25.6	17	18.7
Sometimes	221	61.6	57	62.6
Very often	46	12.8	17	18.7
Total	359	100.0	91	100.0

Table 2 shows the distribution of students' reported frequency of generative AI use in the Türkiye and England samples. In both countries, the most frequently reported category was "sometimes". In the Türkiye sample, 61.6% of students reported using generative AI sometimes, 25.6% reported never using it, and 12.8% reported using it very often. In the England sample, after excluding seven unclear or missing responses, 62.6% of students reported using generative AI sometimes, while 18.7% reported never using it and another 18.7% reported using it very often. A chi-square test of independence indicated that the distribution of AI use frequency did not differ significantly by country, $\chi^2(2, N = 450) = 3.24$, $p = .197$. The effect size was small (Cramer's V = .08), suggesting that students in Türkiye and England showed broadly similar patterns in their reported frequency of generative AI use.

4.3 RQ3. How do Students' Purposes of Using Generative Artificial Intelligence (e.g., Homework, Research, Design, Entertainment) Differ Between Türkiye and England?

To examine students' reported purposes for using GenAI, the open-ended responses were coded using a multiple response framework, allowing for students who reported overlapping usage purposes to be accurately represented.

Table 3: Reported Purposes of Generative AI use by Country (Multiple Response)

Purpose Category	Türkiye (n)	Türkiye % of Cases	England (n)	England % of Cases
Education & Academic Support	87	24.2	18	18.4
Information & Research	103	28.7	32	32.7
Games & Entertainment	91	25.3	21	21.4
Creative Production	75	20.9	1	1.0

Purpose Category	Türkiye (n)	Türkiye % of Cases	England (n)	England % of Cases
Non-users / Unclear	3	0.8	35	35.7
Total Cases (N)	359	100.0*	98	109.2*

(Note: Because a multiple response analysis was employed for the England sample where students reported overlapping purposes, the total percentage of cases for England exceeds 100%. The Türkiye sample did not report overlapping purposes in single responses.)

Table 3 presents the distribution of students' reported purposes for using generative AI across the Türkiye and England samples. In the Türkiye sample, students demonstrated a relatively balanced and diverse distribution of AI usage, with Information & research (28.7%) and Games & entertainment (25.3%) being the most common, followed closely by Education & academic support (24.2%). Notably, Creative production—such as prompting AI to write stories, draw pictures, or code—emerged as a substantial and unique use case in Türkiye (20.9%).

Conversely, in the England sample, the primary use case was information and research, with students positioning GenAI mainly as a functional tool for finding facts or asking questions. The multiple-response analysis also showed that some students used GenAI for games and entertainment and for education-related purposes, sometimes reporting these purposes together (e.g., “work and games”). However, engagement in creative production was very limited among students in England. In addition, a higher proportion of students in the England sample reported non-use or unclear purposes compared with the Türkiye sample. These findings indicate that, while both groups used GenAI for research and educational support, students in Türkiye reported broader and more creative applications, whereas students in England more often used GenAI as a functional information-retrieval tool or reported unclear/non-use.

4.4 RQ4. Is the Presence of Family Members who use AI Related to Students’ Own AI use Behaviours?

To examine the role of family AI use, frequency distributions were calculated and cross-tabulation analysis was conducted between family AI use and students’ own reported frequency of AI use.

Table 4A: Family AI Use by Country

Family AI Use	Türkiye (n)	Türkiye %	England (n)	England %
Yes / at least one family member	359	99.7	61	62.2
No	0	0.0	16	16.3
Unclear / missing	1	0.3	21	21.4
Total	360	100.0	98	100.0

Table 4A shows the distribution of family AI use across the two country samples. In the Türkiye sample, almost all students (99.7%) reported that at least one family member used AI, whereas in the England sample 62.2% reported family AI use, 16.3% reported no family AI use, and 21.4% provided unclear or missing responses. Because the Türkiye family-use variable showed almost no variation, a statistical test of association could not be meaningfully conducted for that sample.

Table 4B: Cross-tabulation of Family AI Use and Students' Own AI Use Frequency in the England Sample

Family AI Use	Never n (%)	Sometimes n (%)	Very Often n (%)	Total
No	7 (43.8)	9 (56.3)	0 (0.0)	16
Yes	3 (4.9)	41 (67.2)	17 (27.9)	61
Total	10	50	17	77

For the England sample, the cross-tabulation presented in Table 4B indicated that students who reported family AI use were more likely to report using AI themselves more frequently. Among students who reported family AI use, 67.2% reported using AI sometimes and 27.9% reported using it very often, whereas none of the students who reported no family AI use reported using AI very often. A chi-square test of independence showed that the association between family AI use and students' own frequency of AI use was statistically significant in the England sample, $\chi^2(2, N = 77) = 19.41$, $p < .001$. The effect size was large (Cramer's $V = .50$), indicating a strong association between family AI use and students' reported AI use frequency.

Taken together, the findings show that children's engagement with GenAI differs not only in terms of conceptual understanding and usage purposes, but also in relation to family context. While students in both countries reported similar overall frequency of use, cross-national differences were observed in how they understood AI and the purposes for which they used it. These findings are interpreted in the following discussion in relation to children's AI literacy, developmental understandings of AI, and informal family-based mediation.

5. Discussion

The findings of this study show that children in Türkiye and England engage with generative artificial intelligence in broadly similar ways in terms of frequency but differ in how they conceptualize and use these tools. Regarding RQ1, students' definitions suggest that many children understand AI through familiar, concrete, and human-centered concepts, such as robots, helpers, intelligent machines, or systems that think like humans. This finding is consistent with previous studies showing that children often attribute human-like characteristics to AI systems and evaluate them as socially responsive entities (Andries & Robertson, 2023; Samuelsson, 2023). From a developmental perspective, such anthropomorphic interpretations may reflect children's attempts to make sense of abstract computational systems through familiar social categories. At the same time, the high proportion of unclear responses, particularly in the England sample, suggests that children's growing exposure to AI does not necessarily correspond to a clear understanding of how these systems work. This supports the view that children's AI literacy may develop more slowly than their practical exposure to AI technologies (OECD, 2026).

Regarding RQ2, most students in both countries reported using GenAI only sometimes. This suggests that, although GenAI is becoming increasingly visible in children's digital lives, it has not yet become a fully routinised everyday practice among this age group. The relatively occasional nature of use may also reflect the developmental stage of the participants, as AI use tends to increase with age and educational level (OECD, 2026). From a social constructivist perspective, children's technology practices develop gradually through repeated exposure, social modelling, and participation in meaningful activities. Therefore, the predominance of occasional use may indicate an emerging stage of engagement in which children are still exploring the possible functions of GenAI rather than using it as a stable learning or production tool.

Regarding RQ3, students in both countries reported using GenAI mainly for information seeking, asking questions, homework support, games, and entertainment. However, students in Türkiye reported more diverse purposes, including creative production such as writing stories, generating images, or coding, whereas students in England more often positioned GenAI as a functional information-seeking tool or reported unclear/non-use. This difference suggests that children's GenAI use may be shaped not only by access to technology, but also by the social and cultural contexts in which technology use is modelled and encouraged. Through a social constructivist lens, technology use can be understood as a socially situated practice: children learn how tools should be used by observing family members, peers, teachers, and wider digital communities. Therefore, cross-national differences in reported purposes may reflect differences in family practices, digital cultures, and opportunities for informal experimentation rather than differences in the technology itself.

Regarding RQ4, the findings highlight the role of family environments in shaping children's engagement with AI. In the England sample, family AI use was strongly associated with students' own frequency of GenAI use. Although this finding does not imply causation, it indicates that children whose family members use AI may have more opportunities to observe, discuss, and experiment with these tools. This interpretation is consistent with Vygotsky's concept of the more knowledgeable other, as family members may introduce children to AI, model its use, and provide informal guidance about its possible applications. In this sense, the family can be understood as an important site of informal AI learning. The findings therefore suggest that AI literacy should not be treated only as a school-based responsibility, but as a broader developmental and social process involving families, schools, and children's everyday digital environments.

Overall, the study demonstrates that children's engagement with GenAI is shaped by the interaction of conceptual understanding, usage frequency, usage purposes, and family context. While children in Türkiye and England showed similar levels of occasional use, their conceptualizations and purposes of use differed in meaningful ways. These findings suggest that early experiences with GenAI are embedded in wider social, cultural, and familial contexts. Consequently, AI literacy initiatives for primary school students should move beyond technical definitions and include age-appropriate discussions of how AI works, what it can and cannot do, and how children encounter these technologies in both school and home settings.

6. Conclusion

This study shows that children's engagement with generative artificial intelligence is shaped by the interaction of usage frequency, conceptual understanding, purposes of use, and family context. Although students in Türkiye and England reported broadly similar levels of GenAI use, differences emerged in how they conceptualised and used these technologies. The findings indicate an important gap between children's emerging use of GenAI and their understanding of how these systems work.

The study also highlights the importance of family environments in shaping children's early AI experiences. AI literacy initiatives should therefore move beyond technical definitions and involve both schools and families in supporting children's responsible and informed use of GenAI. Future research should examine children's GenAI engagement through mixed-method designs, more balanced national samples, and multidimensional AI literacy frameworks that capture not only use, but also understanding, critical evaluation, and ethical awareness.

Ethics Declaration

The confidentiality and anonymity of all participants were strictly maintained. All collected data were stored securely and used solely for research purposes. No identifiable personal information was disclosed.

AI Declaration

This paper was prepared with the assistance of Chatgpt-4.0, which was employed to proofread, edit, and refine the text, enhancing clarity and coherence. The AI tool was used as a supplementary editorial aid, complementing the human editorial process by refining language while preserving the author's original analysis and intent.

References

Andries, V., & Robertson, J. (2023). Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence, 5*, 100176.

Bauer, E., Richters, C., Pickal, A. J., Klippert, M., Sailer, M., & Stadler, M. (2025). Effects of AI-generated adaptive feedback on statistical skills and interest in statistics: A field experiment in higher education. *British Journal of Educational Technology, 56*(5), 1735-1757.

Beneteau, E., Boone, A., Wu, Y., Kientz, J. A., Yip, J., & Hiniker, A. (2020, April). Parenting with Alexa: exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Department for Education. (2023). *Generative artificial intelligence (AI) in education*. https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education

Diliberti, M., Schwartz, H. L., Doan, S. Y., Shapiro, A. K., Rainey, L., & Lake, R. J. (2024). *Using artificial intelligence tools in K-12 classrooms*. Santa Monica, CA: Rand.

Druga, S., Christoph, F. L., & Ko, A. J. (2022, April). Family as a Third Space for AI Literacies: How do children and parents learn about AI together?. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-17).

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate–inanimate distinction as examples. Cognitive Science, 14(1), 79–106.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-millionword gap has serious consequences. Child Development, 90(3), 985–992.

Mansfield, K. L., Ghai, S., Hakman, T., Ballou, N., Vuorre, M., & Przybylski, A. K. (2025). From social media to artificial intelligence: improving research on digital harms in youth. *The Lancet Child & Adolescent Health, 9*(3), 194-204.

Nikken, P., & Jansz, J. (2014). Developing scales to measure parental mediation of young children's internet use. *Learning, Media and technology*, *39*(2), 250-266.

OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative artificial intelligence in education*. OECD Publishing.

Samuelsson, R. (2023). A shape of play to come: Exploring children's play and imaginaries with robots and AI. *Computers and Education: Artificial Intelligence, 5*, 100173.

T.C. Millî Eğitim Bakanlığı. (2026). *Eğitimde yapay zekâ uygulamaları etik kılavuzu*. https://ttkb.meb.gov.tr/meb_iys_dosyalar/2026_01/09152326_egitimdeyapayzekauygulamalari_etik_kilavuzu.pdf

Xu, Y. (2023). Talking with machines: Can conversational technologies serve as children's social partners? Child Development Perspectives, 17(1), 53–58.

Zhao, X., Cox, A., & Chen, X. (2025). The use of generative AI by students with disabilities in higher education. The Internet and Higher Education, 66, 101014. https://doi.org/10.1016/j.iheduc.2025.101014

Zhou, X., Li, Y., Chai, C. S., & Chiu, T. K. (2025). Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interactive Learning Environments, 33*(10), 5766-5788.

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications.

Digital Transformation Where it Matters: A Four-Pillar Maturity Model for Diverse Assessment Practices in Higher Education

Thashmee Karunaratne, Fredrik Enoksson, Richard Lee Davis, Anna Land and Caroline Bern
Digital Learning Unit, KTH Royal Institute of Technology, Stockholm, Sweden

thashmee@kth.se
fen@kth.se
rldavis@kth.se
anland@kth.se
cbern@kth.se

Abstract: Digital transformation in higher education has been widely conceptualised through institutional frameworks such as those developed by EDUCAUSE and Jisc. While these offer robust lenses for analysing organisation-wide change, assessment transformation within disciplines is an ongoing process that requires continuous measurement and validation at a meso level. Assessment, as a core pedagogical and regulatory function, is closely linked to institutional structures but enacted primarily at the course and teacher levels. Consequently, nuanced transformations in assessment practices may not be fully captured by high-level organisational frameworks. This study addresses the gap by introducing the Digital Assessment Maturity Index (DAMI), a framework for analysing assessment transformation within higher education. The model was developed through an iterative abductive process that combined institutional digital transformation frameworks, contemporary assessment literature, and a contextual analysis of assessment practices at a technical university in Sweden. Rather than adopting organisation-wide dimensions, the study selectively synthesised concepts relevant to the local enactment of digital assessment at course and programme levels. This process identified four interrelated constructs: (1) Adoption & Coverage; (2) Efficiency & Throughput; (3) Authenticity & Learning Alignment; and (4) Trust & Reliability. Institution-wide dimensions such as governance, finance, and policy alignment aspects were excluded because they operate beyond the assessment enactment. The resulting four-pillar DAMI model was operationalised using an institution-wide survey of teachers at the bachelor's and master's levels, enabling an initial empirical mapping of the university's digital assessment maturity. The findings from 314 responses from faculty demonstrate that digital assessment at the university in 2024 is characterised by substantial adoption and efficiency gains, alongside persistent trust and reliability concerns and uneven pedagogical transformation. The model also demonstrates sensitivity to both barriers and enablers of assessment transformation, offering actionable insights to inform programme-level improvements. These outcomes reveal that the DAMI model can isolate assessment as a distinct unit of analysis and derive its dimensions through cross-framework conceptual reduction rather than institutional abstraction. As a next step, DAMI will be evaluated longitudinally through repeated institution-wide surveys.

Keywords: Digital Assessment Maturity, Higher Education, Education Transformation, Assessment Practices

1. Introduction

Higher Education institutions increasingly prioritise Digital transformation (DT) in response to evolving demands from education stakeholders for more innovative and improved education, in accordance with Sustainable Development Goal 4 (Singun, 2025). In first principles, "*Digital transformation is a series of deep and coordinated culture, workforce, and technology shifts that enable new educational and operating models and transform an institution's business model, strategic directions, and value proposition*" (Brown et al., 2020, p. 1). Within higher education, digital transformation extends beyond the adoption of new technologies and involves fundamental changes in how teaching, learning, assessment, and administration are designed and enacted (Rodríguez-Abitia & Bribiesca-Correa, 2021).

The complexity of DT, however, also demands systematic approaches for understanding, implementing, and evaluating institutional change across technological, pedagogical, organisational, and cultural dimensions (Brown et al., 2020; Rodríguez-Abitia & Bribiesca-Correa, 2021). Contemporary DT has been widely conceptualised through comprehensive institutional frameworks, such as the EDUCAUSE Dx Framework (EDUCAUSE, n.d.) and the Jisc Framework (Ahuja, 2025; Newman et al., 2025). These frameworks support universities in assessing digital maturity, identifying strategic priorities, coordinating transformation initiatives, and aligning technological development with institutional goals and educational practices (Alfirević et al., 2025; Brooks & McCormack, 2020; Newman et al., 2025). While providing robust lenses for analysing organisation-wide transformation, these frameworks primarily address institutional strategy, governance, infrastructure, and organisational capability at a macro level.

The digital assessment practices, however, operate as a continuous and locally enacted pedagogical process that evolves incrementally through course design, teacher practices, disciplinary requirements, and student interactions (Valdivia-Salazar et al., 2025). Although the assessment process is structurally shaped by institutional leadership, policies, and technological infrastructure, its enactment occurs primarily at programme, course, and teacher levels, where assessment types, design, pedagogical alignment, and perceptions of digital trust and authenticity are negotiated in practice (Sahni et al., 2024). Assessment transformation may involve course-level pedagogical redesign, changes in feedback processes, the adoption of authentic and technology-enhanced assessment formats, evolving trust between teachers and students in digital systems, and greater alignment of assessment methods with intended learning outcomes. These changes are often incremental, uneven, and highly contextual across disciplines and teaching practices. As Ahuja (2025, p. 156) notes, existing institutional DT frameworks, including Jisc, face "*challenges of contextualisation, lack of operational granularity, and limited evaluation evidence*" when applied to specific educational practices. This creates a need and space for a domain-specific, development-oriented framework capable of capturing nuanced changes in assessment practices, identifying barriers and enablers to transformation, and supporting longitudinal evaluation of digital assessment maturity.

This study systematically develops a Digital Assessment Maturity Index, a domain-specific framework grounded in the state-of-the-art DT frameworks and operationalised through a case study from a technical university in Sweden, guided by the following research questions: 1. which dimensions of DT are most relevant to the enactment of assessment practices in higher education, and 2. what patterns, tensions, and areas of uneven development emerge when applying the domain-specific framework to digital assessment practices within a technical university?

2. Background - Digital Transformation Maturity Frameworks

Evidence of the challenges higher education institutions (HEIs) encounter during DT is abundant. A recent systematic literature review (Singun, 2025) identified more than 80 transformation barriers across the areas of (i) Digital Vision, Strategy, and Policies; (ii) Digital Leadership and Management; (iii) Digital Organisation; (iv) Digital Resources; (v) Digital Competence; (vi) Digital Stakeholder Management; (vii) Digital Culture; (viii) Digital Academic; and (ix) Digital Ethics (Singun, 2025, p. 1). The article argues, based on these research outcomes, that many HEIs may feel inadequately prepared to adapt DT due to barriers to implementation (ibid). Inevitably, the ongoing DT within HEIs requires a collaborative and inclusive change process, reviewed and updated regularly (Rodríguez-Abitia & Bribiesca-Correa, 2021). When a systematic and regular approach (Roadmap) is required for any complex process, a strategic framework becomes necessary (Brown et al., 2020). Consequently, contemporary research provides evidence-based frameworks for measuring, monitoring, validating, evaluating, and conceptualising DT in HEIs, such as Jisc (Ahuja, 2025; McGill, 2023).

2.1 EDUCAUSE Framework

EDUCAUSE defines digital transformation (Dx) as "a series of deep and coordinated culture, workforce, and technology shifts that enable new educational and operating models" (Brooks & McCormack, 2020). This means Dx is not just a technological approach driven by the integration of tools, but a holistic institutional change in values, strategy, staffing, and capabilities. The framework provides a strategic, institution-wide lens, focusing on leadership, governance, data, technology, and cross-campus coordination, and offers guidance for the senior leadership on why Dx matters and what shifts must occur. However, because the framework operates at such a macro-organisational level, it offers limited insight into how digital transformation is enacted within specific pedagogical domains such as assessment, where transformation occurs through course-level practices, teacher decisions, and student experiences.

2.2 Jisc Framework for Digital Transformation

The Jisc DT Framework for Higher Education, predominantly verified with HEIs in the United Kingdom, is organised around six interconnected dimensions of Leadership and Governance, Culture, People, Technology, Practice and Processes (McGill, 2023). To this end, the framework provides a more practice-oriented perspective on digital transformation than EDUCAUSE, offering insights into the pedagogical, cultural, and operational factors that shape the effective use of digital technologies across the institution. Importantly, Jisc explicitly acknowledges that its conceptualisation of digital transformation is derived from the EDUCAUSE definition of Dx and adapted to the context of UK higher education (Newman et al., 2025). The two frameworks should therefore

be viewed as complementary rather than competing perspectives. While EDUCAUSE provides a strategic and organisational lens focused on leadership, governance, and institutional change, Jisc extends this perspective into the operational and pedagogical domains where digital transformation is experienced by staff and students. Together, they provide a comprehensive understanding of digital transformation, spanning both institutional strategy and educational practice.

3. Digital Transformation Maturity Model Synthesis

Existing digital transformation frameworks primarily address institutional change, offering limited support for analysing assessment transformation at programme and course levels. This study therefore developed a simpler framework to capture localised and incremental digital assessment change.

Using directed content analysis, with assessment practice as the analytical focus, the two frameworks were examined iteratively to identify and refine relevant dimensions. Hence, the process employed an abductive synthesis integrating DT frameworks and assessment transformation literature, and empirical evidence from teachers' assessment practices. As a process of sensemaking, abductive synthesis is particularly well-suited to addressing complex problems, providing a rigorous approach for integrating theoretical and empirical insights (Kolko, 2010, p. 17). Accordingly, the synthesis focused on the semantic meaning and functional role of framework components within assessment practice in contrast to their original component categorisation. Framework components were therefore reinterpreted through an assessment enactment lens and iteratively grouped according to the aspects of assessment transformation they represented.

3.1 The Synthesis Outcomes – the Four-Pillar DAMI Maturity Model

3.1.1 Adoption and Coverage

This dimension concerns the extent to which digital assessment practices are implemented and embedded across an institution. Within the EDUCAUSE Dx framework, widespread adoption is a requirement for digital transformation to enable shifts in educational and operational models. From this perspective, adoption reflects an institution's progression along the transformation journey, from isolated experimentation to routine, embedded practice. Measuring adoption, therefore, indicates how far an institution has progressed in its assessment-related transformation. Within the Jisc framework, adoption is closely associated with digital maturity and institutional benchmarking. Jisc emphasises the consistent integration of digital practices across departments, programmes, and professional roles. Consequently, adoption and coverage reflect the degree to which digital assessment has become a normalised component of everyday teaching and assessment practice.

3.1.2 Efficiency and Throughput

This reflects the extent to which digital assessment improves operational effectiveness through streamlined processes, reduced administrative burden, faster feedback cycles, and scalable assessment workflows. Within the EDUCAUSE framework, digital transformation involves developing workforce capabilities and redesigning organisational processes to improve efficiency and data-informed decision-making. From this perspective, efficiency serves as an indicator of whether digital transformation is delivering tangible improvements to assessment operations. Similarly, the Jisc framework emphasises process improvement, the reduction of organisational complexity, and the development of sustainable digital workflows. Efficient and evidence-informed practices are considered key indicators of digital maturity. Accordingly, this dimension captures the extent to which digital assessment processes are both operationally effective and sustainable in practice.

3.1.3 Authenticity and Learning Alignment

This dimension captures the extent to which digital assessment supports meaningful, practice-oriented assessment tasks and strengthens alignment between assessment design and intended learning processes. This dimension reflects both the authenticity of assessment tasks (i.e., the degree to which they resemble real-world or professional applications) and the extent to which assessment practices support iterative learning through feedback and structured engagement with course content. Within the EDUCAUSE framework, institutional change in higher education is closely linked to the redesign of teaching and assessment practices, particularly where digital technologies enable more flexible, student-centred, and outcomes-oriented learning environments. Similarly, the Jisc framework emphasises enhancement of teaching, learning, and assessment through digital technologies, with particular attention to assessment design, feedback processes, and alignment

between learning outcomes and evaluation methods. Accordingly, authenticity and learning alignment are treated as complementary indicators of pedagogical transformation, reflecting how digital assessment practices contribute to both real-world relevance and structured learning progression.

3.1.4 Trust and Reliability

This dimension captures teachers' and students' confidence in digital assessment technologies, processes, and assessment conditions. Within the EDUCAUSE framework, successful digital transformation depends upon secure, reliable, and future-ready technological infrastructure supported by effective data governance. Trust is therefore closely linked to institutional readiness, technological robustness, and organisational confidence in digital systems. Similarly, Jisc identifies reliable infrastructure, staff capability, institutional support, and a positive digital culture as foundational conditions for sustainable transformation. Its maturity model explicitly considers user confidence, support provision, and the reliability of digital systems. In the context of digital assessment, trust and reliability reflect the extent to which stakeholders feel confident using these technologies in authentic educational settings.

3.2 The Operational and Pedagogical Alignment of DAMI

The four dimensions of DAMI represent complementary perspectives on digital assessment transformation. Together, they provide a multidimensional view of assessment maturity spanning implementation, operational performance, pedagogical quality, and institutional confidence.

Adoption and Coverage capture the extent to which digital assessment practices are implemented and embedded across courses and programmes. It reflects the institutional reach of digital assessment and indicates how far transformation has progressed beyond isolated or experimental use.

Efficiency and Throughput reflect the operational performance of digital assessment, focusing on workflow efficiency, scalability, administrative burden, and feedback timeliness. It captures whether digitalisation translates into tangible improvements in assessment processes.

Authenticity and Learning Alignment represents the pedagogical quality of digital assessment, capturing the extent to which assessment designs support meaningful learning, align with learning outcomes, and enable authentic demonstrations of knowledge and skills. It reflects whether digital assessment contributes to educational improvement rather than procedural digitisation.

Trust and Reliability capture stakeholder confidence in digital assessment systems, including perceptions of technical reliability, system robustness, institutional support, and fairness of assessment processes. It reflects the conditions necessary for sustainable adoption and long-term transformation. This structure is simpler and practical yet enables moving beyond simple measures of technology adoption and captures the multifaceted nature of assessment transformation. Consequently, the framework provides a practical basis for monitoring digital assessment maturity, identifying areas of uneven development, and supporting evidence-informed improvement initiatives at programme and course levels.

4. Empirical Research Instruments

This research employed an empirical methodological approach grounded in a case study of a Swedish engineering university to operationalise the digital assessment maturity model. The data collection instrument was a semi-structured questionnaire, developed for multiple research studies on digital education, digital transformation, and technology penetration through teaching and learning in the engineering discipline. The survey items were designed as five (or seven)-level Likert-scale constructs for DAMI (cf. section 3.2), with digital assessment maturity treated as a multidimensional construct reflected in observable assessment practices, perceived benefits, pedagogical alignment, and trust in digital assessment environments. The questionnaire was administered to the faculty (teachers in bachelor's and master’s courses) at the HEI in the selected case study. The data were analysed descriptively using the R statistical analysis tool. The empirical analysis focused on identifying levels of digital assessment adoption, perceived operational and pedagogical impacts, barriers and enablers of transformation, and tensions between institutional digitalisation and local assessment enactment. The analysis captured unevenness across dimensions, including differences between operational digital maturity and pedagogical assessment transformation.

The standard ethical aspects of research data collection and analysis were followed in this study. The respondents were informed of the explicit use of data for research outputs, and a formal written consent for participation was obtained. However, no sensitive or personal data were collected in this study. The research process was designed following the ethical guidelines of the HEI where the data is collected.

5. Findings and Discussion

The resulting four-pillar DAMI model was embedded in the institution-wide survey of teachers at the bachelor's and master's levels, enabling an initial empirical mapping of the university's digital assessment maturity. Of the 314 respondents, 68 % were male, 27 % female. The median age was between 50 and 54 years, and 46 % of respondents were either professors or associate professors.

Descriptive and interpretive analyses on the survey data were intended to examine how the four DAMI dimensions manifested within teachers' perceptions of digital assessment practices. The analysis aimed to provide an initial empirical mapping of digital assessment transformation within the university context. The analytical focus was therefore placed on identifying patterns of transformation, examining tensions and contradictions across dimensions and exploring areas of uneven development in digital assessment practices.

5.1 Digital Assessment Adoption Across the HEI

Digital assessment adoption was measured through variables regarding teachers' use of digital platforms, proportion of course activities conducted digitally, whether teachers had created digital assessments last year, and use of digital quizzes and examination tools. Overall, 53.5% (n = 168) of respondents reported having created at least one digital assessment during 2024, while 46.5% (n = 146) reported no use of digital assessment. This suggests that digital assessment has achieved moderate institutional penetration but has not yet become a dominant assessment practice. From a digital maturity perspective, the findings indicate a transitional stage in which adoption has progressed beyond early uptake but remains uneven across the teaching population. \

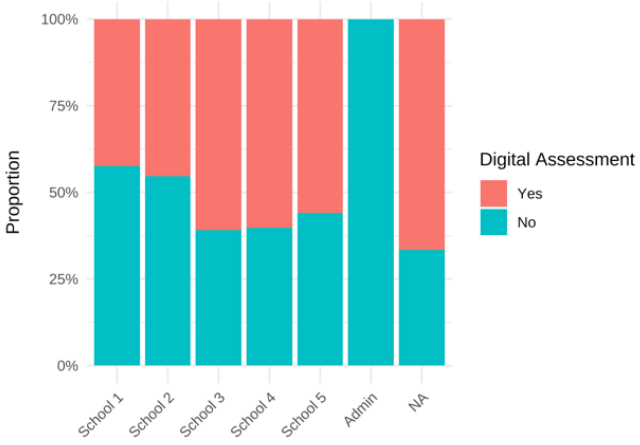


Figure 1: Digital Assessment Adoption by School

As Figure 1 shows, the differences in adoption proportion among schools are insignificant. But teachers' engagement with digital tools in the course activities and the number of activities they conducted digitally during 2024 were unevenly distributed, as shown in Figure 2 (Left). The median of digital course activities 25%, and Mean = 35%, SD = 28%, which indicates that while some teachers conduct a large proportion of their activities digitally, most report relatively modest levels of digital integration. Hence, digital adoption remains uneven across teaching contexts and has not yet become a uniformly embedded aspect of educational practice. As Figure 2 (right) shows, digital tool adoption is very high (Canvas ~97%, Zoom ~90%).

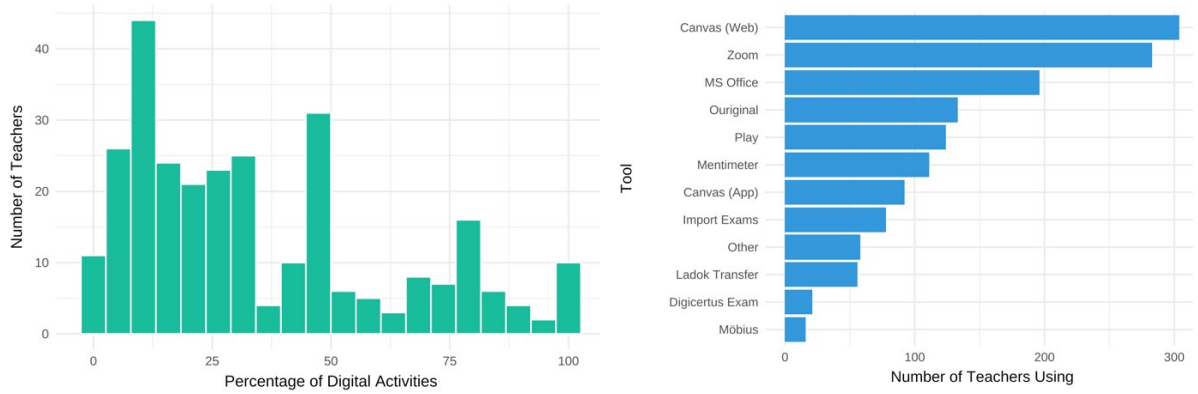


Figure 2: Percentage of digital activites (Left) and, Digital tools adoption (Right)

Considering that 53.5% created at least one digital assessment and the median digital activity level is 25%, it can be concluded that the adoption of digital assessment does not necessarily imply digitalisation of teaching activities. Many teachers just use digital assessment occasionally, not as a part of a broader digital pedagogical approach. Digital quizzes are a popular assessment method among teachers who conducted digital assessments, with 144 out of 168 conducted digital assessments showing that 86.7% of teachers who engage in digital assessment use digital quizzes. This indicates that once teachers adopt digital assessment, quiz-based assessment becomes highly prevalent.

The assessment technology ecosystem was dominated by Canvas, which was used by 96.8% of respondents. Adoption of specialised assessment technologies was considerably lower, including Ouriginal (42.4%), Import Exams (24.8%), Transfer to Ladok (17.8%), Digicertus (6.7%), and Möbius (5.1%). These findings suggest that while digital assessment infrastructure has achieved broad institutional coverage through a common platform, more advanced and specialised assessment technologies remain confined to relatively small groups of teachers. Consequently, assessment maturity appears to be characterised by widespread access and baseline adoption rather than diversified use of digital assessment tools.

The outcomes collectively indicate moderate digital assessment operationalisation concerning adoption and coverage. While over half of respondents reported creating at least one digital assessment, adoption has not yet become universal. Teachers reported relatively modest levels of digital activity overall (median = 25%), and assessment practices were concentrated around a small number of institutional platforms, particularly Canvas. The median teacher used only two assessment-related tools, suggesting that digital assessment is established but not yet characterised by extensive diversification or driven by the opportunities of deep technological integration.

5.2 Change to Practice: Efficiency and Throughput

The second DAMI dimension shifts the analytical focus from the extent of digital assessment adoption to the operational changes associated with its use. This dimension captures the degree to which digital assessment is perceived to support more efficient assessment processes through mechanisms such as timely feedback, resource savings, workflow optimisation, monitoring of student progress, and scalable assessment practices. The resulting distribution was strongly right-skewed (Min = 0, Q1 = 0, Median = 1, Mean = 1.77, Q3 = 3, Max = 9). This means more than one quarter of respondents reported no efficiency-related drivers for adopting digital assessment, while a smaller subset reported multiple efficiency-related benefits. The findings indicate that efficiency-oriented transformation is present within the institution but remains unevenly distributed across assessment practices. Most teachers reported only a limited number of efficiency-related characteristics, suggesting that digital assessment adoption is frequently associated with specific operational benefits rather than a comprehensive transformation of assessment workflows. Consequently, the Efficiency & Throughput dimension reflects a heterogeneous maturity profile in which operational gains are realised selectively rather than systematically across the teaching population.

5.3 Authenticity and Learning Alignment

In DAMI, "authenticity and learning alignment" is operationalised not as curriculum-level constructive alignment, but as the extent to which assessment practices are perceived to support authentic disciplinary tasks and learning-oriented processes such as feedback and iteration. Accordingly, the construct for this dimension is defined by the variables of teachers' perceived assessment accuracy/value, digital quizzes as a formative learning integration, and digital quizzes as an iterative learning (multiple attempts) process.

The most prevalent pattern of assessment and learning alignment is the use of digital quizzes enabling multiple attempts (39.2%), suggesting an emerging orientation toward iterative and mastery-based learning designs, where students are given repeated opportunities to improve performance. 32.9% of teachers reported use of quizzes as preparation or follow-up of laboratory or applied learning activities, indicating that digital assessment is increasingly embedded in discipline-specific learning workflows, particularly in practice-oriented teaching contexts. This reflects a partial integration of digital assessment into course structures rather than isolated use. However, only 16.9% of teachers perceive that digital assessment more accurately evaluates student performance compared to traditional paper-based methods. This weak perception suggests that digital assessment has not yet achieved strong epistemic legitimacy among most teachers.

Overall, the distribution of responses shows that assessment–learning alignment in digital assessment is uneven and selectively adopted rather than systemically embedded. While formative and applied uses of digital assessment are present, particularly through iterative quiz design and lab integration, these practices have not yet translated into a broader shift in perceived assessment validity or unified pedagogical transformation. These findings position Pillar 3 as reflecting an emergent but fragmented form of learning-oriented assessment transformation, where digital tools are primarily used to support learning processes rather than to fundamentally redefine assessment quality or standards.

5.4 Trust and Reliability

The trust risk perception reflects teachers' confidence in the functional stability, pedagogical safety, and integrity of digital assessment systems. It captures both trust in the institutional technology ecosystem and perceived risks related to technical failures, cheating opportunities, and system unreliability.

5.4.1 Distribution of Trust in Digital Assessment Tools

Trust was operationalised using the Likert-scale item *"My overall impression of the currently available digital tools for examinations at the university is positive"*, capturing general evaluative attitude toward institutional digital assessment tools (1 = strongly disagree, 7 = strongly agree, 0 = no opinion). The distribution of normalised trust in digital assessment (0–1 scale) indicates a generally positive but heterogeneous perception of digital assessment tools among teachers. The density plot reveals a concentration of responses in the upper-mid to high trust range, with prominent peaks around 0.70–0.85. A secondary cluster is observed around moderate trust levels (approximately 0.50–0.60), while very low trust values are comparatively low, suggesting that trust in digital assessment tools is not uniformly distributed but instead exhibits a right-skewed clustering structure, as shown in Figure 3.

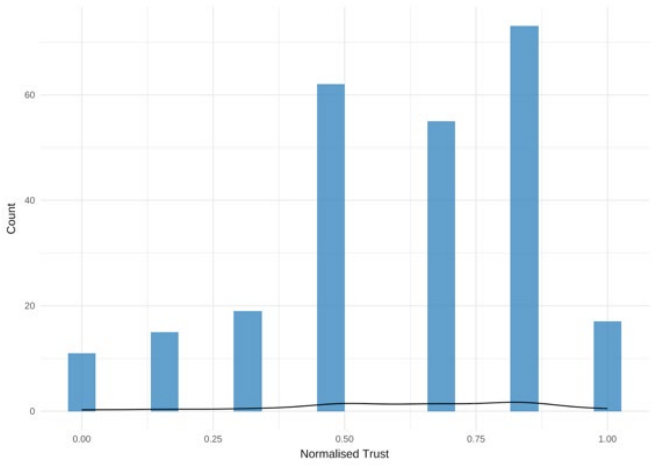


Figure 3: Distribution of trust in digital assessment tools

5.4.2 Distribution of Perceived Risk

Technical risk perception was measured using the binary items *"I'm unsure if digital assessments will function as expected"* and *"I'm worried about technical issues that could arise during the exam"*, capturing concerns related to system reliability and operational stability during digital examinations (0/1 presence of concern) as shown in Figure 4. Accordingly, the risk perception index (0–1 scale) shows a significantly different structure. The distribution is strongly concentrated at the lower end of the scale, with the highest density observed at values close to zero. Smaller secondary concentrations appear around 0.2–0.3 and 0.4–0.5, while high-risk perceptions are rare. This indicates that perceived risks related to digital assessment, such as technical reliability, assessment integrity, and system functionality, are generally weakly expressed across the sample, with a dominant configuration of low perceived risk.

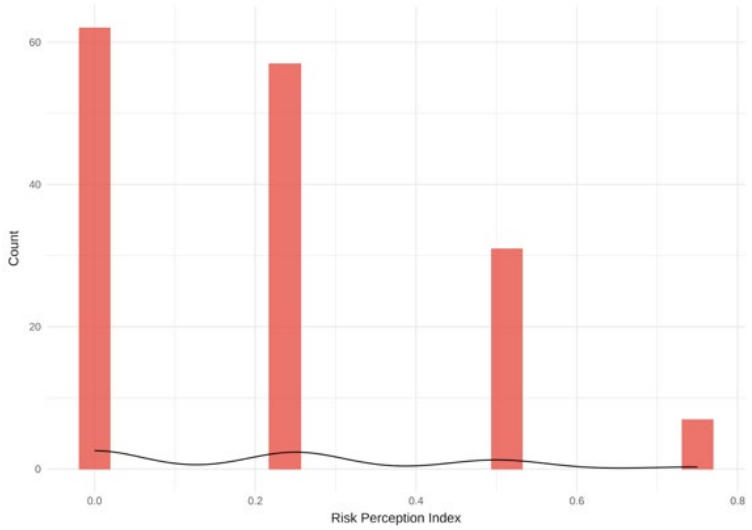


Figure 4: Distribution of Risk Perception Index

5.4.3 Trust vs. Risk

The joint distribution of trust and risk reveals a structured and highly clustered configurational space rather than a linear relationship between the two constructs. As Figure 5 (left) shows, density estimation and two-dimensional bins indicate a dominant concentration of observations in the region characterised by high trust (0.75–0.85) and low-to-moderate risk (0.20–0.30). A broader high-density zone extends across trust values ranging from approximately 0.50 to 0.90, combined with risk values between 0.20 and 0.40. This reveals the presence of an extended "acceptance band" in which trust remains relatively stable while risk perceptions vary moderately. Importantly, the high-density region is not aligned along a diagonal gradient, indicating that trust and risk do not operate as simple inverse variables. Instead, they form a configurational structure with identifiable stable entities. Furthermore, the general distrust was measured using the item "I don't trust digital tools for assessments", capturing explicit scepticism toward digital assessment technologies, alongside the overall trust–risk composite structure derived from the index. The joint distribution of trust and risk (Figure 5 – right) was divided into four quadrants based on median splits of both variables. The results show a dominant concentration of respondents in the high trust / low risk quadrant, with substantially smaller proportions in the other configurations. The high trust / low risk quadrant represents the dominant regime of digital assessment perception. The high trust / high risk quadrant is present but substantially smaller, suggesting conditional acceptance with latent concerns. The low trust / low risk quadrant indicates limited engagement or indifference rather than active resistance. The low trust / high risk quadrant is marginal, indicating that strong scepticism is rare in the dataset. This configuration suggests that digital assessment at the institutional level is largely characterised by technology acceptance, where trust is relatively high, and risk perceptions are contained within.

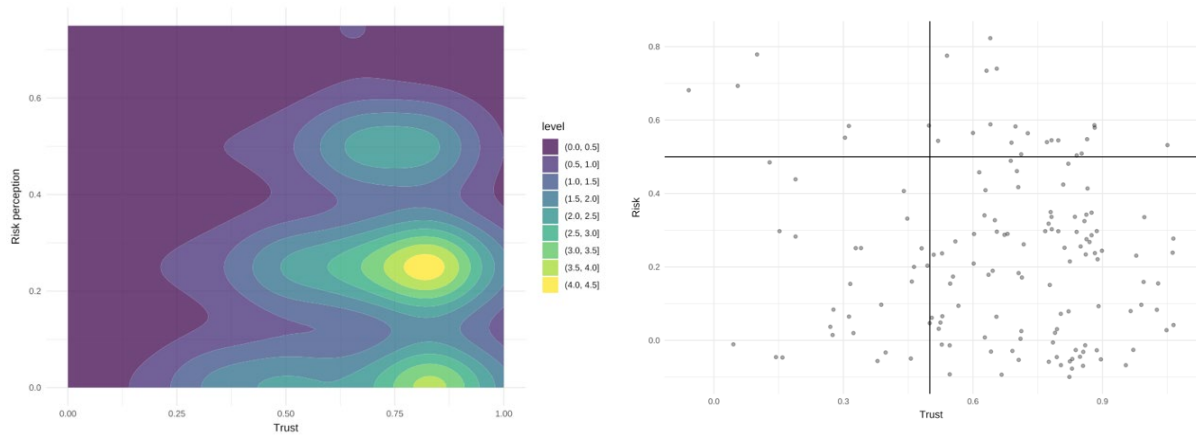


Figure 5: Trust–Risk Joint Density (Left) and Quadrant Distribution of Trust–Risk Configurations

Overall, the digital assessment landscape is characterised by broad adoption of foundational technologies, particularly learning management systems, while the use of more advanced assessment technologies remains selective and uneven. High trust and low perceived risk are associated with this established digital infrastructure, indicating a stabilised foundation for digital assessment. The widespread use of digital quizzes further illustrates that when technologies are perceived as useful, reliable, and pedagogically valuable, their integration into specific assessment practices can become routine and widespread.

6. Concluding Remarks

This study developed and applied the Digital Assessment Maturity Index (DAMI) to examine how digital transformation is enacted through assessment practices in higher education. The study was guided by two research questions: (1) which dimensions of digital transformation are most relevant to the enactment of assessment practices in higher education, and (2) what patterns, tensions, and areas of uneven development emerge when applying a domain-specific framework to digital assessment practices within an engineering university. The findings show that digital assessment transformation can be understood through four interrelated dimensions: Adoption & Coverage, Efficiency & Throughput, Authenticity & Learning Alignment, and Trust & Reliability. Together, these dimensions capture not only the extent of digital implementation, but also how digital assessment is operationalised, how it supports pedagogical practice, and how it is experienced in terms of institutional confidence. The results suggest that transformation cannot be reduced to adoption alone, but instead emerges through the interaction of operational, pedagogical, and perceptual dimensions.

The empirical analysis reveals clear patterns of uneven development across these dimensions. While adoption of digital assessment tools is relatively high, operational efficiency gains are selective and context-dependent rather than systematic. Similarly, authenticity and learning alignment are present but not fully consolidated, indicating an emerging but incomplete pedagogical transformation. Trust in digital assessment tools is generally positive, yet structured by heterogeneous perceptions of risk, particularly relating to technical reliability and academic integrity. Overall, the DAMI framework demonstrates its value as an analytical tool for capturing these multidimensional dynamics. It enables a more granular understanding of digital assessment transformation beyond adoption metrics, highlighting where practices are stabilised, where tensions persist, and where development remains incomplete in the context of engineering education.

Ethics Declaration

No personal data is collected for this research, and therefore, ethical clearance was not required for the research.

AI Declaration

AI tools were used for correcting language and grammar in writing the manuscript.

References

Ahuja, H. (2025). The JISC Digital Capability Framework: Development, Implementation, Evidence, and Future Directions. *Library Waves*, *11*(1), 147–162.

Alfirević, A. M., Vasilj, D., Vasić, D., & Alfirević, N. (2025). Digital Maturity of Academic Teaching and Learning, Digital Culture and Transformation: Empirical Evidence and Implications for Higher Education in Bosnia and Herzegovina and Croatia. *Information (Basel)*, *16*(12), 1084. https://doi.org/10.3390/info16121084

Brooks, C., & McCormack, M. (2020). *Driving Digital Transformation in Higher Education*. https://www.educause.edu/ecar/research-publications/driving-digital-transformation-in-higher-education/2020/defining-digital-transformation

Brown, M., Reinitz, B., & Wetzel, K. (2020). *Digital Transformation Signals: Is Your Institution on the Journey?* EDUCAUSE Review. https://er.educause.edu/blogs/2019/10/digital-transformation-signals-is-your-institution-on-the-journey

EDUCAUSE. (n.d.). *Dx: Digital Transformation of Higher Education*. EDUCAUSE. Retrieved May 5, 2026, from https://www.educause.edu/focus-areas-and-initiatives/digital-transformation

Kolko, J. (2010). Abductive Thinking and Sensemaking: The Drivers of Design Synthesis. *Design Issues*, *26*(1), 15–28. https://doi.org/10.1162/desi.2010.26.1.15

McGill, L. (2023, March 7). *Framework for digital transformation in higher education*. Jisc. https://beta.jisc.ac.uk/guides/framework-for-digital-transformation-in-higher-education

Newman, T., McGill, L., & Knight, S. (2025, February 5). *How to approach digital transformation in higher education: Report and case studies*. Jisc. https://beta.jisc.ac.uk/reports/how-to-approach-digital-transformation-in-higher-education

Rodríguez-Abitia, G., & Bribiesca-Correa, G. (2021). Assessing Digital Transformation in Universities. *Future Internet*, *13*(2). https://doi.org/10.3390/fi13020052

Sahni, S., Verma, S., & Kaurav, R. P. S. (2024). Understanding digital transformation challenges for online learning and teaching in higher education institutions: A review and research framework. *Benchmarking: An International Journal*, *32*(5), 1487–1521. https://doi.org/10.1108/BIJ-04-2022-0245

Singun, A. Jr. (2025). Unveiling the barriers to digital transformation in higher education institutions: A systematic literature review. *Discover Education*, *4*(1), 37. https://doi.org/10.1007/s44217-025-00430-9

Valdivia-Salazar, C., Serquén, O., Guevara, L., Bravo-Jaico, J., Alarcón, R., Germán, N., Aquino, J., Maquen-Niño, G. L. E., & Moreno Heredia, A. (2025). Development and validation of an instrument to assess the maturity of digital transformation in higher education institutions. *Frontiers in Education (Lausanne)*, *10*. https://doi.org/10.3389/feduc.2025.1691178

Ruwad Al-Mustaqbal System: A Knowledge-Behavior Gap-Aware Framework (KBG-Ruwad) for Personalized Responsible AI Education among Omani Undergraduates

Samah Alzidi, Aseel Abdul Nassor Ahmed Al Shizawi and Arunadevi Karuppasamy
Department of Computing Sciences, Gulf College, Muscat, Oman

2310149@gulfcollege.edu.om
2310179@gulfcollege.edu.om
arunadevi@gulfcollege.edu.om

Abstract: The use of generative AI among undergraduate students has surpassed structured instruction in responsible use of generative AI, and this gap is documented, with students' awareness of and knowledge of AI ethics not matching with their ability to apply it in academic contexts. This has been identified as a critical issue in Oman, with Vision 2040 stating AI adoption as a visionary goal, and most Omani HEIs lack formal AI governance, along with faculty AI literacy falling short of visionary goals. AI ethics are structured into progressive levels in competency frameworks like UNESCO's, and adaptive systems already personalize content. This paper proposes a conceptual framework, KBG-Ruwad, that introduces a novel measure of the knowledge-behavior gap (KBG). The framework assesses, for each student and across each dimension, the discrepancy between self-reported ethical knowledge and behavior demonstrated in scenario-based tasks. In addition, it measures students' level of ethical knowledge, their confidence in that knowledge, and the misconceptions they hold. The framework further extends the existing seven-dimensional model. Building on these measures, KBG-Ruwad extends the seven-dimension Responsible AI Competency Vector with discipline-specific, misconception-aware adaptive scenarios, and with a guidance function sensitive to gap, confidence, discipline and learning history; it further introduces a student-owned Learning Passport and a privacy-protected institutional dashboard. The paper reviews the AI-competency, AI-ethics pedagogy and Omani AI-readiness literature; argues for integration over invention; critically analyzes the components and flow of the proposed framework; and identifies gaps in the current specification and challenges for its operationalization

Keywords: AI Ethics Education, Knowledge-Behavior Gap, Awareness Profiling, Personalized Learning, Responsible AI, AI Competency Assessment

1. Introduction

While using generative AI tools in the undergraduate classroom has become commonplace, the use of ethical judgement in working with generative AI tools, what information can be provided to a generative AI tool, who can claim credit for the work created, and when is the work generated by the tool not the work of the student, is not taught as a distinct, practiced skill. In general, universities have responded with citation policies and brief orientation modules instead of ongoing teaching and learning and have not determined which students have which misconceptions or whether a student who can state a principle is using it. This deficit however, in Oman, does not happen in isolation but rather is a part of a specific national agenda. Digital transformation and AI adoption are strategic objectives of Oman Vision 2040, pursued in part through the country's workforce development objectives (Oman Vision 2040, 2019). The national AI and advanced-technologies programme, in turn, focuses explicitly on strengthening Oman's position on international measures of AI-readiness (MTCIT, 2024). Benayoune, Slimi and Al Habsi (2026) conducted a mixed-methods study on the 28 Omani higher education institutions, finding only 5.1% had a formal policy on AI and 74.4% did not, while 59% of the faculty surveyed were concerned about students relying heavily on generative tools. The results of the survey conducted among the 92 faculty members of a private Omani university revealed moderate awareness and usage of AI along with significantly higher acceptance of the technology with positive attitudes, which indicates that the gap between awareness and usage of the technology cannot be addressed by such general awareness campaigns (Natarajan et al., 2026). This paper presents a conceptual model, termed KBG-Ruwad, that measures this gap, dimension-by-dimension, and combines this measurement with confidence, discipline and learning history, to guide towards non-punitive, personalized advice by filling the gap. The elements of its individual components are not new to the literature, as sections 3 explain, but its novelty is the way in which they are integrated around a single measurable discrepancy. The related work is reviewed in Section 2, the gap and objectives are stated in Section 3; the framework is formalized in Sections 4 and 5, a critical, structural analysis of the work is provided in Section 6, Section 7-9 discuss ethics, significance and limitations, and Sections 10 concludes.

2. Literature Review

The power of AI-driven personalization has been proven. A global review of 25 studies found that adaptive content delivery, real-time feedback and progress tracking are already common features of AI-driven personalization (Fortuna et al., 2025). AI-powered virtual assistants are used in a similar way to enhance student engagement, and interest in this application has grown rapidly, although mainly from the perspective of educators rather than students (Sajja et al., 2024). UNESCO's AI competency framework for students already defines what students should know in four aspects: human-centred mindset, ethics, techniques and system design, organized as three development phases (UNESCO, Miao and Shiohira, 2024). Personalization and progression of competency levels are not new, as this literature focuses on the competency of subject-matter or general AI competency, rather than if the behavior correlates with the stated knowledge; and also as the technical and ethical competencies evolve as progressions of their own, and this could be measured separately from the learning of the content, as demonstrated in the early research on the foundation of K-12 education (Casal-Otero et al., 2023). While the studies above focus on the technology of content personalization, a separate strand of literature treats AI ethics as a pedagogical problem: it asks how ethical reasoning about AI use should be taught, rather than how content delivery can be automated. A 2025 EDUCAUSE framework provides an AI ethics foundation based on scenario analysis in real world settings instead of compliance (Georgieva and Stuart, 2025). A systematic review of AI-ethics interventions revealed that several case studies and group projects are already common, but the effects on behavior, as opposed to declarative knowledge, is inconsistent and often missing (Wiese et al., 2025). A review of 68 K-12 studies reached a similar conclusion, finding that ethics-based approaches also support AI-literacy development (Ma et al., 2025). A separate review of 34 studies found limited incorporation of cognitive, behavioral and attitudinal indicators, with most designs favoring dilemma-based tasks (Yan, Liu and Chau, 2025). A mixed-studies review similarly found that very few interventions used an ethics framework as a diagnostic tool for individual learners (Fu and Weng, 2024). Indeed, scenario-based AI-ethics teaching is already common, and the open question is how scenario performance should relate to an individual profile that can be updated, reinforced by a coordination-game analysis which shows that responsible norms rely on the expectations and assessment design of peers and that reflective tasks change norms over time (Ogbo et al., 2026). A research paper addressing AI's role in supporting Oman's knowledge-based economy discusses personalized and adaptive learning as an approach to aligning higher education with Vision 2040 (Sujee, Othmani and Dalwai, 2024). A review suggests that measures for reducing bias should be implemented on an institutional level, not solely within the classroom, and that it is necessary for institutional leadership to take responsibility for an institution's relationship to responsible adoption (Barnes and Hutson, 2024; Tarisayi, 2024). Skill development in general has been proposed to be adaptive and AI-driven, with a personalized-pathway approach (Akintola, Akintayo and Kadri, 2024), while studies underline both the promise and the academic-integrity concerns of generative AI in education have been documented (Adel, Ahsan and Davison, 2024). There are no studies that suggest there is a platform that would be specifically designed to identify and address the AI ethics knowledge and practice gaps that exist in the individual student population.

3. Research Gap and Objectives

Existing frameworks and adaptive learning systems for AI competency already establish several features, including AI-literacy assessment, awareness of ethics and privacy, academic-integrity instruction, case-based scenarios, and personalized recommendations and progress dashboards (Fortuna et al., 2025; Sajja et al., 2024; UNESCO, Miao and Shiohira, 2024). Similarly, AI-ethics research already draws on case studies and group reasoning rather than relying solely on declarative teaching (Wiese et al., 2025). What the literature does not depict is these features brought together to form a mechanism that compares the information a student reports he/she knows to behavior, dimension by dimension, confidence level by confidence level, and guides personalized, discipline-specific guidance from that comparison. While the UNESCO framework outlines knowledge expectations, it does not have a clear pathway on how to consistently measure such knowledge; while adaptive learning systems personalize content, they lack a focus on multidimensional AI-ethics competencies; and while the AI-ethics pedagogy documents scenario-based teaching, they rarely connect performance to an individual, updatable profile (UNESCO, Miao and Shiohira, 2024; Fortuna et al., 2025; Wiese et al., 2025; Yan, Liu and Chau, 2025). This gap is part of an Omani environment where the policy intent of Vision 2040 has been clearly stated and where there is a recognized lack of AI governance (Benayoune, Slimi and Al Habsi, 2026; Natarajan et al., 2026). KBG-Ruwad's main contribution is an integration of existing components, rather than a wholly new component, into a discipline-sensitive, non-punitive framework for undergraduate AI-ethics education in Oman. Within this framework, the knowledge-behavior gap is formalized as a measurable

construct, as presented in Section 4.2. The aim is to propose and conceptually evaluate a knowledge-behavior gap-aware framework for AI-ethics education for Omani undergraduates. The goals of the work are to develop a multidimensional competency model and a formalized measurement of the gap between knowledge and behavior; to develop discipline-specific, adaptive scenario and guidance mechanisms based on these gaps and on confidence and learning history; and to present a conceptual evaluation and phased implementation plan, including governance safeguards, that can take the model from proposal to a piloted system.

4. The KBG-Ruwad Conceptual Framework

4.1 Framework Dimensions and the Responsible AI Competency Vector

The framework sees AI-ethics competency as seven interrelated dimensions, rather than one score: AI literacy (L), even if it's just the basic understanding users have of the process AI tools use to create their output; ethical awareness (E), the knowledge of core ethical concepts, such as autonomy and informed consent; privacy awareness (P), understanding of privacy concepts and what data is safe to share with the AI tool; bias and fairness awareness (F), the ability to identify bias from AI output; academic-integrity awareness (A), knowing how to use AI tools to cite sources and acknowledge others' contributions, or how to disclose its use; critical evaluation (X), the habit of verifying AI output and not taking it as gospel; and responsible AI-use behavior (Bh), the general pattern of a student's decisions in AI-primitive use. These form a Responsible AI Competency Vector for student i:

$$R_i = [L_i, E_i, P_i, F_i, A_i, X_i, Bh_i]$$

These dimensions reflect the distinction that the K-12 AI-literacy literature has made between technical competency and ethical competency (Casal-Otero et al., 2023) and allow the platform to identify students who demonstrate high ethical awareness but low privacy awareness, a distinction that would not be captured by a single composite score.

4.2 Knowledge-Behavior Gap and Confidence-Aware Misconception Indicator

The framework splits a knowledge score $K_{i,d}$, from the questionnaire and a behavior score $B_{i,d}$, from the choices of scenarios , and defines the knowledge-behavior gap:

$$KBG_{i,d} = K_{i,d} - B_{i,d}$$

The sign and size of $KBG_{i,d}$ distinguish four knowledge-behavior combinations. High knowledge with low behavior indicates an application gap: a student who can correctly describe the privacy principle in the questionnaire but still uploads confidential material in the scenario. These students are expected to benefit more from practice-oriented guidance than from additional theoretical content. Low knowledge with high behavior indicates intuitive but incomplete knowledge, which is worth reinforcing conceptually so that it generalizes beyond the scenarios already encountered. Low knowledge with low behavior is treated as the highest priority for support, since neither the concept nor the corresponding practice is yet in place. High knowledge with high behavior indicates that the two measures agree and that KBG i,d is close to zero; because this combination does not indicate a gap, it does not receive the same targeted guidance as the other three. The second combination, low knowledge with high behavior, is expected to be more common among students who already use AI daily, and a single composite score would not distinguish it from the first combination. The second profile (low knowledge and high behavior) is more common among students who already use AI every day, a single composite score is not enough to differentiate it from the first case. Each scenario is also presented with a question about the student's confidence $Cf_{i,d}$, normalized in [0,1] since he/she could get the response right by chance. This helps to create a misconception-risk indicator:

$$MRI_{i,d} = \left(1 - B_{i,d}\right) \times Cf_{i,d}$$

A high $MRI_{i,d}$ indicates that the student chose the wrong action while reporting high confidence in it, which signals a misconception the student is unlikely to correct without guidance, rather than a result of guessing or acknowledged uncertainty. KBG and MRI together provide two separate signals as opposed to one signal for the profiling in Section 5.3.

5. Framework Components and Conceptual Platform Architecture

The six-layer architecture is summarized in Figure 1, which contains a profiling and knowledge assessment layer, a discipline-specific, adaptive scenario layer which includes the representation of behavior and confidence, the core KBG or MRI computation, a layer of personalized student support combining multidimensional profiling, the guidance function, explanation-based feedback and the Learning Passport, and an institutional insight and refinement layer. The details of each block are found in the corresponding subsection of Section 5.

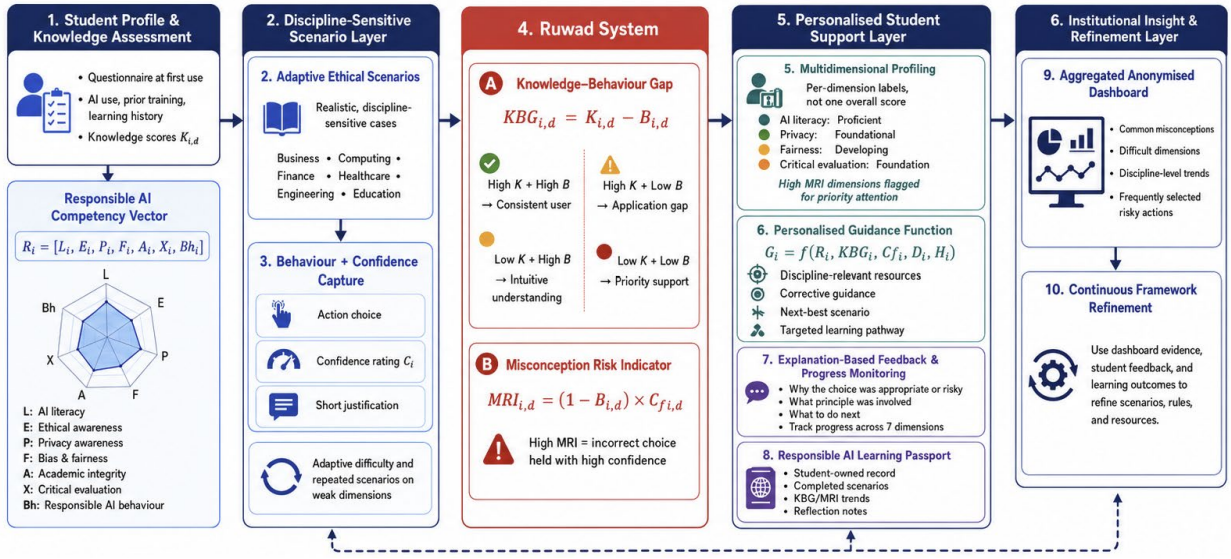


Figure 1. The architecture of the KBG-Ruwad system

5.1 Student Profile and Knowledge Assessment

An initial profile is created from the first- administration of a questionnaire, which asks about the frequency and purpose of using AI tools, understanding of AI hallucination, verification practices, citing AI sources, disclosure practices, processing institutional data, identification of biased outputs, and reliance on answers produced by AI, resulting in $K_{i,d}$ for each dimension. The instrument is conceptualized at this stage which is not used any actual student data but will be proposed to be pilot run and psychometrically validated in the future.

5.2 Discipline-Sensitive, Adaptive Ethical Scenarios

Students are presented with realistic scenarios covering uploading personal information, using AI-generated content without attribution, relying on inaccurate output without verification, acting on biased recommendations, using AI in assessed work, and relying too heavily on AI answers in written responses. Each theme is written as a set of variations, one for each of six sectors: computing, business, finance, healthcare, engineering, and education. Scenario selection is adaptive: if a student struggles on a dimension, they are given a further scenario on that dimension, at an easier or harder level depending on their performance. For each scenario, students choose the action they would take, rate their confidence in that decision, and provide a brief justification. The platform records the consequence of the chosen action, for example that sensitive data becomes exposed and the institution is put at risk, rather than only whether the answer was right or wrong. This branching, adaptive design follows a format that has already been shown to be effective for scenario-based instruction of ethical reasoning (Georgieva and Stuart, 2025; Wiese et al., 2025; Yan, Liu and Chau, 2025).

5.3 Behavioral, Confidence and Multidimensional Profiling

Profiling in this stage draws on a student's scenario choices to build a picture of their behavior, confidence and understanding across dimensions, and to indicate where each of these needs to improve. Scenario choices produce $B_{i,d}$ and $Cf_{i,d}$ for each dimension, from which $KBG_{i,d}$ and $MRI_{i,d}$. are calculated as described in Section 4.2. The labels Foundation, Developing or Proficient are then assigned to each dimension together, to create a multidimensional profile, e.g., AI literacy: Proficient, privacy: Foundation, fairness: Developing, academic integrity: Proficient, critical evaluation: Foundation. If a dimension has a high MRI, regardless of whether it is labeled Foundation/Developing/Proficient, the dimension is flagged for priority attention as a risk due to a

confidently held misconception in the dimension, not because it is labeled Developing/Proficient. This would lose the distinction that was the intent of grouping the seven dimensions in Section 4.1. These are still educational-support categories and not disciplinary designations; the complete profile is not sent to the teacher for the purpose of identifying form, and is not used for the purpose of marking.

5.4 Personalized Guidance and Recommendation Function

Guidance is recommended based on the student competency vector, gap profile, confidence profile, discipline and learning history:

$$G_i = f(R_i, KBG_i, Cf_i, D_i, H_i)$$

A student flagged with high MRI on privacy is prioritized for guidance on that dimension above dimensions the student has already acknowledged as uncertain but has not yet misconceived, and above discipline-relevant privacy dimensions already labeled Proficient.

5.5 Explanation-Based Feedback and Progress Monitoring

In addition to the chosen action, students are asked to submit a brief justification for the action, that will be assessed in a rubric format or by using the NLP model, moving the platform from being a multiple-choice system to an ethical reasoning assessment. Immediate feedback contains information about why a chosen action was right or wrong, which principle the action was based on, and what resources to move to next. Progress monitoring displays activities completed and areas that still need to be developed on the seven dimensions, as a process of self-development and not as a monitoring process.

5.6 Responsible AI Learning Passport

The Responsible AI Learning Passport is proposed to record a student's evolving profile, completed scenarios, dimension-level competencies, KBG and MRI trends, and self-authored reflection notes. The Passport is owned and controlled by the student: it is not visible to instructors by default, and, unlike the aggregated dashboard described below, it is not counted towards grading or used for disciplinary purposes.

5.7 Institutional Insight and Continuous Refinement

An authorized dashboard provides educators with aggregated, anonymized information on common misconceptions, the most challenging dimensions, differences between disciplines, the most frequently chosen risky actions, and changes in awareness over time. Individual-level detail is not routinely visible and remains subject to the safeguards described in Section 7. This evidence, in conjunction with student feedback and the pilot outcomes described in Section 9, is designed not to remain in place after deployment, but to be used to inform rounds of revision of the scenario library, guidance rules and content of the resources, thereby closing the loop in the sense of Figure 1.

6. Findings and Analysis

The findings are organized around four questions: what architecture is structurally built to expose, where it diverges from existing approaches and where gaps remain in its own specification, what practical challenges stand between the specification and a working instrument, and what these findings imply for the next stage of development. Table 1 sets the framework against the established work reviewed in Section 2. The extension in each row is architectural, not conceptual: none of the seven dimensions replaces an existing competency category, but the KBG and MRI measures in Section 4.2 quantify something none of these framework's measures alone, per-dimension discrepancy and confidence, rather than a single composite score. The analysis suggests that diagnosis is not incidental to the diagnostic framework; it is derived from the connections made by the two sections of the framework (Sections 4 and 5). The pipeline is coherent, every downstream component depends on the data it generates, and all components, except for the explanation-based one in Section 5.5, are implementable using methods already described in Section 2, and satisfy feasibility. Consider a business student with moderate privacy familiarity ($K_{i,privacy}$ = high). In the matching scenario, this student uploads a client dataset to a public AI tool to save time, a low-privacy action ($B_{i,privacy}$ = low), and rates their confidence in that choice as high ($Cf_{i,privacy}$ = high). Because $K_{i,privacy}$ is high while $B_{i,privacy}$ is low, the knowledge-behaviour gap ($KBG_{i,privacy}$) is large: the student knew better, so the failure is an application gap rather than a knowledge gap. Because the

action was taken with high confidence despite being wrong, the misconception risk index (MRI $_{i,privacy}$) is also elevated. The rest of this section examines whether that distinction holds up consistently in practice.

Table 1: Comparative Analysis of Existing AI Competency, Ethics, and Adaptive Learning Frameworks and the Proposed KBG-Ruwad Extension

Existing Framework	Existing Contribution	KBG-Ruwad Extension
UNESCO AI Competency Framework for Students	Combines ethics and technical competence in a progressive manner (UNESCO, Miao and Shiohira, 2024)	Incorporates a dimension of measured knowledge-behavior gap and confidence-weighted misconception detection.
AI-ethics pedagogy research	The use of case studies and as everyday practice the use of scenarios for teaching. (Wiese et al., 2025)	Rather than just an exercise, links scenario performance to an individual, updatable profile and adaptive scenario selection
Adaptive learning systems	Personalize content delivery and monitor progress (Fortuna et al., 2025; Sajja et al., 2024)	Focus on personalizing around competency vector and knowledge/behavior gap rather than overall subject progress
Academic-integrity guidance	Citation and disclosure norms for AI-assisted work	Embeds those norms within one of seven dimensions (confidence weighted)

Comparing the framework against Table 1 and against its own specification surfaces two groups of issues. The first concerns severity, not validity. If the confident/uncertain contrast were scored on a composite AI-ethics measure, the design argued for in earlier sections, it would fall into one of the "moderate" bands, and the same is true of the knowledge-only and behavior-only designs. So, while the framework distinguishes these cases correctly, it does not obviously rank them as more severe than a composite score would. The second group covers three assumptions that Section 6.1 relies on but does not itself examine. First, the seven dimensions in Section 4.1 were not derived empirically; they were argued for from the literature and from the problem statement in Section 1. Their coverage is assumed to be complete, and they are treated as independent of one another, neither claim is tested. Second, KBG $_{i,d}$ = K $_{i,d}$ – B $_{i,d}$ treats a given numeric gap as meaning the same thing on a common scale across dimensions and disciplines. Sections 5 establish only that the instruments share a similar convention at a given scale point, not that the gaps themselves are comparable in magnitude or meaning across dimensions. Third, the multiplicative form MRI $_{i,d}$ = (1 – B $_{i,d}$) × Cf $_{i,d}$ is consistent with the modelling approach in Section 4.2, but it is a design choice, not a form derived from or validated against a theory of misconception risk; a different functional form would weigh the same underlying data differently. Taken together, these results support the framework's diagnostic advantage over traditional methods, but they also show that its internal assumptions are untested by the comparative argument itself. They need to be examined separately, through how the framework performs in operation.

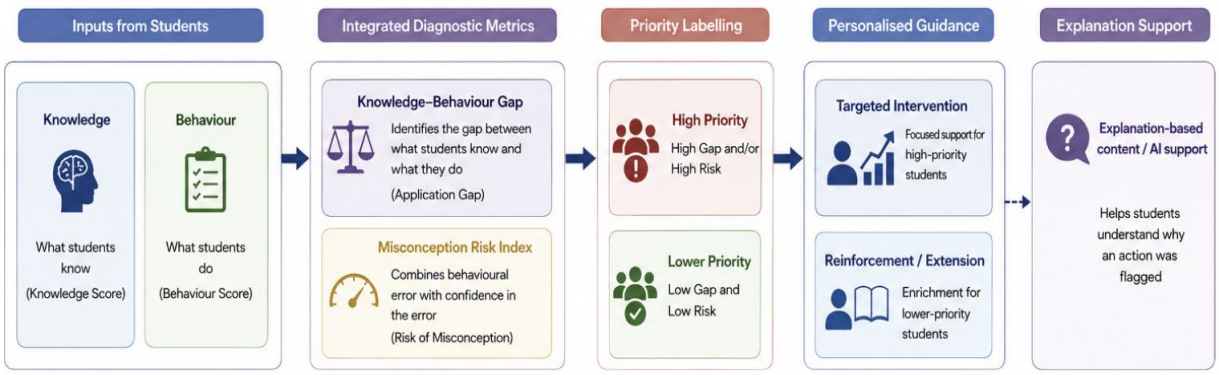


Figure 2: Findings from the Proposed Framework for Diagnostic Assessment and Personalized Student Guidance

In addition to the specification level gaps described earlier, this analysis suggests that there are some practical issues in making the framework a working instrument. It is not an easy task to write scenarios that are valid in the respective fields, without any scenario-writer bias as to what is the "responsible" action: The professional norms vary by field. A longstanding measurement challenge is to elicit reliable confidence ratings, where students may anchor on a fixed scale point, even if the rating is not a true reflection of their confidence. Adaptive difficulty, as suggested in Section 5.2, can become confused with under-competence if based solely on behavior, without taking confidence into account. The explanation-based module of section 5.5 relies on scoring unvalidated for this domain and therefore its reliability against human judgement is not presumed adequate. Creating the Learning Passport and dashboard to be non-punitive, then interpreting the flag as an indicator of risk is both an institutional and cultural challenge, as well as a technical one. These results show overall that operationalizing the framework poses challenges that are not covered by the gaps identified but are specific and additional. The gaps and problems highlighted above suggest five methodological priorities for the validation stage, beyond the general validation from expert review to pilot, described above. The score-scaling and equating procedures for both the $K_{i,d}$ and $B_{i,d}$ should be standardized and reported before any KBG value should be considered significant. The MRI formulation should be regarded as one of several and not the only formulation that should be tested prior to being accepted. Empirical pilot distributions should be used to determine KBG/MRI thresholds, and thresholds should be reported transparently and not determined a priori. A single subject-matter expert should not be used for more than one scenario in each discipline; cross-checks should be made to minimize single author bias. The explanation-based module needs to be tested with independent human raters before going into operation, and if reliability is not achieved, the guidance function does not require the use of the explanation-based module. Taken together, these findings suggest the institutional value put forward in Section 8 depends on addressing the gaps, the value that this framework provides at this point is the set of measures against which the addressing of the gaps earlier can be tested.

7. Ethical, Privacy and Responsible-AI Considerations

A platform profiling students' ethical understanding, confidence and misconception risk carries its own obligations, treated here as a first-order design requirement. Proposed safeguards include that informed consent is obtained before assessment, that data is minimized to items relevant to competency modelling, that data is anonymized before aggregate reporting, that access to identifying profiles and the Learning Passport is restricted (and should not be routinely visible to instructors), that assessment criteria are transparent, that guidance is supported by human oversight, that dashboards are interpreted by humans, that bias is monitored on an ongoing basis (as profiling that was tuned on one cohort may perform differently across linguistic backgrounds, disciplines or prior AI exposure), and that a student right to review recommendations is established, and that data is appropriately stored.

8. Conclusion

Students' use of AI has progressed faster than formal teaching on responsible AI use. Interventions are frequently generic and fail to pinpoint the student, discipline, and learning areas that need assistance. They also do not differentiate between students who do not know how to explain an ethical principle and students who know the ethical principle but are not able to make it a part of their practice. KBG-Ruwad tackles this by creating a clear diagnostic framework of the knowledge–behavior gap, adding confidence into a misconception-risk indicator, and offering discipline-specific and non-punitive guidance. The novelty of it is that the components are integrated in a systematic way around a measurable discrepancy between what the students know and what they do. The results and findings are articulate as a testable specification and state the measurement assumptions, gaps in implementation and challenges to be overcome in operation that require empirical assessment. The proposed order of expert review, prototyping, ethical approval and pilot testing is therefore proposed as a way of moving KBG-Ruwad from conceptual specification to evidence based validation. If the framework is validated, it can equip higher education institutions in Oman with a concrete guide to enhance and build the competence of responsible-AI and lay the groundwork for the workforce development goals of Oman Vision 2040.

Acknowledgment

The research leading to these results from the undergraduate research grant approved by Research and Innovation Authority (RIA) of the Sultanate of Oman under the Block Funding Programme (BFP), Agreement No. BFP/URG/ICT/25/390.

Ethics Declaration

This study proposes a conceptual framework and does not involve the collection, use, or analysis of primary or secondary data. No human participants, personal information, clinical records, or identifiable data were used. Therefore, institutional ethical approval was not required.

AI Declaration

AI-assisted tools were utilized only to improve linguistic quality and readability and creating the figure from my proposed work.

References

Adel, A., Ahsan, A. and Davison, C. (2024) "ChatGPT promises and challenges in education: computational and ethical perspectives", Education Sciences, 14(8), p.814.

Akintola, A.S., Akintayo, M. and Kadri, T. (2024) "Adaptive AI systems in education: real-time personalised learning pathways for skill development", Journal of Artificial Intelligence, Machine Learning and Data Science, 1, pp.2489-2494.

Barnes, E. and Hutson, J. (2024) "Navigating the ethical terrain of AI in higher education: strategies for mitigating bias and promoting fairness", Forum for Education Studies, 2(2).

Benayoune, A., Slimi, Z. and Al Habsi, A. (2026) "Artificial intelligence policy challenges and institutional readiness in Omani higher education", Discover Education, 5, p.86.

Casal-Otero, L., Catala, A., Fernandez-Morante, C., Taboada, M., Cebreiro, B. and Barro, S. (2023) "AI literacy in K-12: a systematic literature review", International Journal of STEM Education, 10(1), pp.1-17.

Fortuna, A., Prasetya, F., Samala, A.D., Rawas, S., Criollo-C, S., Kaya, D., Raihan, M., Andriani, W., Safitri, D. and Nabawi, R.A. (2025) "Artificial intelligence in personalized learning: a global systematic review of current advancements and shaping future opportunities", Social Sciences and Humanities Open, 12, p.102114.

Fu, Y. and Weng, Z. (2024) "Navigating the ethical terrain of AI in education: a systematic review on framing responsible human-centered AI practices", Computers and Education: Artificial Intelligence, 7, p.100306.

Furman, G.D. (2024) "Enhancing engineering education: the role of artificial intelligence in personalizing learning and outcomes", in Proceedings of the International Conference on Big Data Engineering and Education (BDEE). IEEE, pp.61-65.

Georgieva, M. and Stuart, J. (2025) "Ethics is the edge: the future of AI in higher education", EDUCAUSE Review, 24 June.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004) "Design science in information systems research", MIS Quarterly, 28(1), pp.75-105.

Ma, M., Ng, D.T.K., Liu, Z. and Wong, G.K.W. (2025) "Fostering responsible AI literacy: a systematic review of K-12 AI ethics education", Computers and Education: Artificial Intelligence.

MTCIT (2024) Oman AI and Digital Future Program (2024-2026). Muscat: Ministry of Transport, Communications and Information Technology, Sultanate of Oman.

Natarajan, J.R., Ayyappan, J.P., Sharma, K., Krishnasamy, T. and Seshan, V. (2026) "Faculty readiness for AI-driven digital transformation in Omani higher education: examining the attitudes-usage gap", Frontiers in Education, 11, p.1774840.

Ogbo, N.B., Song, Z., Ghareeb, S. and Han, T.A. (2026) Mathematical modelling of ethical AI use in higher education: a coordination game framework for future-facing learning. Preprint. Teesside University.

Oman Vision 2040 (2019) Oman Vision 2040. Muscat: Ministry of Energy and Minerals, Sultanate of Oman.

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D. and Demir, I. (2024) "Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education", Information, 15(10), p.596.

Sujee, S., Othmani, S.H. and Dalwai, T. (2024) "Toward a knowledge-based economy: empowering higher education in Oman with AI for Vision 2040 achievement", in Mansour, N. and Bujosa Vadell, L.M. (eds) Finance and Law in the Metaverse World. Cham: Springer, pp.133-142.

Tarisayi, K.S. (2024) "Strategic leadership for responsible artificial intelligence adoption in higher education", CTE Workshop Proceedings, 11, pp.4-14.

UNESCO, Miao, F. and Shiohira, K. (2024) AI Competency Framework for Students. Paris: UNESCO.

Wiese, L.J., Patil, I., Schiff, D.S. and Magana, A.J. (2025) "AI ethics education: a systematic literature review", Computers and Education: Artificial Intelligence, 8, p.100405.

Yan, Y., Liu, H. and Chau, T. (2025) "A systematic review of AI ethics in education: challenges, policy gaps, and future directions", Journal of Global Information Management, 33(1), pp.1-50.

Zawacki-Richter, O., Marin, V.I., Bond, M. and Gouverneur, F. (2019) "Systematic review of research on artificial intelligence applications in higher education - where are the educators?", International Journal of Educational Technology in Higher Education, 16(1), p.39.

Cracking the Code: Online Education for Modern Leadership

Amna Khaliq
University of Fredericton, Fredericton, Canada

amna.khaliq@ufred.ca

Abstract: This study, Cracking the Code: Online Education for Modern Leadership, examines professors' perceptions of the institutional, pedagogical, and technological challenges associated with online education in Canadian higher education. The literature review suggested that online education is a permanent and scalable approach that can expand exponentially with effective pedagogical methods and quality assurance. Strategic leaders may encourage the adoption of technological AI tools, advanced infrastructure, and training for students and professors. The research employed a phenomenological qualitative methodology, involving discussions with professors across Canadian institutions to address the two research questions. Based on findings, the primary research suggested that online education may help address significant issues for educational institutions by focusing on 1) data privacy, 2) infrastructure, 3) security, and 4) technology. Findings highlight the potential of a blended education system that combines the strengths of online and face-to-face learning, empowering both professors and students. Ultimately, this research contributes to the ongoing discussion of expanding online education in the Canadian education system, considering the needs and expectations of both students and professors, with a focus on business leaders.

Keywords: Blended learning, Educational leadership, Online education, Strategic leadership, Technology Adoption

1. Introduction

This qualitative phenomenological study investigates teachers' experiences of online education in Canadian higher education, with particular attention to pedagogical, technological, and institutional challenges. A sound research project begins with straightforward, uncomplicated ideas that are easy to understand and process (Creswell & Creswell, 2018). The two research questions addressed in this study are: How do online and face-to-face (FTF) strategies differ for educational leaders at higher education institutions in Canada? How does technology challenge online education at higher educational institutions in Canada? While previous studies have examined online learning broadly, limited research has explored professors' lived experiences regarding leadership, infrastructure, privacy, and technology challenges in Canadian higher education following the COVID-19 pandemic. This study seeks to address that gap through a phenomenological exploration of professors' perspectives.

1.1 Research Questions

This study addresses the following research questions:

> RQ1: How do online and in-person teaching approaches differ for Canadian higher education professors?
>
> RQ2: What technological challenges affect online education in the Canadian higher education system?

2. Literature Review

In essence, the study aims to understand the challenges faced by strategic leaders in Canada's online education system. It provides information on transformational theory within online education for higher education leaders, upper management, professors, and staff. This section reviews relevant prior literature on the online education system. The literature review offers a deeper understanding of the various challenges faced by strategic leaders in online education in Canada. It highlights the factors contributing to these challenges and provides effective strategies to overcome them. The section provides an overview of the strategic challenges and benefits of online education leadership, compares online learning with face-to-face education, and examines how leaders, professors, and students can effectively navigate online education. Additionally, it discusses the lack of pedagogical approaches for professors and emphasizes the importance of quality assurance in online learning.

2.1 Online Education Leadership's Strategic Challenges

Face-to-face education is a traditional learning system that has been used for centuries. This method is preferred for providing quick, on-the-spot feedback, assigning tasks promptly, respecting diverse learning styles, and fostering clear communication between professors and students. It allows professors and students to quickly and effectively understand more information and richer content by observing body language, interests, and

voice cues. The opportunity to connect with students, network, and collaborate as a team is more evident in face-to-face education (Fish & Snodgrass, 2019). Face-to-face learning encourages students to feel motivated and well-connected with their peers, promoting real-time interaction. They are less distracted and more focused because professors hold them accountable in real time for their projects. It also encourages critical thinking outside the box, as real-time class discussions are highly valued. Professors are also responsible and accountable for students' learning, as they receive timely feedback and can adjust the course content accordingly. This setting may provide a familiar environment for leaders, a comfort zone for leaders, professors, and students to engage in traditional learning methods (Stoian et al., 2022).

Strategic leaders in educational institutions initially did not focus on online education because they believed face-to-face education was superior and more conducive to learning. Their strategies, policies, and procedures may aim to increase gender equity, especially for women and people of colour in educational institutions, and to include refugees and international students from marginalized communities and developing countries (Bystydzienski et al., 2017). Developing this set of strategies, policies, procedures, processes, and rules is only possible when strategic leaders understand their target student demographics and students' needs. This is much easier in face-to-face education, as students are more visible to professors and staff. Their daily interactions help strategic leaders discover what is needed to develop a holistic and effective strategy that benefits everyone at the macro level within the institutions. Therefore, a strategy centered on diversity and inclusion has a positive impact. Educational institutions may strive to create a diverse and inclusive environment in both online and face-to-face modes of delivery, where everyone feels welcome (Kizzie, 2022). Upper management and professors are agents of cultural change, given their daily interactions with students. Meaningful change may occur when the needs and desires of the target market are clearly understood by leaders and upper management (Bystydzienski et al., 2017). Diversity, equity, and inclusion may be strengthened when gaps are identified, enabling us to find, understand, and address issues with practical solutions, and to implement realistic plans that bring everyone together with respect, kindness, and dedication (Khaliq, 2022).

However, strategic leaders can effectively collaborate with upper management, professors, staff, and students because they can quickly see the results of their policies and rules; the actions and outcomes are clear to all internal and external stakeholders. Engaging both internal and external stakeholders is essential for the success of educational institutions. Internal stakeholders include students, professors, staff, strategic leaders, the community, and the board of governors. External stakeholders encompass the government, partners, and customers. They play a significant role in institutional effectiveness and sustainability (IGI Global Disseminator of Knowledge, 2019).

2.1.1 Advantages of Face-To-Face Learning

The most significant advantage of face-to-face learning is the opportunity for synchronous engagement. Face-to-face classes occur at scheduled times, demonstrating the commitment of both professors and students. If neither party attempts to make lectures more engaging by using relatable, current information, students may not attend class or might use their cell phones or play video games on their laptops during lessons. Leaders can observe results in real time and adjust educational strategies accordingly, enabling timely improvements. They and professors can identify what works and what does not, creating social benefits through learning and applying knowledge to case studies and practical projects. Classroom training helps build personal relationships, confidence, engagement, and increased focus, while providing instant clarification of doubts and individual attention when needed (Feng et al., 2021).

Face-to-face learning supports personal leadership growth. Face-to-face environments may encourage greater engagement and accountability among participants. They foster a sense of community and offer opportunities to think differently or take risks that might not be available in online learning (Keane et al., 2022). Face-to-face learning keeps everyone connected, aligned, and informed. It may help reduce miscommunication and misunderstandings. It is faster and more efficient, and it fosters peer trust (Henny, 2016).

2.1.2 Comparing Online Learning to Face-to-Face Education

However, online learning cannot replace face-to-face learning activities, as they are distinct parts of the education system. They may serve complementary roles within higher education (Bhutoria, 2022). Students prefer face-to-face classes for math, general science, and laboratory work; online courses cannot fully replace face-to-face instruction, as they provide both theoretical and practical learning. This suggests that face-to-face learning will remain important for the next century, unless technology significantly transforms our mindsets and

lifestyles. “Interpersonal interaction remains an important component of collaborative learning and student engagement and face-to-face learning helps promote interaction and open communication. Students’ experiences within the learning community and their identities as students are shaped by teaching methods and technology (Nortvig et al., 2018).

2.2 Navigating Online Education: Leaders, Professors, and Students

The mission of higher education is evolving in tandem with the changing nature of the workplace. The reach, impact, and relevance of educational content must be updated to reflect a changing world. Some scholars and policy commentators argue that traditional degree structures are being challenged by alternative credentials, flexible learning pathways, and labour-market demands (Belsky, 2019). Teaching and learning are complex, and research has shown that factors beyond teaching style alone influence them. Teachers play an integral role in students’ mental and emotional growth, as they interact directly with them daily alongside their parents, siblings, and peers (Khaliq, 2021). The presence of professors in online settings, student interactions, teaching and content, the connection between online and offline activities, and campus- and practice-related activities define online education for a student (Nortvig et al., 2018). Dealing with the workload, teaching preparations, technological issues, and other pandemic-related challenges has exhausted professors, as they are expected to deliver higher-quality content than before (Power et al., 2022).

2.3 Lack of Pedagogical Approaches for Professors

Developing leadership in students is a vital part of higher education; however, the lack of effective pedagogical methods by experienced professors has made it more difficult for students to reach their full leadership potential in their chosen fields (Skalicky et al., 2020). Professors and leaders must understand the importance of delivering programs that develop the skills and capabilities employers expect in the post-pandemic world. Employees who add extra value at work are valued and have better opportunities for future growth because they are committed and willing to go the extra mile when needed (Armstrong Jr., 2020). Recognizing different leadership styles and their varied applications across diverse target audiences, such as students, could help close the gap between the demand for and supply of workers in the market (Skalicky et al., 2020). Leaders may focus on long-term growth in the emerging online education trend, given the large student market, the post-pandemic shortage of skilled workers, and ongoing technological advancements (Khaliq, 2018).

2.4 Quality Assurance in Online Learning

Leaders in online education may help ensure quality across online programs by leveraging technology to streamline operations. They may also regard degrees, diplomas, certificates, and short courses as valuable and high-quality, alongside traditional education programs (Bliss, 2019). Student satisfaction and the quality of education are directly linked, which the Perceived Session Quality Scale can measure in online synchronous learning. These findings suggest that designing a well-structured session flow, encouraging discussion and idea exchange among participants, providing evidence-based knowledge, maintaining up-to-date information, and using diverse online learning tools enhance the quality of online education (Czahajda & Černko, 2021). Students often multitask during online classes, either using their cell phones or attending to other chores while keeping their cameras turned off. Therefore, the implemented retrieval practice (RP) suggests that Students may retain learning content more effectively, even after 7 days, without multitasking. This may contribute to improved online learning experiences. In retrieval practice, students purposely recall information without using their cell phones; the fewer distractions, the more effective the learning (Khaliq, 2022). However, students may experience loneliness, lack of support, technical difficulties, weak internet connections, and low motivation. Nonetheless, online learning encourages students to develop technological literacy (Khaliq, 2018). The COVID-19 pandemic has sparked a global educational movement, compelling students and professors to utilize technology to continue teaching and learning (Khaliq, 2020).

3. Methodology

This qualitative phenomenological study examined professors’ experiences with online learning in Canadian higher education before, during, and after the COVID-19 pandemic. Conducted through Zoom interviews, the study explored pedagogical, technological, and institutional challenges associated with online education. The study included 18 professors from two Canadian institutions with experience teaching online: Cambria College in British Columbia and Fanshawe College in Ontario.

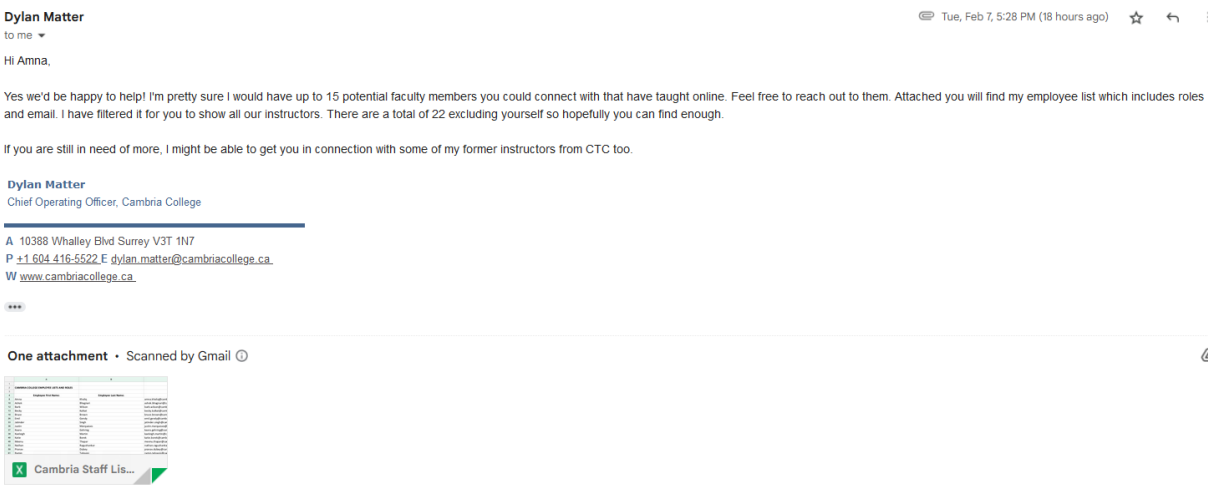


Figure 1: Permission E-mail from Cambria College

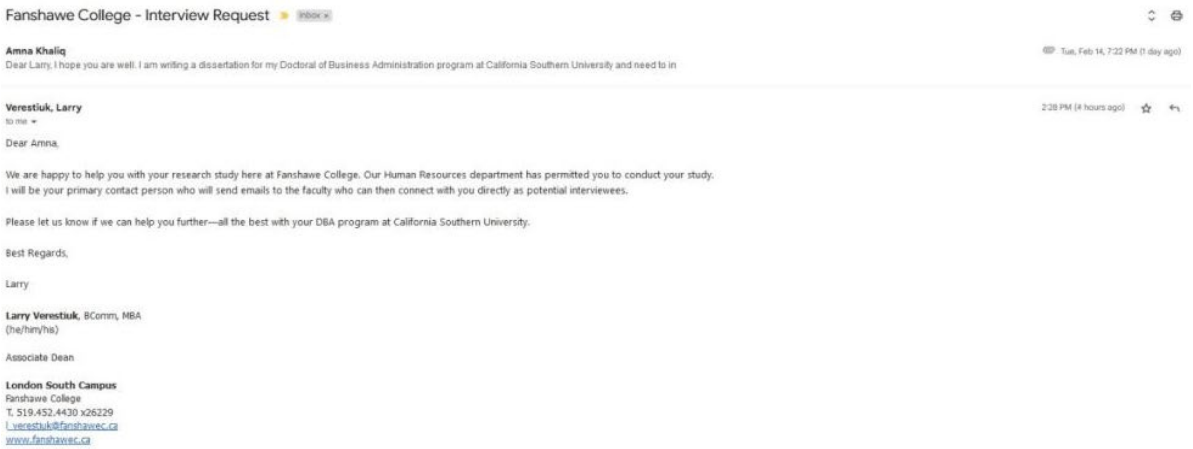


Figure 2: Permission E-mail from Fanshawe College

Participants were recruited through purposive sampling based on their experience with online education. Interview responses were analyzed using thematic analysis to identify recurring patterns related to pedagogical engagement, digital literacy, infrastructure, privacy, security, and blended learning. Themes were reviewed and refined to ensure alignment with the research questions. Participants received an invitation letter outlining the study's purpose, scope, data collection procedures, and ethical considerations prior to participation.

3.1 Qualitative Results

This study addresses the following research questions:

> RQ1: How do online and in-person teaching approaches differ for Canadian higher education professors?

3.1.1 Student Engagement and Classroom Interaction

All 18 participating professors reported that online teaching posed greater challenges than face-to-face teaching, particularly regarding student engagement, technological reliability, and classroom interaction. Students appeared more comfortable in face-to-face courses; however, many developed greater confidence in online learning following the COVID-19 pandemic. Professors indicated that online course content may need to be kept up to date and relevant to maintain student interest. Various teaching methods and strategies were used to encourage peer interaction and engagement in online classrooms. Professors also expressed interest in greater control over online learning management systems, including camera visibility during tests and monitoring of participation during lessons.

Professors suggested that students' personality types influenced their online learning experiences and participation levels. Extroverted students appeared more comfortable asking questions and interacting with peers. Participants emphasized the importance of creating a supportive, comfortable online environment in which students felt encouraged to raise concerns and participate without fear of judgment. Weekly virtual sessions and anonymous polls were identified as possible strategies to support student engagement and communication.

3.1.2 Digital Literacy and Technology Training

Professors indicated that digital literacy and technological advancement remain important components of online education. Students and professors were required to address technical challenges during virtual lessons, creating a need for updated technology training and support. Participants also noted the importance of ergonomic home office setups for online teaching and learning environments.

Furthermore, professors indicated that course content may benefit from regular updates, and learning objectives may need to reflect diverse learning methods and experiences throughout the course. International students, particularly those with limited prior exposure to digital learning environments, may benefit from introductory training in Microsoft Office, Zoom, and internet navigation before beginning online classes.

3.1.3 Multitasking and Participation Challenges

Professors reported challenges related to student multitasking during online lectures, including eating, completing household chores, or disengaging from class discussions while cameras remained off. Repeated interruptions and reduced participation were identified as concerns that affected the online classroom experience for both professors and students. Participants suggested that institutions may consider clearer policies regarding multitasking, participation, and academic integrity in online classrooms to support student attentiveness and engagement.

3.1.4 Blended Learning and Practical Skills

Professors strongly believed that online education may not always provide students with sufficient practical work experience and hands-on training. Participants suggested that virtual and augmented reality technologies may support innovative thinking and strengthen connections between education and evolving workplace expectations. Professors described using presentations, discussions, simulations, and games to assess students' theoretical knowledge alongside practical applications.

Participants also highlighted the importance of career preparation within online education. Professors may support students in developing resumes, cover letters, and interview skills to help prepare them for future employment opportunities. Demo interviews and experiential online activities were identified as potential approaches to strengthen practical learning.

RQ2: What technological challenges affect online education in the Canadian higher education system?

3.1.5 Technology, Accessibility, and Institutional Growth

Rapid technological advancement continues to reshape the delivery of online education and institutional operations within Canadian higher education. Participants indicated that technology may support student enrollment growth, institutional collaboration, and accessibility through partnerships with local and international institutions. Professors also noted that technology provides increased flexibility in course delivery and communication while supporting automation in grading, communication, and learning management processes. Additionally, online learning may create opportunities for students to work full-time while pursuing their education independently.

3.1.6 Digital Literacy and Technology Access

Professors reported that technology also creates challenges within online education. Students may require digital literacy skills to navigate online learning platforms and effectively adapt to new technological tools. Participants emphasized that unequal access to reliable internet services and advanced technological devices may contribute to digital inequality among students. Continuous technological training for professors was also identified as important for the effective use of online educational tools, although participants noted that such training may increase institutional costs.

3.1.7 Privacy, Security, and Online Safety

Participants identified privacy and cybersecurity concerns as ongoing challenges in online education. Viruses, hacking, and breaches of personal information were viewed as significant risks within educational institutions. Professors indicated that institutions may benefit from implementing robust backup and security systems to help address potential online threats and protect sensitive academic information.

3.1.8 Student Isolation and Institutional Competition

Professors also reported that online learning environments may contribute to student isolation and reduced motivation because peer interaction and support are less visible than in traditional classrooms. Additionally, participants noted that expanding online education may intensify competition among educational institutions seeking to attract students. Concerns were also raised regarding the authenticity and quality of some self-directed online certificates offered through external educational platforms. Participants emphasized that institutions may require reliable technological infrastructure to support the long-term sustainability and credibility of online education systems.

4. Recommendations

For the research paper "Cracking the Code: Online Education for Modern Leadership," the author believes that data privacy, infrastructure, security, and technology play a significant role in expanding the online education system in Canada.

1. **Data Privacy:** Online education collects a lot of information and data on students, including their personal and professional information, learning attitude, and grades. It requires consent and control by educational institutions, which can raise concerns among students about how to navigate complex privacy regulations. Therefore, educational sanctions indicated a prioritization of data privacy to prevent the compromise of sensitive information. They could only collect necessary data for academic purposes and refrain from storing sensitive information unless absolutely required.
    a. **Access Control:** Only authorized administrators and teaching professors may have access to students' information. Audit third-party edtech tools for compliance with privacy standards, such as PIPEDA in Canada, is suggested. Professors and students may receive cybersecurity training at the start of each semester to refresh. And it is important to obtain explicit student consent for third-party tools such as LMS, Zoom, Teams, and proctoring software. A data breach response team must notify students within 72 hours if the data is compromised.
2. **Infrastructure:** Education institutions may need to have a reliable infrastructure ready to adapt and expand online education. The internet bandwidth may need to be sufficient to handle the load when students are simultaneously logged in to their online LMS. Additionally, educational institutions may benefit from having extra devices, such as tablets, mobile phones, and wearables, so students can borrow them when their own devices are out of order, thereby avoiding disruptions to their education. They could require device encryption, antivirus software, and regular security patches for bring-your-own-device (BYOD) policies. AI technologies may be used to flag regularly occurring potential breaches. Backups may help protect against data loss, and cloud providers may need to be thoroughly vetted for their adherence to strong privacy and security standards.
3. **Security:** Educational institutions may only give access to students by authorizing them to take online classes and use their LMS system. A double-verification system may help support academic integrity in online learning. IT teams could ensure that multi-factor authentication (MFA) is implemented for students, staff, and professors. The systems are protected by strong passwords, anti-virus, anti-malware, and firewall protections that may be installed and regularly updated. Awareness of phishing and cyber threats may help strengthen institutional cybersecurity practices. Students' records and sensitive academic information may need to be protected from unauthorized access and hacking. The IT department may play a role in maintaining the confidentiality of lectures and discussions. Regular security audits may help ensure the IT environment remains safe and resilient.
4. **Technology:** Technology and online LMS create challenges and opportunities simultaneously for students and professors. However, Technology training for both professors and students may be important for effective online learning. This will ensure they are compatible with various online

learning methods, technologies, and devices, and know how to work effectively on the system when technological systems are unreliable, inaccessible, or insufficiently supported. Promoting digital skills may be considered a priority within online education systems. Students and professionals may receive training in various operating systems to create a seamless learning environment for all. All tools may help support data encryption and privacy settings. Integration between systems must be secure and seamless to help prevent glitches and other security issues. Institutions may benefit from adopting emerging technologies and providing accessible technical support to students, staff, and professors in online education to ensure their smooth and effective use within the education system.

4.1 Blended Learning

Based on these recommendations, blended learning is the most effective approach for both professors and students to explore in the online education system, as it leverages the strengths of both learning methods to create a personalized learning experience that utilizes technology effectively. It will help students and professors navigate the education tools more effectively. This may support the long-term sustainability and accessibility of Canadian higher education. Online learning emphasizes flexibility, accessibility, cost-effectiveness, increased engagement and feedback, customized learning patterns, and the effective use of technology. Given the recent changes and visa restrictions affecting international student enrollment in 2024, as well as the requirement to take a certain percentage of online courses, this policy may need to be reconsidered to create a fair educational system that minimizes discrimination based on gender, race, and socioeconomic status.

5. Limitations

This study was limited by its qualitative design and relatively small sample size of 18 professors, which may affect generalizability. Additionally, the findings reflect participants' perspectives within Canadian higher education and may not generalize to all institutional contexts.

6. Conclusion

Addressing these challenges in navigating the expanding online education for Canadian business leaders requires a support system from educational institutions to professors, staff, and students. This research paper compared the online and face-to-face strategies for professors in the higher education system. While face-to-face learning provided an interpersonal learning environment, it required physical presence. Online learning may expand accessibility and convenience by breaking geographical barriers for students.

However, navigating these challenges requires business leaders who must make online education easy by adapting to various strategic programs, such as expanding technological infrastructure, implementing security protocols and privacy regulations, and providing training for students and professors. This may contribute to a more inclusive and supportive digital learning environment. It will foster an engaging online education that promotes technological integration, meaningful virtual interaction, and online assessments within the digital landscape.

Education in Canada has continued to evolve since the COVID-19 pandemic, and it now requires modern leadership to adopt an online learning approach that combines the strengths of both online and face-to-face learning. The findings suggest that blended learning may offer a promising approach if institutions invest in pedagogical design, digital infrastructure, staff development, student support and robust privacy safeguards. Professors' development and student support programs will empower all stakeholders in the education system to effectively leverage the strengths of each strategy, benefiting students and expanding the higher education system in Canada.

Acknowledgement

I would like to express my deepest gratitude to my parents, Ami and Abu, and my sister for their unwavering love, support, and encouragement throughout my academic journey. I am also sincerely thankful to Dr. Marci Stone for her guidance and support. Their constant belief in me has been a source of strength and motivation in completing this research paper.

Ethics Declaration

Trust and confidentiality were maintained throughout this phenomenological qualitative research study involving professors and their respective educational institutions. The participants were all volunteers interested in the study's purpose. Some participants had the opportunity to review questions in advance to ensure they were comfortable speaking about the topic. The researcher followed all IRB recommendations before data collection. The consent forms were e-mailed to participants with the institution's permission. Institutional permissions were obtained from participating colleges prior to data collection (see Appendix A). The researcher made a conscious effort to minimize bias by setting aside her thoughts and beliefs to ensure bracketing of the data.

AI Declaration

This research paper was created with the help of Grammarly and ChatGPT software for word-smithing, language editing, and grammar. All analyses, interpretations, findings and final academic decisions remain the author's original work.

References

Armstrong Jr., N. (2020) *The socio-economic impact of the 2020 pandemic: A phenomenological exploration of nonessential business owners in Tennessee*. Doctoral dissertation, California Southern University. Available at: https://library.calsouthern.edu/writing?key_id=75

Belsky, L. (2019) 'Where online learning goes next', *Harvard Business Review*, 4 October. Available at: https://hbr.org/2019/10/where-online-learning-goes-next

Bhutoria, A. (2022) 'Dealing with online and blended education in modern, challenging times', *Decision*, 49(2), pp. 179–180. doi:10.1007/s40622-022-00323-y

Bliss, K. (2019) 'The future of education and technology', *eLearning Industry*. Available at: https://elearningindustry.com/future-of-education-and-technology

Bystydzienski, J., Thomas, N., Howe, S. and Desai, A. (2017) 'The leadership role of college deans and department chairs in academic culture change', *Studies in Higher Education*, 42(12), pp. 2301–2315. doi:10.1080/03075079.2016.1152464

Creswell, J.W. and Creswell, J.D. (2018) *Research design: qualitative, quantitative, and mixed methods approaches*. 5th edn. Sage.

Czahajda, R. and Černko, M. (2021) 'Perceived session quality scale: what contributes to the quality of synchronous online education?', *International Journal of Training & Development*, 25(3), pp. 217–243. doi:10.1111/ijtd.12216

Feng, X., Ioan, N. and Li, Y. (2021) 'Comparison of the effect of online teaching during COVID-19 and pre-pandemic traditional teaching in compulsory education', *Journal of Educational Research*, 114(4), pp. 307–316. doi:10.1080/00220671.2021.1930986

Fish, L.A. and Snodgrass, C.R. (2019) 'Age and gender and their influence on instructor perspectives of online versus face-to-face education at a Jesuit institution', *Journal of Education for Business*, 94(8), pp. 531–537. doi:10.1080/08832323.2019.1595499

Henny, C. (2016) '9 things that will shape the future of education: what will learning look like in 20 years?', *eLearning Industry*. Available at: https://elearningindustry.com/9-things-shape-future-of-education-learning-20-years

IGI Global Disseminator of Knowledge (2019) 'What is academic engagement and retention', 30 October. Available at: https://www.igi-global.com/dictionary/academic-engagement-and-retention/63553

Keane, T., Linden, T., Hernandez-Martinez, P. and Molnar, A. (2022) 'University students' experiences and reflections of technology in their transition to online learning during the global pandemic', *Education Sciences*, 12(7), p. 453. doi:10.3390/educsci12070453

Khaliq, A. (2018) 'Online learning: education conventions have changed', *The Dawson Creek Mirror*, 10 April. Available at: https://www.dawsoncreekmirror.ca/opinion/online-learning-education-conventions-have-changed-3498801

Khaliq, A. (2021) 'How informal networks of international Indian students affect the learning culture of Canadian institutions', paper presented at *International Perspectives on University Teaching and Learning*, 2–4 June. Available at: https://www.auburn.edu/academic/international/_assets/docs/iputl/essays/2021/khaliq-how-informal-networks-of-international-indian-students.pdf

Khaliq, A. (2022) 'Why do millennials prefer to move to smaller towns in Alberta?', paper presented at *Canadian Economic Association*, Ottawa, Canada, 31 May.

Kizzie, S. (2022) *The effects of Black/African American millennials' perceptions on the acceptance of mobile advertising*. Doctoral dissertation, California Southern University. Available at: https://library.calsouthern.edu/writing?key_id=75

Nortvig, A., Petersen, A.K. and Balle, S. (2018) 'A literature review of the factors influencing e-learning and blended learning in relation to learning outcome, student satisfaction and engagement', *Electronic Journal of E-Learning*, 16(1), pp. 46–55.

Power, A., Park, V., Owens, M. and Palapal Sy, M. (2022) 'Academics' experiences of online interprofessional education in response to COVID-19', *British Journal of Midwifery*, 30(4), pp. 222–228. doi:10.12968/bjom.2022.30.4.222

Skalicky, J., Warr Pedersen, K., van der Meer, J., Fuglsang, S., Dawson, P. and Stewart, S. (2020) 'A framework for developing and supporting student leadership in higher education', *Studies in Higher Education*, 45(1), pp. 100–116. doi:10.1080/03075079.2018.1522624

Stoian, C., Farcasiu, M., Dragomir, G. and Gherhes, V. (2022) 'Transition from online to face-to-face education after COVID-19: the benefits of online education from students' perspective', *Sustainability*, 14(19), pp. 1–18. Available at: https://ideas.repec.org/a/gam/jsusta/v14y2022i19p12812-d935858.html

Reflex, Refraction and Reflection: A Prism for Dialogue

Thomas Kjærgaard
University College North, Aalborg, Denmark

tmk@ucn.dk

Abstract: This paper develops and refines the concepts of reflex, reflection and refraction as a framework for analysing and designing reflective dialogue in higher education. The argument responds to a practical problem identified across prior studies of blended and in-class learning that students frequently participate in practical activities without moving beyond routine action unless the activity is deliberately slowed down and analysed. The paper therefore asks: how can educational activities be designed so that students have something to reflect on, a framework within which to reflect, someone to reflect with, and a purpose for reflection? Methodologically, the paper is based on a qualitative document analysis of a corpus of prior studies in Danish higher education that examined dialogue, reflection and lecturer scaffolding in hybrid learning settings. The analysis used iterative coding to identify recurrent forms of reflection, pedagogical arrangements and mediating conditions that supported reflective dialogue. Three cross-case conditions were identified: the sequencing of online and in-person dialogue, the use of structured dialogue formats such as rubrics and question prompts, and the use of technological affordances aligned with students' communicative needs. On this basis, the paper proposes a three-phase model in which reflex denotes undisturbed routine action, refraction denotes a didactic interruption that decomposes experience into analytically manageable elements, and reflection denotes the subsequent synthesis of experience and theory. The paper further develops a typology of hard, soft, diffuse and absorbed reflection to describe different degrees of lecturer involvement. The contribution of the paper is a clearer conceptual vocabulary and a design-oriented framework for linking practice, dialogue and theory in blended higher education.

Keywords: Reflection, Refraction, Dialogue, Blended Learning, Higher Education, Rubrics.

1. Introduction

Reflection and dialogue have long been treated as central to educational practice, from Socratic questioning to contemporary theories of reflective practice and dialogic learning (Padesky, 1993; Paul & Elder, 2007). Yet reflective intentions do not automatically become reflective activity. In higher education, students often complete tasks, placements and practical exercises efficiently, but they do not necessarily examine the assumptions, choices and theoretical implications embedded in what they are doing. This challenge becomes particularly visible in blended learning designs, where students move between independent online activity, peer interaction and classroom teaching. When reflective tasks are experienced merely as compliance requirements, the pedagogical ambition of linking practice and theory remains under-realised.

Prior studies from our research group (Hachmann et al., 2023; Kjaergaard & Hachmann, 2022; Kjærgaard & Dau, 2018; Kjærgaard & Andersen, 2023; Kjærgaard et al., 2025) identified this problem through recurring observations in in-class and blended-learning designs. The studies argue that students often move from practical action to more reflective dialogue only when their routine activity is interrupted by a purposeful pedagogical disturbance. This current paper preserves that core insight while reorganising the argument into a conventional academic structure. The central claim is that lecturers can design for reflective dialogue by creating moments of didactic deceleration in which complex activity is broken down into analytically manageable elements.

This paper asks the following research question:

> *How can educational activities be designed so that students have something to reflect on, a framework within which to reflect, someone to reflect with, and something to reflect for?*

To answer this question, the paper synthesises findings from a corpus of prior studies concerning dialogue and reflection in higher education and uses that synthesis to develop a conceptual framework for educational design.

The paper makes two contributions. First, it clarifies the relationship between reflex, refraction and reflection, thereby offering a vocabulary for describing what happens when routine action is transformed into analytical dialogue. Secondly, it identifies design conditions that support such transformation in blended and peer-learning environments. The paper proceeds with a review of relevant studies and theoretical foundations, followed by an account of the analytical procedure, the presentation of findings, and a discussion of implications for teaching and learning in higher education.

2. Literature Review and Theoretical Framework

2.1 Reflection, Self-Reflection and Dialogue

The paper adopts a broad understanding of reflection as analytical thinking oriented towards experience, action and understanding (Di Stefano et al., 2023; Dohn et al., 2018; Kjærgaard & Georgsen, 2021; Turns et al., 2015). Within that broad category, self-reflection is treated as a more specific form of reflection directed towards one's own role, judgement and possibilities for action. This distinction matters because learning designs may support reflection on an external object, reflection on one's practice, or reflection on oneself as a participant in practice. In higher education, these dimensions might often overlap, particularly in professional education where students continuously move between doing, interpreting and evaluating.

Thus, the prior studies indicate that questioning has a foundational role in reflective pedagogy. The Socratic tradition illustrates how clarifying, assumption-testing and implication-oriented questions can prompt learners to articulate and examine what would otherwise remain tacit (Padesky, 1993; Paul & Elder, 2007). In this paper, such questioning is not understood as a generic teaching technique but as a design principle, thus, reflective learning requires prompts that focus attention on selected features of experience and sustain inquiry rather than closure.

Dialogue is defined here as thinking together through language (Baucal, 2010; Kjærgaard, 2016; Littleton & Howe, 2010; Mercer et al., 2017; Sorensen & Kjærgaard, 2016). This definition distinguishes dialogue from forms of conversation centred on winning, defending or merely exchanging positions. In dialogic learning, participants listen, respond, reconsider and elaborate their understandings in relation to one another (Howe, 2010). Dialogue is in this context, therefore, both a medium and a method for reflection. It is the medium because reflective thought becomes tangible when expressed through language, and it is a method because carefully structured dialogue can guide learners towards the essential aspects of an experience, problem or text.

2.2 From Reflection to Refraction

The notion of refraction is adopted from Pagano and Roselle's use of the term to describe a process through which learning is broken down into intermediate processes that support critical thinking (Pagano & Roselle, 2009). They propose the following:

> *"We propose to move towards the concept of refraction through reflective practice. Refraction focuses on critical thinking skills and problem solving. The refraction process helps students understand and identify the intermediate processes of learning that aid the development of critical thinking skills" (Pagano & Roselle, 2009, p. 220)*

In the present framework, refraction names the pedagogical moment in which a lecturer, peer or mediating artefact slows down a learner's activity and separates it into constituent elements (Wagenschein, 2012). The metaphor of the prism is useful because it captures both interruption and analytic differentiation. Thus, it becomes a complex process and where the activity is made available for examination in parts.

This idea is compatible with Schön's account of reflection-in-action, where practitioners think within action rather than only after it (Schön, Donald A., 2017). However, the present paper specifies the pedagogical conditions that make such reflection more likely in educational settings. Whereas routine action may remain tacit and self-confirming, refraction introduces a deliberate disturbance that invites analysis. The disturbance can take the form of targeted lecturer questions, a rubric, a digital prompt, a peer-feedback protocol or a sequence in which online dialogue prepares later classroom analysis.

The framework also draws on Vygotskian perspectives on mediation and the role of the More Knowledgeable Other (MKO) (Vygotsky, Lev S., 2012). If dialogue is the primary medium through which reflection becomes socially available, then the lecturer's role is not only to ask questions but to mediate attention, sequence participation and help learners identify what is educationally significant. In this sense, refraction is a mediated and relational process rather than an individual cognitive event.

3. Methodology

3.1 Research Design

The paper uses qualitative document analysis to re-examine a corpus of prior studies on dialogue and reflection in higher education. The purpose of the analysis was not to aggregate effect sizes or evaluate interventions quantitatively, but to develop a coherent conceptual framework from a set of related studies that addressed comparable pedagogical concerns. Document analysis was selected because the source material consisted of published papers and associated analytical accounts that already contained empirical observations, design descriptions and interpretations relevant to the research question.

The corpus comprises prior studies undertaken in Danish higher education and reported across publications between 2016 and 2023 (Kjaergaard & Hachmann, 2022; Kjærgaard & Sorensen, 2016; Kjærgaard & Dau, 2018; Kjærgaard, 2019, 2020; Kjærgaard & Georgsen, 2021; Kjærgaard & Andersen, 2023; Rasmussen et al., 2022; Valbak-Andersen et al., 2021). Across the corpus, the studies examine dialogue and reflection in hybrid learning settings that combine online interaction with in-person teaching, and they share an explicit interest in how lecturers scaffold student dialogue and reflection through designed activities. The studies vary in immediate emphasis, covering online forums, inter-lesson reflections, peer-feedback rubrics and emergent learning networks, but they are connected by their common concern with mediated reflective practice.

3.2 Selection of Studies

The selection strategy was purposive rather than exhaustive. Studies were included when they met three criteria: first, they focused explicitly on dialogue, reflection or reflective practice in higher education; secondly, they described a designed pedagogical arrangement, such as a hybrid sequence, structured dialogue format or digital mediation; and thirdly, they contained interpretive material that made it possible to analyse how reflective processes were scaffolded.

This purposive selection was appropriate because the aim of the paper was conceptual refinement rather than comprehensive mapping of the field. The studies functioned as a theoretically meaningful corpus through which recurrent pedagogical mechanisms could be identified and compared.

3.2.1 Introduction to Selected Studies

The papers *"Reflection, Dialogue, and Hybrid Learning Spaces"* (Kjærgaard & Georgsen, 2021) and *"Inter-lesson Reflections: Learning through Interrogative and Reflective Dialogue"* (Kjærgaard & Andersen, 2023) investigated how a combination of online dialogues and in-class dialogues can scaffold deeper reflection. These studies concluded that preceding online dialogues effectively scaffold and enrich subsequent in-class dialogues. Furthermore, the archived online dialogue itself becomes an object for reflection, as students later use it as a springboard or prism for further discussion.

Another set of studies, *"Utilising Digital Technology for Dialogue and Evaluation"* (Kjærgaard, 2016) and *"Designing for Dialogue and Digitality in Higher and Continuing Education"* (Sorensen & Kjærgaard, 2016), explored how an asynchronous online discussion could be structured with specific didactic guidelines (such as preparatory readings, structured response formats, etc.) to support deeper dialogues reminiscent of classic scholastic conversation. These studies found that framed and directed dialogues, with clear structures and roles, serve as explicit objects of reflection, and that a well-defined framework for discussion can act as a prism that reveals the elements of the dialogue as discrete parts for examination. However, these particular studies focused on written, asynchronous academic dialogues and did not incorporate real-world practice elements, so they primarily highlight the potential of structured scholastic dialogue.

Additional studies, in particular *"Backchannels: Covert Digital Backchannels in the Overt Classroom"* (Kjaergaard & Hachmann, 2022) and *"Emerging Networks – Disturbed Learning Trajectories and New Ways of Connectivity at University Colleges"* (Hachmann et al., 2023), investigated dynamic learning networks that emerged from student-initiated communication in specific situations. These studies found that certain technological affordances (like online forums or backchannels) combined with an authentic student *need* for dialogue created excellent conditions for networked learning. Notably, the students' actual needs for understanding the topic and task in the context acted as an impetus for reflection, effectively functioning as a prism through which key elements were brought into focus.

Synthesising insights from these empirical studies identified three key conditions that seem to function as a prism that breaks the learning process into elements and thereby focusing students' reflection:

- Blending online and in-person dialogue: A combination of online dialogues (preceding a class session) and subsequent face-to-face dialogues provides scaffolding. The online discussion effectively refracts the learning experience into elements that the later in-class dialogue can analyse more deeply.
- Structured dialogue formats: Dialogues that are intentionally framed with clear guidelines or roles (for example, using rubrics or predetermined question prompts) create better conditions for students to explore and reflect on distinct elements of the learning process. The structure itself refracts the learning content into pieces that can be separately examined.
- Technology aligned with student needs: Thoughtfully chosen technological tools, aligned with students' authentic needs for communication or feedback, can highlight specific aspects of the content or experience, turning them into opportunities for reflection. The technology's affordances can act as a prism that draws attention to elements of the learning process which might otherwise remain implicit.

3.3 Coding and Theme Development

The analysis proceeded in four stages. First, the publications were read to identify passages describing reflective activity, student participation, lecturer interventions and the role of mediating artefacts or technologies. Secondly, these passages were coded descriptively for instances of reflection, self-reflection, dialogue, questioning, scaffolding, timing, technological mediation and the object of reflection. Thirdly, the coded material was compared across the selected studies in order to identify recurring pedagogical conditions that appeared to support reflective dialogue. Fourthly, these patterns were synthesised into broader themes and connected to the theoretical concepts of reflex, refraction and reflection.

Coding was iterative and comparative rather than strictly linear. Initial codes were refined as the analysis progressed, particularly where similar pedagogical mechanisms appeared under different labels across the studies. For example, online questioning, rubric-guided peer feedback and digital posting activities were initially coded separately, but later brought together under the broader analytical theme of structured refraction because each design decomposed experience into discussable elements. The analysis was informed by the iterative, data-responsive logic associated with grounded theory, although the study did not claim to generate formal grounded theory.

3.4 Analytical Procedure and Limitations

The analytical procedure was abductive, and the themes were derived from the source material and then interpreted in dialogue with theory. This meant that empirical regularities in the corpus were used to refine the conceptual framework, while concepts from Schön, Vygotsky, Pagano and Roselle, and dialogic learning research helped to specify the educational significance of the patterns identified. The outcome is therefore a practice-oriented theoretical framework grounded in prior studies rather than a report of newly generated primary data.

This methodological choice has limits. Because the analysis synthesises a bounded corpus of related studies, the findings should be understood as conceptually generative rather than universally representative. Moreover, the corpus is anchored in higher education contexts in Denmark, and the framework should therefore be transferred to other settings with appropriate contextual sensitivity.

4. Findings

4.1 Three Design Conditions that Function as Pedagogical Prisms

The cross-case analysis identified three recurring conditions that helped transform routine participation into reflective dialogue. Together, these conditions functioned as pedagogical prisms because they directed attention towards selected elements of a learning experience and made those elements available for discussion.

The first condition was the sequencing of online and in-person dialogue. In several studies, online discussion before a class session helped students externalise questions, observations and tentative understandings that could later be revisited and analysed in class. The archived online dialogue became an object of reflection in its

own right. Rather than treating online interaction as preparatory administration, these designs used it to create a first analytical layer that enriched subsequent face-to-face dialogue.

The second condition was the use of structured dialogue formats. When discussion was organised through explicit rubrics, prompts, role descriptions or questioning protocols (answers to posts should end in new questions), students were better able to identify what they were meant to examine and how they were expected to contribute. Structure did not close down reflection; instead, it provided a frame within which reflection could become more precise. Rubrics were particularly powerful because they named evaluative dimensions and thereby refracted a complex performance into discussable criteria.

The third condition concerned technology aligned with students' authentic communicative needs. Digital platforms such as discussion posts, Padlet boards and polling tools became productive when their affordances matched the communicative task at hand. In such cases, technology did not merely deliver content but mediated the reflective process by preserving traces of thinking, enabling comparison of contributions or foregrounding student questions. When the technology was meaningfully linked to the students' need to ask, respond, compare or prioritise, it supported more focused reflective activity.

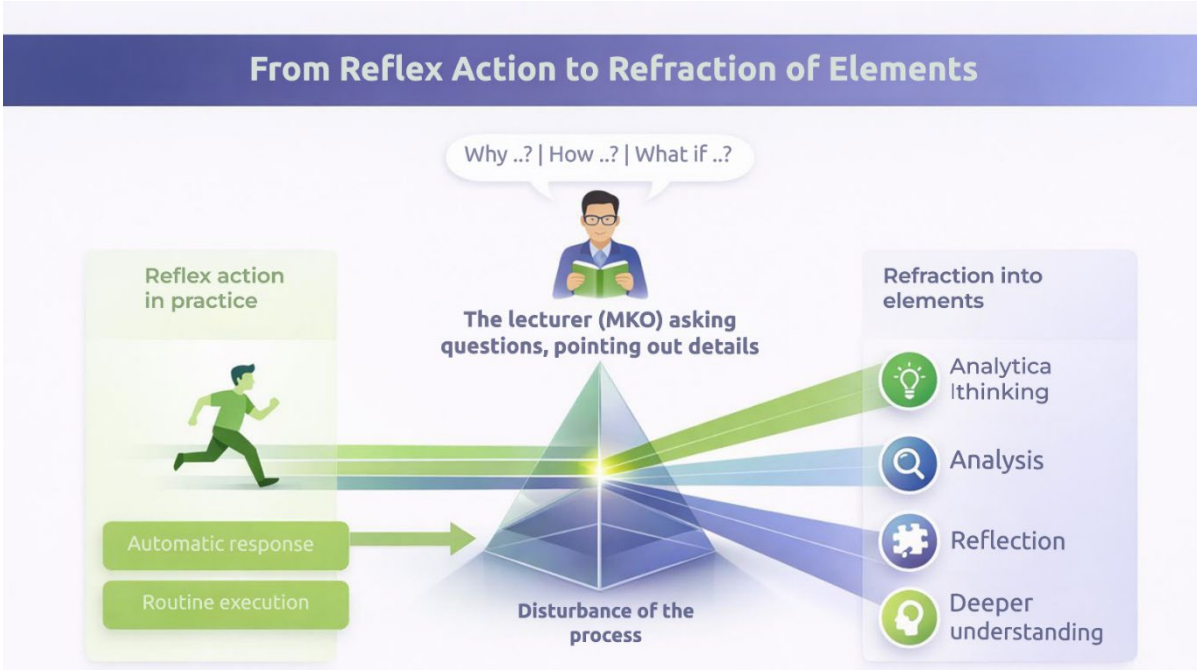


Figure 1: Conceptual model: from reflex action to refraction of elements (made with CoPilot).

4.2 The Reflex–Refraction–Reflection Framework

The thematic synthesis supports a three-phase framework. Reflex denotes routine, often efficient action in which the learner responds within practice without stepping back from the immediate demands of the situation. Such action is not deficient; indeed, professional work often depends upon routinised competence. The educational problem arises when routine action remains pedagogically opaque and learners do not examine what informs it.

Refraction denotes the intervening pedagogical act that slows down the process and separates it into elements. It is the moment of didactic deceleration in which a More Knowledgeable Other (MKO), such as a lecturer, peer or artefact introduces a question, category, pause or comparison that makes a previously tacit activity discussable (Vygotsky, Lev S., 2012; Vygotsky, Lev Semenovich & Cole, 1978). Refraction is therefore neither identical with reflex nor with reflection. It is the analytically generative turning point between them. In the refraction stage, the lecturer's task is to break the student's experience or expression (an utterance, a written text, a demonstration, etc.) into elemental parts, helping the student examine and understand each of these fundamental components and how they connect (as depicted in Figure 1). This approach is inspired by Wagenschein's *exemplary principle* (Wagenschein, 1951; Wagenschein & Horton-Kriiger, 2000), which advocates selecting teaching content that can be broken down into fundamental, representative elements.

Reflection denotes the subsequent process of analysis and synthesis through which the learner connects experience, concepts and possible future action. Reflection may occur during action, shortly after it or across a more extended period, but in each case, it depends on some means of revisiting and reorganising experience (Schön, Donald A., 2017). In the framework proposed here, refraction is the pedagogical mechanism that makes such revisiting more likely and more disciplined.

Table 1: Decelerations of learning process through the prism of refraction

Phase	Primary Learner Activity	Role of Lecturer or Mediating Artefact	Pedagogical Outcome
Reflex	Routine action in practice; tacit or automatic responding	Observation, task framing or background support without direct interruption	Continuation of activity without explicit analytical breakdown
Refraction	Attention shifts from acting to examining selected elements of the activity	Questions, prompts, rubrics, peer protocols or technologies interrupt and focus the process	Didactic deceleration; elements of the activity become discussable
Reflection	Analysis, comparison, conceptualisation and synthesis	Support for linking experience to theory, alternatives and implications	New understanding, reframed judgement and possible adaptation of practice

4.3 Types of Reflection

The analysis also indicated that reflection varies according to the timing and intensity of pedagogical support. To describe this variation, the paper retains the typology of hard, soft, diffuse and absorbed reflection, but presents it more explicitly as a design heuristic rather than as a claim about fixed learner types. The typology is inspired by Schöns work on the Reflective Practitioner (Schön, Donald, 1987; Schön, Donald A., 2017)

Hard reflection refers to immediate and systematic lecturer involvement. It occurs when questioning or feedback is closely coupled to the action itself, enabling reflection in action. Soft reflection refers to delayed but still intentional support, for instance when students formulate reflections independently and then return to them in a facilitated dialogue. Diffuse reflection occurs when learners are expected to reflect largely on their own, with limited or inconsistent guidance. Absorbed reflection designates situations in which no meaningful pedagogical support for reflection is present, so any reflective work depends entirely on the learner's own initiative.

Table 2: Reflection typology

Type of Reflection	Degree of Support	Typical Timing	Illustrative Educational Use
Hard reflection	High and direct	During or immediately within the activity	Live questioning, coached performance, tightly guided inquiry
Soft reflection	Moderate and indirect	Shortly after the activity	Inter-lesson writing, facilitated sharing, rubric-guided peer feedback
Diffuse reflection	Low or irregular	After activity or over time	Independent journalling with sparse follow-up
Absorbed reflection	Absent	No structured reflective timing	Task completion without dialogic or analytical support

4.4 Illustrative Examples from the Corpus

One group of studies used a questioning protocol in which students ended posts with further questions and responded to peers through additional questions. This design foregrounded inquiry itself and pushed students to unpack a topic progressively. In analytical terms, the question protocol worked as a mechanism of hard reflection because it required participants to interrogate the topic continuously and made the boundaries of the inquiry explicit.

A second example was the use of inter-lesson reflections. Students completed written reflections between meetings and then discussed these materials in the following class session. This design shifted part of the refractive work into the time between lessons. The written posts served as objects of reflection that could be revisited, compared and elaborated with peers. The process is best understood as soft reflection because the students engaged initially on their own but later returned to their entries in structured classroom dialogue.

A third example concerned rubric-mediated peer feedback. Here the rubric itself acted as a prism by naming the dimensions through which student work should be examined. Because attention was directed towards specific categories, peer dialogue could move beyond general praise or criticism and towards more precise analytical judgements. The design redistributed some of the lecturer's refractive role to students and artefacts, illustrating how structured peer-learning can support reflection without constant lecturer intervention.

Across these examples, technologies were not pedagogically important in themselves but because of how they stored, organised and foregrounded contributions. Discussion platforms preserved the trace of online dialogue; Padlet boards assembled reflections into a shared visual field; and polling tools enabled collective prioritisation of contributions. In each case, the technology became educationally meaningful when it helped make selected elements of the learning process visible and discussable.

5. Discussion

The findings suggest that the chief pedagogical value of refraction lies in didactic deceleration. The analysis of the prior studies indicate that students do not necessarily fail to reflect because they lack knowledge or experience, however, they often fail to reflect because educational designs do not create adequate occasions for pausing, focusing and analysing. Refraction names the intervention through which complexity is temporarily reduced to discussable elements. This matters in professional and practice-based education, where learners are often required to respond quickly and may therefore remain within action rather than gaining perspective on it.

The framework also clarifies the lecturer's role. The lecturer is not positioned as the sole source of knowledge but as the designer or mediator of reflective conditions. In some designs, the lecturer asks the refractive questions directly. In others, the lecturer delegates part of that role to peers, rubrics or digital artefacts. What remains constant is the need for a pedagogical arrangement that identifies what should be examined and keeps the dialogue oriented towards understanding rather than mere task completion.

The distinction between hard, soft, diffuse and absorbed reflection is useful because it makes visible the consequences of different levels of pedagogical support. Strong guidance may be necessary when learners are new to a field, when tasks are conceptually demanding or when tacit practice needs to be surfaced quickly. Softer forms of reflection may be more appropriate when the aim is to extend and compare learner-generated interpretations. Diffuse and absorbed forms can also occur, but the analysis suggests that relying on them as default arrangements is unlikely to support deep reflective dialogue consistently.

These points have practical implications for blended learning. Productive blended design is not achieved simply by adding online activities to face-to-face teaching. Instead, the analysis of the prior studies indicates that the sequence and function of activities matter. Online dialogue becomes powerful when it creates material that can be revisited and discussed later, rubrics become powerful when they name dimensions of analysis, and technologies become powerful when their affordances align with the reflective task. In this regard, refraction provides a useful design lens for deciding how a learning activity will move from action to analysis and from analysis to synthesis.

6. Conclusion

This paper seeks to reformulate and strengthen a conceptual argument about reflective dialogue in higher education by presenting it within a conventional academic structure and by making the analytical procedure explicit. The synthesis of prior studies indicates that reflective dialogue is most likely to emerge when educational designs provide a clear object of reflection, a frame for interpreting that object, interlocutors with whom to think, and an educational purpose for the reflective work.

The proposed framework distinguishes among reflex, refraction and reflection. Reflex refers to routine action; refraction refers to the pedagogical interruption that decomposes activity into analytically manageable elements; and reflection refers to the subsequent synthesis of experience, dialogue and theory. Together with the typology of hard, soft, diffuse and absorbed reflection, the framework offers a vocabulary for analysing existing learning designs and for planning new ones.

The most important implication is that reflection should not be treated as a vague aspiration appended to teaching, but as an outcome requiring deliberate design. Further research could test the framework in additional educational contexts and examine how different pedagogical prisms shape the quality of reflective dialogue across disciplines and levels of expertise.

Ethics Statement

The paper is based on previously reported studies. The original datasets are described as anonymised and reported without personally identifiable information. No new empirical data were collected for the present synthesis.

AI Declaration

AI was used in the original submission to generate one illustrative figure. AI is an integral part of the search tools (NotebookLM) used for this study for wordings and alternative formulation.

References

Baucal, A. (2010). Can you think with me? Educational Dialogues: Understanding and Promoting Productive Interaction, 64.

Di Stefano, G., Gino, F., Pisano, G. P., & Staats, B. R. (2023). Learning by thinking: How reflection can spur progress along the learning curve. Harvard Business School NOM Unit Working Paper, (14-093)

Dohn, N. B., Cranmer, S., Sime, J., De Laat, M., & Ryberg, T. (2018). Networked learning: Reflections and challenges. Springer.

Hachmann, R., Rasmussen, H. F., & Kjærgaard, T. (2023). Emerging Networks-Disturbed learning trajectories and new ways of connectivity at University Colleges during the covid-19 lockdown. Sustainable networked learning: Individual, sociological and design perspectives. Springer.

Howe, C. (2010). Peer dialogue and cognitive development: A two-way relationship? Educational Dialogues (pp. 44–59). Routledge.

Kjaergaard, T., & Hachmann, R. (2022). (2022). Backchannels: Covert Digital Backchannels in the Overt Classroom. Paper presented at the European Conference on E-Learning, 21(1) 192–198.

Kjærgaard, T. (2016). Utilising digital technology for dialogue and evaluation – a quasi-scholastic method in action. CEPRA-Striben - Tidsskrift for Evaluering i Praksis, (20)

Kjærgaard, T. (2019). (2019). Hybrid Homework–Blending Blended Learning and Face to Face in four Undergraduate Education Programmes. Paper presented at the World Conference on Online Learning,

Kjærgaard, T. (2020). Hybride læringsrum: Digitale kompetencer gennem læringscirkler.

Kjærgaard, T., & Andersen, C. E. M. (2023). Inter-lesson reflections: Learning through interrogative and reflective dialogue. ECRPL23 Proceedings,

Kjærgaard, T., & Dau, S. (2018). (2018). 21 century learning skills revisited-a conceptual paper on leaving'gaps' and going deep. Paper presented at the Proceedings of the International Conference on Networked Learning, 11 72–79.

Kjærgaard, T., & Georgsen, M. (2021). (2021). Reflection, dialogue, and hybrid learnings spaces: Refleksion, dialog og hybride læringsrum. Paper presented at the 127–144.

Kjærgaard, T., Jespersen, A. W., & Lassen, A. (2025). (2025). Where Technological Literacy and Literacy Meet: New Terms for Old Competencies. Paper presented at the European Conference on E-Learning, 195–204.

Kjærgaard, T., & Sorensen, E. K. (2016). Designing for Dialogue and Digitality in Higher and Continuing Education. Proceedings for Networked Learning 2016,

Littleton, K., & Howe, C. (2010). Educational dialogues: Understanding and promoting productive interaction. Routledge.

Mercer, N., Hennessy, S., & Warwick, P. (2017). Dialogue, thinking together and digital technology in the classroom: Some educational implications of a continuing line of inquiry. International Journal of Educational Research,

Padesky, C. A. (1993). (1993). Socratic questioning: Changing minds or guiding discovery. Paper presented at the A Keynote Address Delivered at the European Congress of Behavioural and Cognitive Therapies, London, , 24

Pagano, M., & Roselle, L. (2009). Beyond reflection through an academic lens: Refraction and international experiential education. Frontiers: The Interdisciplinary Journal of Study Abroad, 18, 217–229.

Paul, R., & Elder, L. (2007). Critical thinking: The art of Socratic questioning. Journal of Developmental Education, 31(1), 36.

Rasmussen, H. F., Hachmann, R., & Kjærgaard, T. (2022). (2022). Emerging Networks: A study on learning networks during the Covid-19 lockdown. Paper presented at the

Schön, D. (1987). Educating the reflective practitioner.

Schön, D. A. (2017). The reflective practitioner: How professionals think in action. Routledge.

Sorensen, E. K., & Kjærgaard, T.Designing for Dialogue and Digitality in Higher and Continuing Education. Paper presented at the Networked Learning 2016, Lancaster.

Turns, J., Sattler, B., Quigley, L. D. T., Atman, C., Bankhead, R. B., Carberry, A. R., Csavina, K., Cunningham, P., Faust, D., Harding, T., & Yasuhara, K. (2015). Reflecting on reflection: How educators experience the opportunity to talk about supporting student reflection.

Valbak-Andersen, C., Kjærgaard, T., & Bundgaard, S. B. (2021). Kernebegreber i refleksiv praksislæring. (pp. 15–27). Aalborg Universitetsforlag.

Vygotsky, L. S. (2012). Thought and language. MIT press.

Vygotsky, L. S., & Cole, M. (1978). Mind in society: Development of higher psychological processes. Harvard university press.

Wagenschein, M. (1951). Das Tübinger Gesprch. Die Pdagogische Provinz,

Wagenschein, M. (2012). Om begrebet eksemplarisk undervisning. MONA-Matematik-Og Naturfagsdidaktik, 4(4)
Wagenschein, M., & Horton-Kriiger, G. (2000). Teaching to understand: on the concept of the exemplary in teaching. Teaching as a Reflective Practice.the German Didaktik Tradition, 161–175.

Societal Transformation Related to Emerging Technologies in Industry Through Bi-Lateral Collaboration

Marisol Kokko Rico[1], Mikael Ericsson[2], Prakruthi Hareesh[3], Pb Venkataraman[3] and Kerstin Johansen[1], Lennart Malmsköld[2]

[1]Product Development, Production and Design, School of Engineering, Jönköping University, Sweden
[2]Institutionen för Ingenjörsvetenskap, Högskolan Väst, Trollhättan, Sweden
[3]Work Integrated Learning Programme, BITS Pilani, India

marisol.kokko.rico@ju.se
mikael.ericsson@hv.se
prakruthi.hareesh@pilani.bits-pilani.ac.in
pb.venkataraman@pilani.bits-pilani.ac.in
kerstin.johansen@ju.se
lennart.malmskold@hv.se

Abstract: Emerging Industry 4.0 technologies—such as automation, additive manufacturing, and artificial intelligence—are accelerating the need for continuous upskilling in industrial workforces. While upskilling is not a new societal phenomenon, the current pace and breadth of technological change interact with national education systems, industrial structures, and funding models in ways that differ substantially across countries. This creates both challenges and opportunities for international collaboration in lifelong learning and work-integrated learning (WIL). This paper examines societal transformation related to emerging technologies through a bilateral lens, focusing on learning initiatives and institutional practices in Sweden and India. Drawing on observations and interactional collaboration between higher education institutions in both contexts, we illustrate how upskilling demands are shaped by differing national trajectories. For example, historical Swedish adult education initiatives—such as evening studies undertaken to respond to increasing internationalization—can be contrasted with contemporary patterns in India, where engineers often engage in weekend-based learning to meet rapidly evolving industrial competence needs. Despite the global nature of industrial markets, priorities regarding who funds upskilling (industry, academia, individuals, or the state), how learning time is allocated, and how competence development is evaluated vary across contexts. We benchmark a WIL-oriented model associated with Indian experiences against Swedish experiences of WIL and educational quality enhancement, including structured follow-up with participants (e.g., focus groups) to iteratively improve course design. Based on this comparative analysis, we propose an initial conceptual framework to support bilateral collaboration in upskilling emerging technologies. The framework highlights critical dimensions to align across contexts, including technology maturity, competence capability, and resource conditions such as time and financial support in both industry and academia. The paper concludes with a gap-oriented discussion of how virtual laboratories and shared e-learning platforms could enable scalable bilateral collaboration for work-integrated upskilling.

Keywords: Work-Integrated E-Learning; Work-Integrated Learning (WIL); Lifelong Learning; Industry 4.0 Upskilling; Reskilling, Virtual Laboratories; International/Bilateral Collaboration; Course Design and Evaluation.

1. Introduction

Emerging Industry 4.0 technologies—such as automation, additive manufacturing, and artificial intelligence—are accelerating the need for continuous upskilling in industrial workforces. While upskilling is not a new societal phenomenon, the current pace and breadth of technological change interact with national education systems, industrial structures, and funding models in ways that differ substantially across countries. This creates both challenges and opportunities for international collaboration in lifelong learning and work-integrated learning (WIL).

The diffusion of Industry 4.0 technologies, including automation, artificial intelligence, cyber-physical systems, and data-driven manufacturing, is fundamentally reshaping industrial work and competence requirements worldwide (Khan & Mourad, 2025). As production systems become increasingly complex and interconnected, industrial professionals are required to engage in continuous learning throughout their careers.

While lifelong learning has long been discussed in policy and educational discourse, the speed, complexity, and interdisciplinary of current technological change represent a qualitative shift compared to earlier industrial transitions (Asiedu & Tenakwah, 2025) (Forum, 2025). This shift challenges traditional models of education that assume a clear separation between formal education and professional practice.

Upskilling does not occur in isolation but is embedded within national education systems, industrial labor structures, cultural expectation, and funding mechanisms (Li, 2022). As a result, similar technological drivers may lead to markedly different institutional responses across countries. Understanding these differences is essential for designing effective international collaboration in work-integrated learning (WIL) and lifelong learning (LLL) (Zegwaard & Pretti, 2023).

This paper examines societal transformation related to emerging technologies through a bilateral lens, focusing on learning initiatives and institutional practices in Sweden and India. Drawing on observations and interactional collaboration between higher education institutions in both contexts, we illustrate how upskilling demands are shaped by differing national trajectories. For example, historical Swedish adult education initiatives—such as evening studies undertaken to respond to increasing internationalization—can be contrasted with contemporary patterns in India, where engineers often engage in weekend-based learning to meet rapidly evolving industrial competence needs. Despite the global nature of industrial markets, priorities regarding who funds upskilling (industry, academia, individuals, or the state), how learning time is allocated, and how competence development is evaluated vary across contexts.

This paper adopts a bilateral comparative perspective, focusing on Sweden and India, two contexts with strong engineering traditions and global industrial integration, yet distinct socio-institutional trajectories.

Through collaborative interactions between higher education institutions in both countries, the study explores how upskilling practices are organized, financed, and evaluated.

2. Background

2.1 Industry 4.0 and the Upskilling Imperative

Research consistently highlights that Industry 4.0 is not merely a technological transformation but also a profound workforce transformation (Li, 2022), which is significant for manufacturing SMEs in their strive to implement new emerging technologies (Onu, 2021) Challenges related to emerging technologies are the lack of skilled persons (Mittal et al., 2018). (Mittal, 2018) However, studies estimate that a significant proportion of current skills will become obsolete within a decade, requiring both reskilling and upskilling across industrial sectors (Forum, 2025). Recent literature emphasizes that effective upskilling must integrate technical competencies with transversal skills such as system thinking, adaptability, and digital literacy (Garcés, Sanz-Benlloch, Montalbán-Domingo, & Díaz-Lantada, 2025). This has led to increased attention to learning models that combine academic knowledge with workplace practice.

2.2 Work-Integrated Learning (WIL)

Work-Integrated Learning (WIL) is widely regarded as a pedagogical and institutional approach that intentionally integrates learning and work through structured collaboration between higher education and industry (Zegwaard & Pretti, 2023). It encompasses a range of models including project-based learning, professional courses, and work-based education. However, comparative research demonstrates that WIL is implemented differently across national contexts, shaped by regulatory frameworks, labor-market conditions, and institutional cultures (Sunnemark, o.a., 2024). These differences affect funding responsibility, learning time allocation, and quality assurance practices. Involving students actively in WIL enhances professional growth, offering scalable alternatives to placements (cases) and contribute to learning by doing (Garner, 2026) (Garner & Wilson, 2026). However, it is at least three different stakeholders in the WIL process – the employer, the employees and the education provider (Rikala P, 2024). The students might be regular program students, employees and job seeking persons (Braun Greta, 2024), which can challenge the WIL organisation and add a possibility to engage the students with different backgrounds and experiences (Garner & Wilsson, 2026). BITS Pilani WILP model, for example, has a mature and large-scale system, it was established in the early 1970s with strong integration with industry, around 95% of campus students move into industry and employers actively sponsor and support education.

2.3 National Trajectories of Adult and Continuing Education

Sweden has long tradition of adult and continuing education supported by public policy, including structured evening studies and employer-supported learning initiatives aimed at supporting industrial transformation and social inclusion (Lundh Snis, Per, & Irene, 2025). In contrast, temporary Indian upskilling patterns often reflect

rapid industrial expansion combined with competitive labor markets. Engineers frequently engage in weekend-based on parallel learning, often driven by individual or employer initiative, to maintain employability in fast-evolving technological domain (Warnhoff, 2025). However, the emerging technologies are global, and it is important that training contribute to bridging the competence gap and is valuable for the company (Johansen & Akay, 2022) and the introduction of new technologies in combination with development of skills needs to be continously discussed (Brynjolfsson, 1993)

3. Methodology

We benchmark a WIL-oriented model associated with Indian experiences against Swedish experiences of WIL and educational quality enhancement, including structured follow-up with participants (e.g., focus groups) to iteratively improve course design. Based on this comparative analysis, we propose an initial conceptual framework to support bilateral collaboration in upskilling emerging technologies.

This study employs a qualitative, exploratory, and comparative approach, consistent with prior research on WIL and international education collaboration (Braun Greta, 2024). The aim of the study is to analyze analytic generalization rather than statistical representativeness when it comes to understanding the similarities and differences between the two national contexts.

The analysis draws on:

- Observations from bilateral collaboration between higher education institutions in Sweden and India
- Interaction with educators, administrators, and industry-aligned partners
- Reflections on implemented or piloted WIL-oriented initiatives
- Structured follow-up activities, including focus groups used for educational quality enhancement (Sunnemark, et al., 2024)

4. Comparative Findings: Sweden and India

This comparative has been done based on literature, observations, discussions, and recently focus group reflections where the authors have had the opportunity to visit both countries and our respective faculties.

BITS Pilani WLP model is mature, large-scale education system established in the early 1970s, currently serves ten of thousands of students and demonstrates a strong alignment between education and industry needs. A key feature is its high employability outcome, with approximately 95% of campus students transitioning into industry, supported by active employer engagement in education.

Jönköping University (JTH) and University West (HV) both implement Work-Integrated Learning (WIL), but with distinct emphases. At JTH, WIL is primarily embedded through CDIO-based engineering education, project-based learning, and structured industry collaboration (e.g., host company programs), enabling students to engage in real-world challenges while remaining largely campus-based (Berglund, et al., 2008). In contrast, University West adopts WIL as a core institutional philosophy, integrating academic and workplace learning across education, research, and collaboration, and emphasizing reflection, active participation, the co-creation of knowledge between academia and society. While JTH focuses on innovation-driven projects and CDIO-aligned competences development, HV provides a more holistic and a research field (Areskoug Josefsson, et al., 2024). Table 1 showed a comparison between BITS Pilani, JU and HV, the first column indicates the dimensions based on meetings and a recently focus group. The main differences highlighted are: The Role of Industry, the scale and flexibility, the Pedagogical Focus, the Students Profile, and the Business Model.

4.1 The Role of Industry: Funding Responsibility and Degree of Work Integration

Funding models vary markedly across contexts. Swedish WIL initiatives typically involve shared responsibility between the state, higher education institutions, and employers, whereas in India funding is often shared between industry and individuals, with limited public support for continuing professional education (Asiedu & Tenakwah, 2025). BITS WILP is Fully work-integrated; employment is mandatory and the workplace is the primary learning environment, and employers co-own the education (funding, mentoring, infrastructure). While JTH model is partially integrated and has strong project-based collaboration but is still mainly campus-centered. Employers act as collaborators in projects and innovation activities. HV model is systematically integrated; work-

based learning is embedded but not always tied to employment and employers are co-creators of learning within a structured WIL framework.

4.2 The Scale and Flexibility: Organization of Learning Time

In the Swedish context, upskilling is often embedded within formal educational structures, including part-time and evening studies supported by employers or public funding, reflecting an institutionalized commitment to lifelong learning (Lundh Snis, et al., 2025). In contrast, Indian professionals frequently engage in weekend-based or parallel learning while employed full-time. These arrangements provide flexibility but also place significant responsibility on individuals to manage learning time alongside professional demands (Warnhoff, 2025). Regarding BITS WILP has a large- scale, moderate flexibility (blended and project-based), while HV has smaller scale, increasing flexibility but still more structured than WILP.

4.3 The Pedagogical Focus, the Student Profile and Education Quality Enhancement

According to the findings based on Table 1, BITS WILP model has experience-driven, mentor based, real industrial problem solving and the student profile is employee-students (working professionals). JTH has a pedagogical focus on CDIO, innovation, theory and practice and the student profile is traditional full-time students with exposure. The HV model pedagogical focuses on reflection-driven WIL pedagogy integrating theory and practice. The student profile is mixed (students and practitioners), but less systematically employment based. Swedish WIL models commonly include systematic follow-up mechanisms such as course evaluations and focus groups, enabling iterative educational improvement (Sunnemark, et al., 2024). Indian models tend to prioritize immediate industrial relevance and skill acquisition, sometimes with less formalized academic quality assurance but strong alignment to industry needs (Warnhoff, 2025)

4.4 The Business Model

The BITS WILP model had Company-funded education with clear ROI, while JTH business model has public funding + project-based external funding (EU/regional) and structured trainee programs, and the HV business model has public funding with partial employer co-financing and large-scale industry engagement. Both JTH and HV have similar business model and common trainee programmes and courses.

Table 1: Comparison of dimensions between India vs Sweden

Dimension	BITS Pilani – WILP	Jönköping University (JTH/JU)	University West (HV)
Overall Model	Fully work-integrated education system at scale (since 1970s)	Predominantly campus-based with project and industry collaboration elements	Explicit Work-Integrated Learning (WIL) university model (core identity)
Scale & Maturity	Very large-scale (tens of thousands), long-established	Medium scale, strong program-level innovations (e.g., MASTech)	Mature WIL model in Sweden, but smaller scale than WILP
Entry Requirement	Employment mandatory (students are working professionals)	No employment requirement; students are full-time learners	Mixed: students + professionals, but employment not always mandatory
Industry Integration	Very strong and systemic (employers' co-owned education)	Strong but project-based (partners, Science Park, industry)	Strong, institutionalized collaboration with public and private sectors
Role of Employers	Employers must sponsor, mentor, and provide learning environments	Employers act as project partners, case providers, advisors	Employers act as co-creators of learning, often embedded in courses
Learning Arena	Workplace = primary learning lab	Hybrid: campus labs + Living Labs + industry projects	Workplace-based learning integrated with academic study
Pedagogical Approach	Situated learning + PBL + mentor-based learning + Imersion elements	CDIO + PBL + Human-Centered Design + Living Lab approaches	WIL pedagogy: reflection + work practice integration

Dimension	BITS Pilani – WILP	Jönköping University (JTH/JU)	University West (HV)
Online Education	Workplace learning Remote/online teaching Physical and virtual labs Teaching staff Includes both academic faculty and industry experts	Teaching staff Includes both academic faculty and industry workers	Teaching staff Includes both academic faculty and industry workers
Student Role	Employee-student (dual identity)	Primarily student, increasingly exposed to real-world contexts	Mix of student + practitioner, depending on program
Faculty Role	Continuous industry supervision + curriculum feedback loop	Faculty guide projects, research integration, and industry collaboration	Faculty facilitate reflection between theory and practice
Curriculum Development	Direct feedback from industry practice continuously integrated	Industry input present but less systematically embedded	Strong integration of practice-based insights into curricula
Programme Formats	- Degrees (2–5 years)	- Bachelor/Master programs	- Degrees + professional education
	- Certificates	- Some professional courses	- Short courses and upskilling
	- Cluster education	- Emerging lifelong learning formats	- Emerging lifelong learning formats
Flexibility	Highly flexible, asynchronous, designed for working professionals	Moderate flexibility; some blended and online learning	Increasing flexibility, but less than WILP
Lifelong Learning (LLL)	Core strategic focus	Growing importance (aligned with EU strategies)	Strong orientation toward continuous professional development
Use of Technology	AI tools, simulations, digital platforms	Strong in engineering tools, prototyping, simulation	Applied digital tools, but less emphasis on advanced AI in learning
Infrastructure	Workplace as lab + supplementary competence centers	Advanced engineering labs + Living Labs (e.g., Assistive Tech)	Workplace + university facilities; strong practice environments
Assessment Challenge	Measuring workplace learning outcomes is still a challenge	Assessment aligned with CDIO/ILOs; less workplace-based assessment	Emphasis on reflection-based assessment, but similar challenges
Industry engagement and Business Model	Company-funded education (high ROI expectation) High competition for skilled employees	Mostly public-funded education + project funding (EU, regional). Large-scale industry engagement.	Public funding + some employer co-financing. Large-scale industry engagement.
Impact Evidence	Strong quantitative results (efficiency, savings, innovation outcomes)	Impact visible on projects, student outcomes, research outputs	Impact on regional development and workforce competence
Unique Strength	Full integration of education-work-industry at scale	Integration of education, research, and innovation	Institutionalized WIL pedagogy and reflection-based learning
Key Limitation	Scalability relies on strong employer ecosystem	Not yet fully work-embedded (employment not required)	Less scale and less deep integration into company structures than WILP

5. Benchmarking Work-Integrated Learning Models

Benchmarking Indian WIL-oriented practices against Swedish quality-enhancement approaches reveals complementary strengths. Indian models demonstrate high responsiveness to rapid industrial change, while Swedish models emphasize pedagogical structure, reflection, and institutional learning (Braun Greta, 2024) (Zegwaard & Pretti, 2023). This comparison suggests that bilateral collaboration can benefit from combining flexibility with systematic quality assurance.

6. A Preliminary Conceptual Framework for Bilateral Upskilling Collaboration

Based on comparative analysis, an initial conceptual framework is proposed. The framework identifies three key alignment dimensions:

1. Technology Maturity: Stability versus rapid evolution of industrial technologies (Khan & Mourad, 2025)
2. Competence Capability: Existing skill baselines and professional experience (Garcés, et al., 2025).
3. Resource Conditions: Time allocation, financial support, and institutional capacity (Asiedu & Tenakwah, 2025)

Explicit alignment across these dimensions is necessary to avoid uncritical transfer of single-context educational models.

7. Discussion and Conclusions

The findings underscore that upskilling challenges are global, but institutional responses are locally shaped. WIL emerges not only as a pedagogical approach but as an interface between societal transformation, industrial change, and educational governance (Zegwaard & Pretti, 2023) (Lundh Snis, et al., 2025). Bilateral collaboration offers opportunities for mutual institutional learning, rather than one-directional knowledge transfer.

7.1 Future Opportunities: Key Challenges and Reflections for Sweden

Transferring the WILP model to Sweden faces structural challenges, including lower company incentives to sponsor education, limited SME engagement, and a tendency to address skills shortages through hiring rather than training, which puts traditional degree programs at risk of losing relevance as skills gain importance. Rather than replicate WILP, the discussion highlights the need for a context-adapted Swedish WIL model that maintains strong integration between education and the workplace while aligning with local economic and industrial conditions. Key opportunities include developing SME-oriented models, leveraging applied sciences and AI-enhanced learning, and using WIL to support industrial transformation and international collaboration.

Virtual laboratories and shared e-learning platforms are increasingly recognized as enabling infrastructures for scalable, international WIL collaboration (Sellber & Mads, 2024) (Zhang & Yin, 2024). Such environments can support asynchronous learning, shared curriculum development, and reduced dependency on physical co-location. However, research also highlights the need for robust assessment frameworks and contextual adaptations to ensure educational quality and relevance (Dominik, et al., 2023).

7.2 Conclusions

This paper has examined industry 4.0-driven upskilling through a bilateral comparison of Sweden and India. Differences in learning organizations, funding responsibility, and quality assurance highlight the importance of context-sensitive WIL models. The proposed framework provides a structured foundation for future bilateral collaboration. The BITS WILP model shows that large-scale, industry-integrated education is both feasible and highly impactful, generating interest in adapting a similar approach to developed economies like Sweden. Key success factors include strong employer engagement and funding, flexible learning formats, real industrial problem integration, and continuous mentorship. However, the conclusion emphasizes that while WIL and lifelong learning will be essential for future education systems, their implementation in Sweden requires system-level adaptation, new incentive structures, and stronger commitment from industry partners

Acknowledgements

The authors acknowledge STINT - The Swedish Foundation for International Cooperation in Research and Higher Education, for funding of the EWICE project as a collaboration between University West, Jönköping University both in Sweden and BITS Pilani in India. Furthermore, our industrial partners are acknowledged for their time and insights in workshops.

Ethics Declarations

Ethical clearance was not required for the research.

AI Declarations

Copilot was used to structure the paper and improve the language.

References

Areskoug Josefsson, K., Näverå, E., Wilner, A. & Masterson, D., 2024. Work-integrated learning at University West: Bibliographic review of the first 20 years with work-integrated learning. Waterloo, Canada, Wace Inc,.

Asiedu, E. & Tenakwah, E. S., 2025. Future-proofing your workforce: upskilling and reskilling as HR's top priorities Open Access. Strategic HR Review, 24(4), pp. 169-173.

Berglund, E., Birgersson, J. & Cederfeldt, M., 2008. Ten+ Years of Successful Workplace Learning through Host Company Collaboration. Gent, Belgium, 4th International CDIO Conference, Hogeschool Gent , June 16-19, 2008.

Braun Greta, 2024. Bridging Skill Gaps A systematic Literature Review of Strategies for Industry. Sustainable Production through Advanced Manufacturing, Intelligent Automation and Work Integrated Learning.

Brynjolfsson, E., 1993. The Productivity Paradox of Information Technology.. Commun ACM., pp. 36:66-77..

Dominik, M., Isa, J. & Stephanie, M., 2023. Online laboratories and virtual experimentation in higher education from a sociotechnical-pedagogical design perspective. Journal of Computing in Higher Education.

Forum2025.[Online] Available at: https://www.weforum.org/stories/2025/01/ai-and-beyond-how-every-career-can-navigate-the-new-tech-landscape/

Garcés, G., Sanz-Benlloch, A., Montalbán-Domingo, L. & Díaz-Lantada, A., 2025. Future engineering competencies for a sustainable world: an integral framework for engineering education in the Industry 4.0 era Available to Purchase. International Perspectives on Inclusive Education. IGI Global.

Garner, A. W. A., 2026. Creative work-integrated learning: Student-driven, cross-disciplinary, non-placement work-integrated learning in a changing educational landscape,. International Journal of Work-Integrated Learning,, Volume 27(1), pp. 151-163.

Johansen, K. A. S., 2022. Emerging technologies: Facilitating resilient and sustainable manufacturing,. SPS2022, In: Advances in Transdisciplinary Engineering 2022, (Eds: Amos, H.C., Syberfeldt, A., Högberg, D. and Holm, M.),, Volume Vol 22,, pp. .751-767.

Khan, S. H. & Mourad, A. H. I., 2025. Integrating Industry 4.0 in engineering education: Implementation patterns and pedagogical strategies.. Cogent Education, 12(1).

Li, L., 2022. Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. Information Systems Frontiers.

Lundh Snis, U., Per, A. & Irene, B., 2025. Editorial on work-integrated learning for global engagement and societal development. Current Issues in Work-Integrated Learning.

Mittal, S. R. D. W. T., 2018. Towards a smart manufacturing toolkit for SMEs.. IFIPInternational Conference on Product Lifecycle Management. Springer, Cham., pp. 476-487.

Onu, P. M. C., 2021. Industry 4.0 opportunities in manufacturing SMEs: Sustainability outlook. Materials Today: Proceedings, Volume 44, pp. 1925-30..

Rikala P., et al 2024. Understanding and measuring skill gaps in Industry 4.0 — A review.. Technological Forecasting and Social Change. , Issue 201:123206.

Sellber, C. & Mads, S., 2024. Virtual Laboratories in STEM Higher Education: A Scoping Review. Journal of Systematic Review in Education.

Sunnemark, L. et al., 2024. Bridging theory and practice through work-integrated learning in Swedish higher education.. Critical Studies in Education, 65(4), pp. 403-420.

Warnhoff, K., 2025. Work-integrated learning in Industry 4.0: A qualitative analysis of assistance systems.. Industry 4.0 Science, 41(2), pp. 20-29.

Zegwaard, K. E. & Pretti, T. J., 2023. The Routledge international handbook of work-integrated learning (3rd ed.).. s.l.:s.n.

Zhang, N. & Yin, L., 2024. Design and implementation of virtual laboratories for higher education sustainability. Frontiers in Education.. Frontiers in Education.

A Redesigned Peer Learning Model for Fair Assessment in Distance Education

Vishaka Nanayakkara, Tharushi Kurukulasuriya, Buddhika Karunarathne and Jageshkar Sivahar
Centre for Open and Distance Learning, University of Moratuwa, Sri Lanka

vishaka@cse.mrt.ac.lk
tharushiha33@gmail.com
buddhika@cse.mrt.ac.lk
jageshkar@gmail.com

Abstract: Learning is widely adopted in distance education to promote collaboration. However, Peer group-based assessments often suffer from unequal contribution and free-riding. This study examines the redesign of a programming group project in a distance learning environment to address these concerns. In the previous implementation, students worked collaboratively to develop a web application, and the assessment was based on a shared group outcome. Analysis of the results revealed anomalies: students with weaker performance in other modules received disproportionately high grades, suggesting a potential dependence on peers. To mitigate this issue, the project structure was redesigned to balance collaboration with individual accountability. While students were randomly assigned to groups to maintain peer interaction, all project deliverables were completed individually. A structured peer-learning component was introduced through the Moodle Workshop module, where students submitted their Software Requirements Specification (SRS) and engaged in peer assessment by evaluating and providing feedback on group members' submissions. This workshop activity was embedded as a compulsory intermediate assessment within the project sequence, with progression strictly conditional: students were required to complete both their individual submission and the associated peer-review tasks before proceeding to subsequent stages of the project. Failure to complete this component prevented advancement, thereby ensuring full participation in both individual and peer-learning activities. The study investigates the impact of this intervention on assessment validity, student performance, and engagement. Findings suggest that the redesigned model reduces grade inflation, enhances individual accountability, and preserves the benefits of peer learning through structured feedback. The results highlight the potential of integrating peer review mechanisms in distance education to improve both learning outcomes and assessment integrity.

Keywords: Peer Learning, Distance Education, Assessment Integrity, Group Assessment, Project-Based Learning.

1. Introduction

Distance education increasingly adopts collaborative and project-based learning approaches to improve student engagement and encourage peer interaction. In software-related modules, group projects are commonly used to simulate industry practices and develop teamwork skills. However, maintaining fairness and assessment integrity in online group assessments remains a significant challenge.

One major issue in group-based assessments is unequal contribution, where some students rely heavily on stronger peers while contributing minimally to the actual implementation process. This issue is particularly critical in technical modules where practical software development skills are key learning outcomes. Although collaborative learning promotes discussion and peer support, fully shared assessment structures may unintentionally encourage free-riding behaviors and reduce individual accountability.

This study examines the redesign of the module ITE2953 Programming Group Project conducted at the Centre for Open and Distance Learning in 2026. In previous offerings of the module, students completed the project as a fully group-based assessment where major deliverables, including the project proposal and Software Requirements Specification (SRS), were submitted collectively. Although students were individually evaluated during the final viva, instructors observed that some students focused mainly on documentation tasks without actively participating in the implementation of the web-based system.

To address these concerns, the assessment structure was redesigned to preserve peer learning while improving individual accountability. Students were assigned to random peer groups for discussions and stakeholder analysis activities, while all assessed deliverables were completed individually. A structured peer-learning component was introduced through the Moodle Workshop module, where students evaluated peer SRS submissions using a rubric-based assessment process.

The redesigned module also introduced multiple assessment checkpoints, enabling continuous monitoring of student participation and technical progress throughout the project lifecycle. This paper presents the redesigned

assessment model, discusses the integration of structured peer-learning activities in a distance education environment, and reports preliminary observations from the ongoing implementation. The study focuses on peer-learning effectiveness, participation patterns, continuous assessment monitoring, and assessment integrity in technical project-based learning.

2. Literature Review

Peer learning has become an important pedagogical strategy in higher education, particularly within online and distance learning environments where student interaction and collaborative engagement support participation and learning outcomes. Research has shown that peer interaction contributes positively to students' sense of community, motivation, collaborative learning experiences, and continued engagement with online learning systems (Hidinger & Borland, 2023; Alshammari & Alshammari, 2024). These findings highlight the importance of collaborative engagement in effective online learning environments.

In computing and programming education, collaborative learning is commonly implemented through group projects that simulate real-world software development practices. Such projects help students develop teamwork, communication, problem-solving, and employability skills while preparing them for professional environments that increasingly rely on virtual collaboration (Donelan & Kear, 2024). However, despite these benefits, group-based assessments continue to present significant challenges related to participation and assessment fairness.

A recurring issue in collaborative learning literature is unequal contribution among group members. Studies frequently report problems such as low participation, uneven workload distribution, and weak communication within online group projects (Donelan & Kear, 2024). In programming projects, these challenges become particularly significant when technically stronger students undertake most development tasks while weaker students receive equivalent grades despite limited contribution. Such free-riding behaviours raise concerns regarding assessment validity and fairness.

Assessment of integrity has therefore become an important consideration in digitally mediated learning environments. Delvenne (2026) emphasized that assessment validity is compromised when submitted work does not accurately represent a student's authentic learning or reasoning process. In collaborative programming projects, this concern emerges when assessment outcomes fail to reflect individual technical competency. Consequently, there is increasing interest in assessment models that preserve the benefits of collaboration while strengthening individual accountability.

Peer assessment has been widely proposed as a mechanism for addressing these concerns. Research suggests that peer review can improve accountability, engagement, reflective judgement, and collaborative responsibility while encouraging students to critically evaluate the work of others (Yang et al., 2022). However, peer assessment is not without limitations. Challenges relating to reliability, consistency, friendship bias, leniency, and student resistance have been reported in the literature (Yang et al., 2022; Rasooli et al., 2025). These concerns highlight the importance of careful pedagogical design when implementing peer-learning activities.

Recent studies advocate for structured assessment approaches that balance collaboration with accountability. Valencia-Forrester (2026) demonstrated that peer-review frameworks can support assessment integrity and student engagement when embedded within a carefully designed learning process. Similarly, Learning Management System tools such as Moodle Workshop provide opportunities to facilitate structured peer-review workflows through automated allocation, feedback exchange, and progression management.

Although previous research has explored collaborative learning, peer assessment, and assessment fairness in higher education, limited attention has been given to assessment redesign models that combine structured peer learning with individually assessed project deliverables in distance programming education. This study addresses that gap by examining a redesigned programming project model that integrates Moodle Workshop peer-review activities while maintaining individual assessment requirements. The study aims to explore how collaboration can be preserved while strengthening accountability, assessment validity, and participation within an asynchronous distance learning environment.

3. Research Methodology

3.1 Research Approach

This study adopted an Action Research approach combined with a mixed-methods research design. The research was conducted within the ITE2953 Programming Group Project module offered by the Centre for Open and Distance Learning (CODL), University of Moratuwa. The study involved the redesign and implementation of a project-based assessment framework intended to improve individual accountability while preserving peer-learning opportunities in a distance education environment. The action research approach was selected because the intervention was implemented by the teaching team within the module and evaluated through ongoing observation and analysis of student participation and assessment outcomes

3.2 Research Questions

The study was guided by the following research questions:

RQ1: How does the redesigned peer-learning model influence student participation in a distance learning programming project module?

RQ2: How does the redesigned assessment framework improve the monitoring of individual student engagement compared with a traditional group-based assessment structure?

RQ3: Can structured peer-review activities preserve the benefits of collaboration while increasing individual accountability in project-based learning?

RQ4: What are the preliminary implications of the redesigned model for assessment integrity in distance education?

3.3 Participants and Context

The study was conducted during the 2025S1 offering of the ITE2953 Programming Group Project module. A total of 210 students were enrolled in the module. Students were required to develop an individual web-based application based on a common project theme while participating in structured peer-learning activities. Students were randomly assigned to peer-learning groups to support discussion, stakeholder analysis, and peer-review activities.

For comparative purposes, assessment data from the previous module implementation, which followed a traditional group-based assessment model with 185 enrolled students, were also considered.

3.4 Previous Assessment Model

The previous module structure included:

- Fully group-based project activities
- Shared submission of SRS, project report, and final project submission
- Shared grading for early assessments
- Individual viva evaluation at the final stage

Although the viva assessed students individually, observations indicated that some students contributed minimally to implementation activities while relying on peers during earlier stages.

3.5 Redesigned Assessment Model

The redesigned structure preserved peer learning while introducing individual accountability.

The redesigned assessment framework consisted of five major stages:

- Individual Project Proposal Submission
- Moodle Workshop Peer Assessment Activity
- Final Software Requirements Specification (SRS) Submission
- Individual Project Implementation and Final Submission
- Practical Examination and Viva Evaluation

Students were required to complete each stage successfully before progressing to subsequent assessment activities. The Moodle Workshop activity required students to submit a draft SRS document and complete four anonymous peer reviews using a rubric-based assessment framework. Students subsequently revised their SRS based on the feedback received.

3.6 Data Collection

Data were collected from multiple sources generated during module delivery, including:

- Moodle assessment completion records
- Moodle Workshop participation records
- Project submission data
- Practical examination attendance records
- Final viva eligibility records
- Instructor observations recorded throughout module delivery

At the time of writing, final viva evaluations were still ongoing. Therefore, the present analysis focuses primarily on participation and progression data obtained from continuous assessment activities.

3.7 Data Analysis

Quantitative data were analyzed using descriptive statistical techniques to examine participation patterns, progression rates, and assessment completion across the project lifecycle. Comparative analysis was performed between the previous group-based assessment model and the redesigned individually assessed peer-learning model. Qualitative observations recorded by instructors during module delivery were used to interpret participation trends, identify implementation challenges, and evaluate the effectiveness of the redesigned assessment structure.

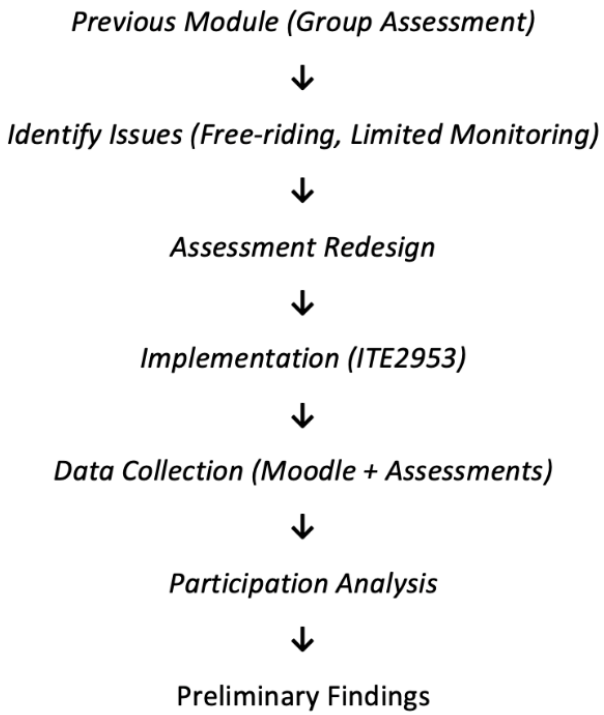


Figure 1: summary of the research process

4. Redesigned Peer Learning Framework

4.1 Assessment Workflow

Table 1: A summary of the workflow that has been redesigned for the module

Stage	Activity	Assessment Type
1	Project Proposal	Individual
2	SRS Draft Submission (Moodle Workshop)	Individual with Peer Learning
3	Peer review (Moodle Workshop)	Individual with Peer Learning
4	Final SRS Submission	Individual
5	Final Project Implementation and Submission	Individual
6	Final report submission	Individual
7	Practical Exam	Individual
8	Final Viva Evaluation	Individual

4.2 Moodle Workshop Integration

Students submitted their draft SRS documents through Moodle Workshop and were automatically assigned four peer submissions for evaluation using a five-level rubric.

The workshop process included:

- Step 01 - Individual draft submission
- Step 02 - Anonymous peer allocation
- Step 03 - Peer evaluation
- Step 04 - Feedback review
- Step 05 - Final SRS resubmission

Progression to subsequent assessments was conditional upon completing both submission and peer-review tasks.

5. Results and Findings

5.1 Comparative Analysis of Assessment Outcomes and Individual Accountability

To evaluate the effectiveness of the redesigned peer learning model, assessment outcomes from the current implementation (ITE2953) were compared with those from the previous curriculum (ITE2952), where a conventional group assessment approach was followed. The comparison focused on indicators related to individual accountability, including viva participation, viva performance, and successful module completion. While completion rates provide an overview of student progression, individual viva performance provides stronger evidence of whether students could demonstrate ownership and understanding of their project contributions.

Table 2: Comparison of Assessment Outcomes and Individual Accountability Indicators Between Conventional and Redesigned Assessment Models

Indicator	Previous Curriculum (ITE2952 – Conventional Group Assessment)	Redesigned Peer Learning Model (ITE2953)
Total enrolled students	185	210
Students eligible for viva assessment	173	164
Students who participated in viva	165	150
Viva participation rate (among eligible students)	95.4%	91.5%

Indicator	Previous Curriculum (ITE2952 – Conventional Group Assessment)	Redesigned Peer Learning Model (ITE2953)
Students successfully completing the module	146	137
Overall completion rate (based on enrolment)	78.9%	65.2%
Students achieving viva marks above 50	102	108
Percentage of viva participants achieving above 50	61.8%	72.0%
Highest viva mark	92.05	100
Assessment approach	Shared group marks for selected assessments + viva	Peer assessment + individual accountability checkpoints + practical assessment + viva verification

Although the previous curriculum demonstrated a higher overall completion rate, this metric alone does not indicate the accuracy of assessment outcomes or the extent of individual learning. Under the conventional group assessment structure, students could achieve satisfactory overall grades through shared group marks, even when their individual viva performance indicated limited understanding. In contrast, the redesigned model introduced multiple accountability mechanisms that required students to demonstrate individual knowledge and ownership of their contributions. The comparison of viva outcomes provides stronger evidence of improved assessment integrity. While 61.8% of viva participants achieved marks above 50 in the previous curriculum, this increased to 72.0% under the redesigned model. Additionally, the highest viva mark increased from 92.05 to 100, with several students demonstrating exceptional understanding of their implementations and confidently explaining their design decisions. These findings suggest that the redesigned model strengthened the alignment between assessed performance and actual individual competence.

While Table 2 provides a comprehensive comparison of assessment outcomes, viva performance provides a more direct indication of individual understanding because students were required to independently explain their implementation decisions and project contributions.

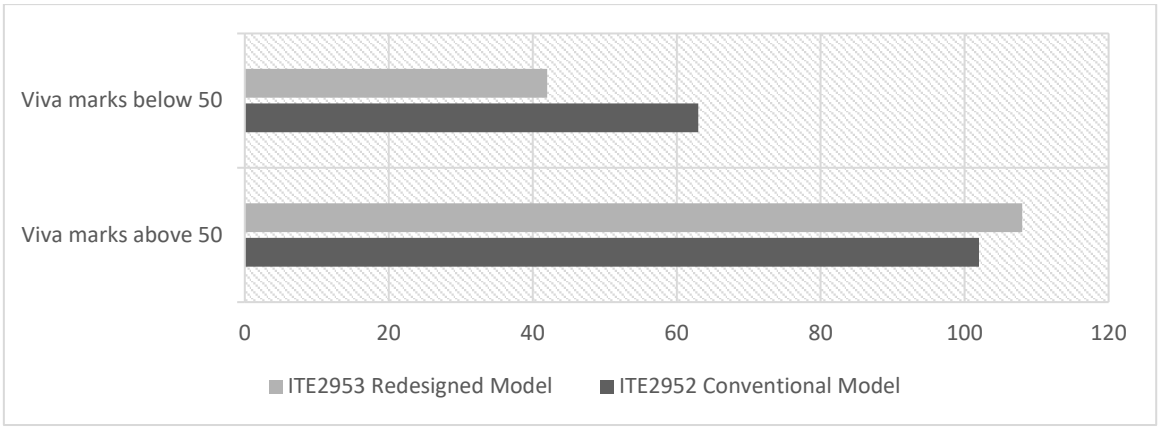


Figure 2: Comparison of Viva Performance Outcomes Between Conventional and Redesigned Assessment Models

As shown in Figure 2, the redesigned model resulted in a higher proportion of students achieving satisfactory viva performance compared with the conventional assessment approach. Under ITE2952, 102 out of 165 viva participants achieved marks above 50 (61.8%), whereas under ITE2953, 108 out of 150 viva participants achieved marks above 50 (72.0%). Additionally, the number of students receiving lower viva marks decreased from 63 to 42. These findings indicate that the redesigned model provided stronger mechanisms for ensuring that final assessment outcomes reflected individual understanding rather than solely group performance.

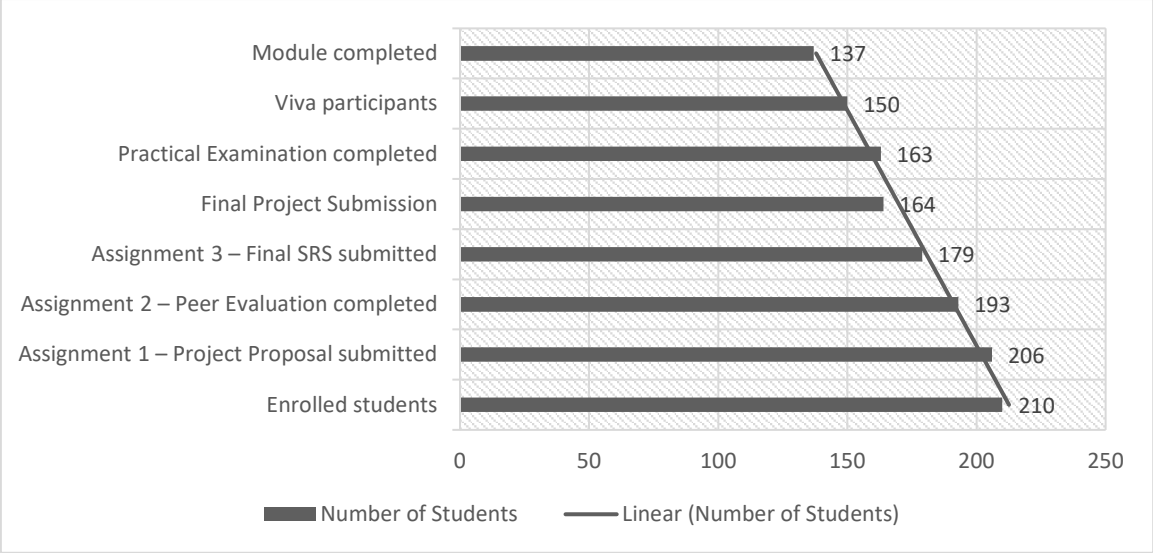


Figure 3: Student Progression Across Assessment Stages in the Redesigned Peer Learning Model

Figure 3 illustrates student progression across the different assessment stages of the redesigned model. The gradual reduction in participation reflects the expected attrition pattern in distance education environments, while most students continued through multiple individual and collaborative assessment checkpoints. 163 students completed the practical examination, demonstrating sustained engagement until the individual competency verification stage. The final viva and module completion outcomes further indicate that the redesigned structure provided multiple opportunities to monitor participation and identify students requiring additional support.

5.2 Comparative Analysis of Assessment Monitoring and Student Engagement

5.2.1 Assessment Structure Comparison

The two Figures show how assessment monitoring changed between the old and new curriculum.

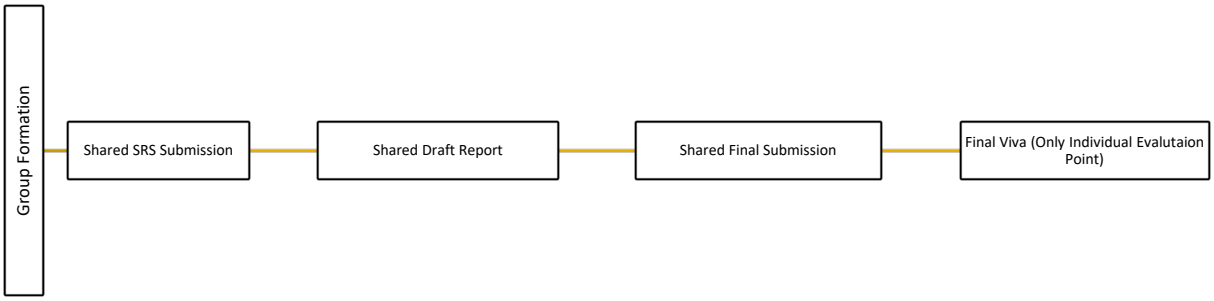


Figure 4: The old curriculum (ITE2952) flow

Here are the key issues of this flow

- Limited visibility of individual contribution
- Individual progress is monitored only during the final viva
- High possibility of opportunities for free riding

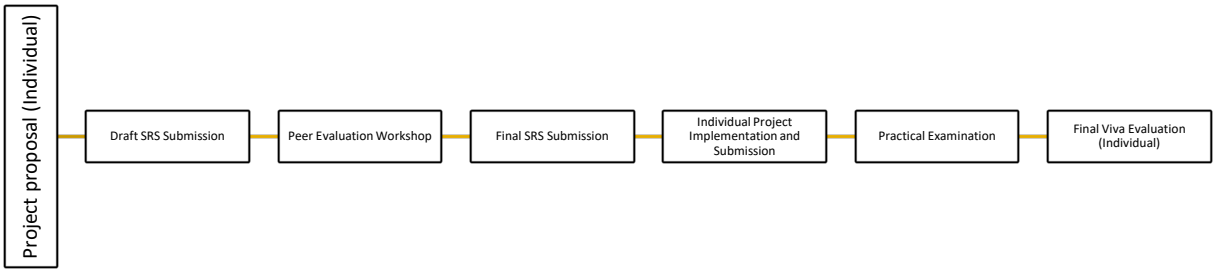


Figure 5: The new curriculum (ITE2953) flow

Here are the key advantages of this flow

- Continuous monitoring of student progress
- Multiple individual checkpoints
- Peer-learning with accountability
- Early identification of inactive students

5.2.2 Comparative Analysis of Assessment Structures

Table 3: Compares the previous fully group-based assessment structure with the redesigned peer-learning-based individual assessment model.

Assessment Aspect	Previous Model (ITE2952)	Redesigned Model (ITE2953)
Project Structure	Fully Group-Based	Individual Deliverables with Peer Learning
Proposal Submission	Shared Group Submission	Individual Submission
SRS Submission	Shared Group Submission	Individual Submission
Peer Learning	Informal Collaboration	Structured Moodle Workshop
Feedback Process	Limited Instructor Feedback	Peer + Instructor Feedback
Implementation Responsibility	Shared Among Group Members	Fully Individual
Assessment Progression	Flexible	Mandatory Completion-Based Progression
Practical Evaluation	Final Viva Only	Practical Exam + Final Viva
Accountability Level	Moderate	High
Risk of Free-Riding	High	Reduced
Assessment Integrity	Difficult to Monitor	Improved Through Individual Evaluation

5.2.3 Monitoring Visibility Comparison

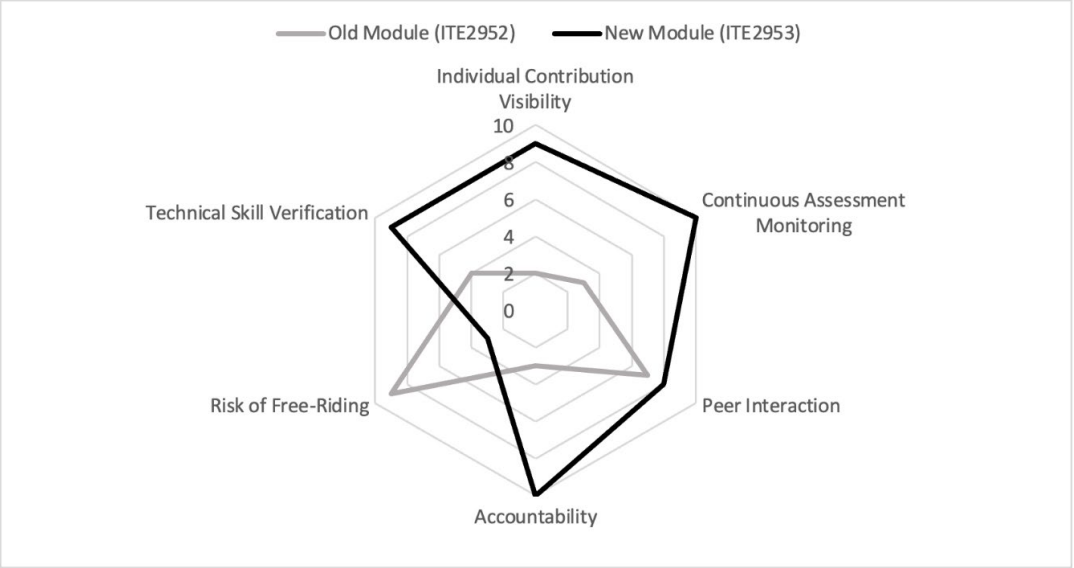


Figure 6: Visually demonstrates how the redesigned assessment structure improved the visibility of student engagement and technical competency throughout the project lifecycle.

5.3 Observed Outcomes

Initial observations suggest several positive outcomes:

- Increased individual engagement
- Reduced dependency on peers
- Better accountability across assessment stages
- Improved participation consistency
- Reduced opportunities for free riding

The structured progression model also encouraged students to actively complete every stage of the project. Although the redesigned module currently shows stronger engagement, accountability, and monitoring capability, the final effectiveness analysis will continue after the completion of the final viva evaluations. Future analysis will include comparative grade distributions, viva performance, technical competency validation, and long-term assessment integrity outcomes.

6. Discussion

The redesigned peer learning model demonstrates how structured peer interaction can be integrated into distance education environments while maintaining individual accountability and assessment integrity. The findings indicate that the model provided greater visibility of student engagement by introducing multiple accountability checkpoints, including peer evaluation, individual project responsibilities, practical examination, and viva assessment. Rather than relying solely on the final group output, the redesigned approach enabled continuous monitoring of student participation and created opportunities to verify individual understanding.

The comparison of viva assessment outcomes provides further evidence of the redesigned model's effectiveness in reducing opportunities for free riding. In the previous curriculum, which used a conventional group assessment structure, students could receive identical marks for selected group-based assessments regardless of their individual contributions. Although 146 out of 173 eligible students successfully completed the module, analysis of individual viva performance showed that only 102 students achieved marks above 50, while several students obtained considerably lower viva marks. This indicates that some students were able to successfully complete the module despite demonstrating limited individual understanding during the viva assessment, as group assessment marks contributed significantly to their overall performance.

In contrast, the redesigned model strengthened the relationship between student contribution and final achievement by requiring learners to demonstrate ownership of their work. Among the 164 students who

reached the viva stage under the redesigned model, 137 successfully completed the module. The highest viva mark increased to 100, with several students demonstrating excellent understanding of their implementation decisions and confidently explaining their individual contributions. Furthermore, 108 students achieved viva marks above 50, while the remaining students obtained marks above 35. These outcomes suggest improved alignment between assessed performance and actual individual understanding, indicating that the redesigned approach provided a more reliable mechanism for evaluating student contribution within group projects.

Despite these positive outcomes, several limitations and challenges should be acknowledged. Peer assessment remains vulnerable to potential bias, as students may provide overly favorable evaluations to maintain relationships with peers or may lack sufficient experience to critically evaluate the quality of another student's work. Although the use of anonymous peer review and structured rubrics helped reduce these concerns, further reviewer training and calibration activities could improve the reliability of peer-generated assessments. Additionally, some students may initially resist peer assessment due to unfamiliarity with evaluating peers or expectations of traditional instructor-led assessment methods. The additional workload associated with reviewing peers may also create challenges, particularly for students managing multiple academic or professional commitments in distance education settings.

The study is also limited by its implementation within a single module and a specific student cohort. Therefore, the findings should be interpreted within the context of the ITE2953 Programming Group Project module, and further research across different disciplines and larger student populations is required to establish the broader applicability of the model. Future implementations could incorporate more extensive student perception surveys, automated analytics from learning management systems, and longitudinal comparisons across multiple offerings to further evaluate the impact of structured peer learning on engagement, fairness, and learning outcomes.

Overall, the findings suggest that a carefully designed peer learning model can address some of the common challenges associated with group assessment in distance education. By combining peer evaluation with individual accountability mechanisms, the model creates a more transparent assessment environment where collaboration is encouraged while individual learning and contribution remain visible.

7. Conclusion and Future Work

This study presented a redesigned peer learning model implemented within the Centre for Open and Distance Learning programming project module. The redesigned framework addressed the limitations of fully group-based assessments by combining individual accountability with structured peer-learning activities facilitated through the Moodle Workshop module.

Unlike the previous assessment structure, where major deliverables were submitted collectively, the redesigned model required all assessed components to be completed individually while still preserving peer interaction through discussion activities, stakeholder analysis, and rubric-based peer evaluation. The integration of structured peer-review activities encouraged reflective learning, continuous engagement, and constructive feedback throughout the software development lifecycle.

Preliminary findings from the ongoing implementation suggest improved student participation, stronger individual accountability, and reduced free-riding behaviors. The progression-based assessment structure enabled continuous monitoring of student engagement and technical progress across multiple assessment stages rather than relying solely on the final viva evaluation.

However, this study represents a work-in-progress evaluation, as the final viva examinations are still ongoing. Therefore, the current findings are primarily based on continuous assessment participation data and early instructional observations. Future work will include analysis of final viva performance, comparative grade distribution analysis, student perception surveys, evaluation of peer feedback quality, and longitudinal analysis across future semesters.

The results are expected to provide deeper insight into the effectiveness of structured peer-learning frameworks in distance technical education and contribute to the development of fairer and more engaging assessment models for online project-based learning environments.

Ethics Declaration

This study was conducted as part of an educational assessment redesign within the module ITE2953 Programming Group Project at the Centre for Open and Distance Learning. The research used anonymized participation and assessment data collected during normal teaching and learning activities. No personal, sensitive, or identifiable student information was collected or disclosed. Therefore, formal ethical clearance was not required for this study.

AI Declaration

Artificial Intelligence (AI) tools were used during the preparation of this paper to support language refinement, content organization, grammar improvement, and academic writing enhancement. The authors reviewed, edited, and validated all generated content to ensure the accuracy, originality, and academic integrity of the final manuscript. AI tools were not used for data generation, analysis, or fabrication of research findings.

References

Alshammari, S.H. and Alshammari, M.H., 2024. Modelling the effects of emotional engagement and peer interaction on the continuous intention to use asynchronous e-learning. *PeerJ Computer Science*, 10, e1990.

Delvenne, J.F., 2026. Navigating GenAI in psychology education: assessment validity, academic integrity, and the realities of teaching in an AI-rich era. *Psychology Learning & Teaching*, pp.1–14.

Donelan, H. and Kear, K., 2024. Online group projects in higher education: persistent challenges and implications for practice. *Journal of Computing in Higher Education*, 36, pp.435–468.

Hidinger, K.B. and Borland, K.W., 2023. A study of peer interaction and community in accelerated online learning. *Journal of Online Learning Research and Practice*, 9(1), pp.5–24.

Rasooli, A., Turner, J., Varga-Atkins, T., Pitt, E., Asgari, S. and Moindrot, W., 2025. Students' perceptions of fairness in groupwork assessment: validity evidence for peer assessment fairness instrument. *Assessment & Evaluation in Higher Education*, 50(1), pp.111–126.

Valencia-Forrester, F., 2026. A peer review model for supervised work integrated learning assessment: design principles and implementation framework. *Assessment & Evaluation in Higher Education*.

Yang, A., Brown, A., Gilmore, R. and Persky, A.M., 2022. A practical review for implementing peer assessments within teams. *American Journal of Pharmaceutical Education*, 86(7), Article 8795.

Prompt-as-Lesson: Authoring AI-Delivered Lessons for Fidelity and Observability

David Lefevre
Imperial College Business School, Imperial College London, UK

david.lefevre@imperial.ac.uk

Abstract: This paper reports the development of the prompt-as-lesson: a lesson encoded as a single prompt and run in the student's own LLM tool, with the model cast throughout as a tutor delivering the teacher's lesson, and the session closing with a report of the dialogue returned to the teacher. The purpose of the prompt-as-lesson is to enable students to pursue teacher-specified learning outcomes through tutorial dialogue with an LLM. The interactive, conversational nature of an LLM makes Laurillard's Conversational Framework the natural interpretive lens. Under this lens, the paper argues, the LLM is the first single medium capable of enacting the iterative teacher–learner dialogue the framework holds essential to formal teaching, yet this dialogue is frequently lacking in higher education. Introducing LLM activities into the formal learning environment risks inconsistency as, left to its defaults, the model teaches from its own conception and constructed learning environment rather than the teacher's. The prompt-as-lesson is a response. Utilising the prompt-as-lesson within the teacher's formal environment makes three demands: consistency, so that what is taught is drawn from the teacher's conception and pedagogy; adherence, so that the enacted lesson remains the authored one; and reporting, so that the dialogue is not opaque to the teacher. Design-based iteration on two contrasting modules to address these demands yielded two primary findings: contemporary LLMs struggled to adopt the learner's perspective and would not reliably hold to an unstructured lesson. The format that emerged seeks to address both issues and comprises three artefacts: a control surface of fourteen dimensions; an eight-section structure; and a four-phase authoring method. The paper claims no effect on learning; its contribution is a control surface, a structure and an authoring method. A formal evaluation of the format's fidelity of implementation begins in September 2026.

Keywords: Prompt-as-Lesson; Generative AI in Education; Conversational Framework; Fidelity of Implementation; Learning Design; Design-Based Research

1. Introduction

1.1 The Second Learning Environment

Students increasingly use generative AI in their studies. In a recent survey, 95% of UK undergraduates reported using AI and 94% reported using generative AI to help with assessed work (Stephenson and Armstrong, 2026). The AI investigated here is the large language model (LLM). It is an interactive medium, accessed via conversation, and, this paper argues, capable of enacting a formal teaching environment.

This interactive nature makes Laurillard's (2002) Conversational Framework an appropriate interpretive lens. This framework models formal education not as linear transfer but as an iterative dialogue across four components: the teacher's conception, the teacher's constructed learning environment, the learner's conception and the learner's actions. Here, the teacher's conception refers to the expert knowledge, theories and mental models that the teacher aims to communicate. The teacher's constructed learning environment refers to the pedagogical tasks, tools and settings designed for students to interact with and test those concepts.

Students have always utilised learning media outside the formal environment convened by the teacher, for example websites and MOOCs. However, Laurillard's pre-generative-AI analysis of learning media reached a precise verdict: no single extant medium enabled the iterative model 'in the way that a teacher in a practical session could'. The extant media covered the essential activities only in combination (Laurillard, 2002: 174). The combining required the work of a teacher.

The LLM is the first single medium capable of enacting the entirety of the formal model. It holds the four properties the framework demands, being discursive, adaptive, interactive and reflective. It converses at the level of descriptions, sets and adapts tasks, gives feedback on actions and prompts reflection. It requires no assembly, and therefore no assembler.

The LLM may be a means by which iterative conversation is expanded within the teacher's formal learning environment. The Conversational Framework holds iterative conversation to be essential to learning, yet iteration is frequently lacking in formal higher education. As an example, coursework typically comprises a single feedback cycle rather than a dialogue.

This challenge has a long lineage in tutoring research, from Bloom's (1984) tutorial advantage to intelligent tutoring systems (VanLehn, 2011), and more recently educators have begun to exploit the capability of generative AI. Mollick and Mollick (2023; 2024) deploy single-activity prompts for use as class activities and persona-based assistants. Prompts embedded in custom agents have been shown to be valuable in the field of physics education (Kestin et al., 2025). Recent systematic reviews of prompt engineering in education (Lee and Palmer, 2025; Qian, 2025) document a literature in which the model is cast as the teacher's assistant or as a tutoring persona, but not as the enactor of a teacher-authored lesson.

A student using an LLM now has a second environment fulfilling the Conversational Framework's description of teaching. However, this raises the issue of inconsistency for students: left to its defaults, the model teaches from its own conception and constructed learning environment rather than the teacher's.

1.2 The Prompt-as-Lesson Format

This paper investigates a response to this potential inconsistency, the prompt-as-lesson. This is a lesson plan authored by the teacher to be consistent with the teacher's conception and constructed learning environment (Laurillard, 2002). It is delivered via the student's own LLM tool with the model cast throughout as a tutor with a brief to deliver the lesson as a one-to-one tutorial. A report of the resulting dialogue is returned to the teacher, maintaining the iterative dialogue the Conversational Framework holds essential to formal teaching.

To this end, the lesson is bounded by a beginning and an end and proceeds as a sequence of timed teaching phases, described in Section 4 as 'acts'. Rather than being left to the model's defaults, it is specified against an explicit control surface, the dimensions of which are defined in Section 5.

The purpose of the prompt-as-lesson is to enable students to pursue teacher-specified learning outcomes through tutorial dialogue with an LLM, conducted within the teacher's formal learning environment rather than outside it. The following sections describe how the approach was chosen and how the subsequent format developed. They then present the resulting structure, a generalisable authoring method, and a planned evaluation.

2. Requirements and the Challenge of Fidelity

Bringing the LLM into the teacher's formal learning environment makes two demands. The first is consistency. The knowledge base, pedagogy and tone need to be aligned with the teacher's conception and pedagogical approach. Left to its defaults the model will draw upon its own knowledge base, which may differ from the teacher's conception and the curriculum curated for a class. Pedagogy may also differ. For example, LLMs are currently optimised for helpfulness, whereas desirable difficulties (Bjork, 1994) and productive failure (Kapur, 2008) hold that instruction frequently requires assistance to be withheld. The second demand is reporting. When a student uses an LLM the conversation and its outcomes are typically opaque to the teacher; maintaining the teacher–student dialogue requires some report of them.

The prompt-as-lesson format introduced in Section 1.2 adds a third demand: adherence. The lesson must be enacted as specified, adapt to the individual learner only within bounds the teacher sets, and bring the session to its planned completion. This is demanding because the model performs two roles at once: within an activity it is the instrument the student works with; across the lesson it is the tutor directing that work and holding the lesson to its course. Whether a programme is delivered as designed is the subject of an established literature termed fidelity of implementation (Stains and Vickrey, 2017).

These demands must be met under technical, resource and user-experience constraints. Technically, the platform must be capable of ingesting a prompt, delivering the lesson as intended and providing a means for activity to be reported to the teacher, and it must not conflict with institutional IT policy. On resource constraints, the format must be affordable to every party and represent value for money. Finally, the user experience must be sufficiently intuitive and easy to use such that it does not derail the lesson.

3. Arriving at the Technical Format

Initial investigations addressed the technical question of how students should access an LLM. Three options merited serious inquiry.

The first was to develop a bespoke platform sitting between the learner and the LLM. This method offers the greatest control as the interface can be specific to this context; transcripts resulting from the lesson would be

easily accessible; a student's progress through the lesson could be continually monitored and interventions could be made to keep a lesson on track. It is, however, the most complex and expensive option.

The second option considered was to embed a lesson configuration within the LLM platform provided by the institution. This would have the advantage of sitting within existing IT infrastructure and, as per option one, offer significant control over the lesson delivery. However, given the sensitivities surrounding so new a technology, engagement with institutional ICT would introduce significant organisational complexity.

The third, the prompt-as-lesson, sends the lesson to students as a prompt they run in their own LLM. This approach is relatively inexpensive and simple to implement. In addition, students use tools they already hold and do not need to learn a new interface. Reporting, however, is manual (the student must submit a report of the conversation) and control ends at the specification: unlike the first two options, the lesson cannot be monitored at run time. Provider-hosted configurations such as custom GPTs are considered here to be a delivery variant of the prompt-as-lesson.

The prompt-as-lesson was selected. It offers the least run-time control, but simplicity of implementation and reach into students' existing tools were, on balance, considered decisive.

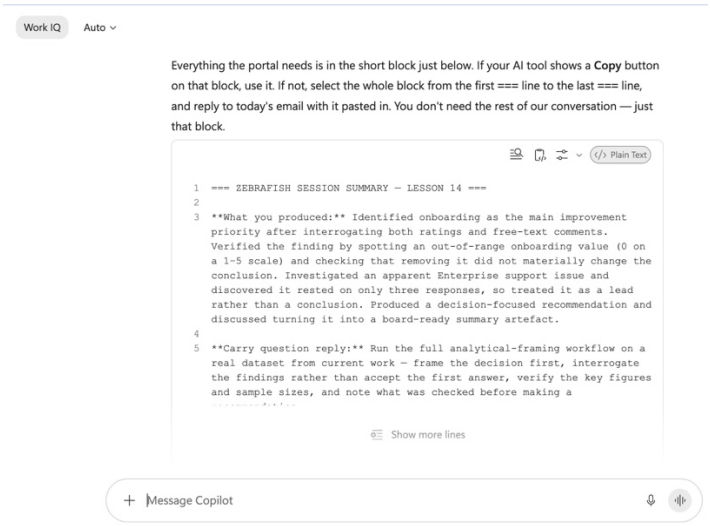


Figure 1: The close of an author test session in Microsoft Copilot: the summary block returned for submission

4. Formative Development

Once the technical approach was determined, the focus moved to the form the prompt-as-lesson should take. A format emerged from the development of prompt-as-lessons on two modules differing in audience, content and setting. The first, 'Sparks', is a set of self-contained micro-lessons on the topic of AI fluency for use in executive education. The second, the 'AI Companion', was developed for a ten-week strand of a postgraduate entrepreneurship module at a UK business school, in which nine weekly prompt-as-lessons each guide the student through a 30–45-minute tutorial applying that week's teaching to the student's venture.

The resulting format resembles a lesson plan stated in natural language but written knowing the teacher executing the plan will be an LLM. The specification was developed by iteration: scripts were drafted, run in repeated informal tests across the LLMs students typically use (ChatGPT, Gemini, Claude and Copilot), and revised wherever the model's behaviour departed from the intended plan. The revision history forms the development record on which Sections 4 to 6 draw. The design responses that follow, and those summarised in Table 1, double as authoring rules. Each records not only what failed but what an author must now write. Two main themes emerged.

4.1 The Model did not Adhere to an Unstructured Lesson

The central finding concerns structure. Early scripts were flat lists of desired learning outcomes and the model would not hold to them. It skipped ahead and followed tangents, seemingly preferring to pursue its own exploratory path.

The response was to write the lesson plan as a series of 'acts'. More specifically, an act is a named phase with a purpose, a pedagogic mode and a rough duration, and the acts are arranged in a fixed and deliberate order. As examples, an act may comprise a knowledge diagnosis, the application of new knowledge or a test of understanding. Figure 2 reproduces one act from the AI Companion.

> ACT 2 — WHY YOU, WHY NOW
>
> *(6–8 minutes)*
>
> Ask: "Why do you want to build a venture? Not the polished version — the real one. …"
>
> Wait for the answer.
>
> [IF THE ANSWER IS SPECIFIC AND HONEST — names a real driver …]: Reflect it back in one line and affirm the specificity. Then probe one level deeper. …
>
> [IF THE ANSWER IS GENERIC — "I want to be my own boss", "make an impact"]: Name it gently. …
>
> Wait. If the answer is thin, probe once. Then move to Act 3.

Figure 2: One act, verbatim, from the first AI Companion lesson (ellipses mark cuts)

4.2 Perspective: the Model's Failure to Adopt the Learner's Position

The development process revealed a further general limit: the model could not put itself in the learner's place. This limitation revealed itself in four ways. First, the model did not recognise the need for signposting across the acts. Notably, introducing more structure into the prompt led to a less coordinated experience at the lesson level. The learner experience was less fluent than when the model was left to take an exploratory approach. The commencement of each act frequently seemed abrupt, its connection with the previous act left unclarified. Structure was bought at the cost of fluency. In response, the segues between the acts and the required instructional text were written into the prompt.

Second, the LLM seemed unable to register struggle. When given a specific instruction such as 'ask the student for an example of', the LLM would adhere to the instruction overdiligently and continue past the point at which a teacher would naturally recognise struggle and change tack.

Third, the model proved poor at estimating the time an interaction would require, sometimes asking the learner to complete substantial work in minutes and at others overestimating heavily.

Fourth, the model's failure to adopt the learner's position impacted pedagogy. While seemingly fluent in educational psychology and the principles of learning design, the LLM struggled to determine an appropriate pedagogy for each act. In response, the activity type and pedagogical approach for each act are explicitly stated in the prompt. The process of determining pedagogy for acts identified a repertoire of activity types that seek to capitalise on the affordances of the LLM. These include Socratic questioning, didactic exposition, role-played practice, application to the learner's own material and a consolidating review.

An inability to adopt the learner's position also revealed itself in the testing phase. One component of testing was the creation of 16 synthetic AI user personas, each prompt-as-lesson script being run twice with each persona. However, this proved a poor replacement for testing with humans: the personas behaved as fluent, cooperative learners and did not reproduce the hesitation, misreading and struggle that informal piloting surfaced.

4.3 Further Findings and Design Responses

Further problems identified during the development phase are summarised in Table 1 together with the resulting change to the design of the prompt format.

Table 1: Problems met in development, and the responses encoded in the specification

Finding	Problem Observed	Design Response
Broken turn-taking	Iterations conflated together; content followed by silence; turns ended in the wrong places.	One question or observation per response, each ending with a handoff; turn management written into the acts.
Indiscriminate praise	The default tutor voice was sycophantic: unearned praise and flattery devalued feedback.	A fixed teaching persona, defined partly by an explicit list of what the voice never does (Figure 3).
Digressions	Students asked digressive questions, for example administration questions, mid-lesson which drew the model away from the lesson plan.	Instruction on meta-question handling: answered in one sentence, then the lesson resumes.
Completing the learner's work	After setting a task, the model too quickly produced the finished product itself; the learner observed rather than worked.	Withholding rules: ask before telling, caps on output length.
Voice confusion	The model moved unannounced between a practice role and the teaching voice; learners lost track of which was speaking.	Role signposting: entry into and exit from a practice role is announced.
Privacy concerns	Learners' real work carried names, clients and commercial detail into the conversation, and thence into the report.	Privacy rules stated within the lesson: people described by role, not name.
Broken completion	Lesson ending before completion or not ending at all.	Introduction of the keyword commands 'Continue' and 'End'.
Reviewing instead of delivering	When text preceded the persona (a title, a covering note), the model treated the lesson as a document to critique rather than a lesson to deliver.	The prompt begins with the persona; nothing precedes it.

5. The Resultant Format: Structure and the Control Surface

5.1 The Control Surface

The development process yielded what is termed here the control surface: the fourteen dimensions an educator must specify for the prompt-as-lesson to proceed as intended (Table 2).

Table 2: The control surface: fourteen dimensions an educator specifies in a prompt-as-lesson

Category	#	Dimension	What the Educator Specifies
Participants	1	Teacher persona	The teaching voice, its character and its prohibitions
	2	Student profile	The learner-specific state injected each session: name, cumulative document, carry question
	3	Expected student engagement	What is expected of the student: how they are expected to interact with the tutor and with the lesson's materials
Curriculum and pedagogy	4	Context	The lesson's place in a wider programme
	5	Learning outcomes	The desired lesson-level learning outcomes
	6	Teaching model	The lesson-level pedagogic patterns to be adopted (optional)
	7	Knowledge base	Description of the knowledge base to be adopted (e.g. theoretical framework, data, case studies). Include files if technology allows
	8	Interaction rules	Per-response conduct: one question per turn, one probe only, the handoff
Rules	9	Withholding	What assistance is refused, when probing stops, what is never performed for the learner

Category	#	Dimension	What the Educator Specifies
	10	Boundary handling	How off-lesson queries are contained and returned. Meta-questions: one-sentence answers to course questions, and the return to the lesson
	11	Controls	For example, the End command
Structure	12	Sequence and pacing	The ordered acts through which the session moves, and the time given to each
	13	Feedback	The closing feedback given to the student
	14	Reporting	What must be reported back to the educator, in what form

5.2 The Structure of the Prompt-as-Lesson

Fundamentally, the prompt-as-lesson is a text document of two parts (Table 3): the control surface, comprising standing instructions to the model, and the script, specifying the components of the lesson experienced by the student.

Note that the structure is stated in the control surface and restated within each act. Development showed the model holds to a lesson more closely when its path is explicit, and this repetition is designed to reinforce the path through the session.

Table 3: The structure of a prompt-as-lesson: eight sections in two parts, with the dimensions of Table 2 located in italics

Part	#	Section	Carries
The control surface	1	Participants	Teacher persona; Student profile; Expected student engagement
	2	Curriculum and pedagogy	Context; Learning outcomes; Teaching model; Knowledge base; Interaction rules
	3	Rules	Withholding; Boundary handling; Controls
	4	Structure	Sequence and pacing; Feedback; Reporting
The script	5	Opening	The prescribed welcome: greeting by name, the opening line, statement of goals, how the session works, its length (*Sequence and pacing*), the Continue recovery command, the readiness check
	6	Acts	Each act is a named teaching phase. The prompt states: the goal; the target duration; the pedagogic mode (*Teaching model*); its place in the lesson's sequence; the opening move, often a verbatim question; explicit wait and hand-off instructions; branches prescribing the response to a strong, thin, stuck or ahead answer; the one-probe limit (*Withholding*); the standard to be achieved before the act closes (*Expected student engagement*); and the segue to the next act.
	7	Close	The prescribed ending: naming what was done, a carry question, a bridge to the next lesson, instruction to submit the End command
	8	Transmission	What is produced when the End command is triggered: confirmation of lesson end, a delimited summary block (*Reporting*), instructions for submitting the report, and optionally a feedback question (*Feedback*)

Each of these sections is written in plain language. Figure 2 above reproduces one act; Figure 3 below shows an extract from the Persona section.

> Your job is not to have the ideas for the students. It is to make them think harder, see more clearly, and be honest with themselves about their venture. … You are not a cheerleader and you are not a critic. You are the most useful half-hour of thinking they will do this week.

Figure 3: The voice: from the teaching persona, reproduced in every lesson (ellipses mark cuts)

Figure 4 below reproduces the transmission section which acts as the reporting mechanism. The scope of the return is itself a choice. The lessons in the Sparks programme return a summary block alone, as shown here, while the AI Companion requires the whole conversation to be returned.

WHEN [STUDENT_NAME] TYPES END

Complete the steps below before the session ends: the recap, then the submission block. …

“Your role-assignment prompt saves to your Playbook the moment you submit. Everything the portal needs is in the short block just below. … Reply to the email you got this morning with the block pasted in. You don’t need the rest of our conversation — just that block.”

Then output the block. … with exactly these lines between the fences:

=== ZEBRAFISH SESSION SUMMARY — LESSON 2 ===

What you produced: [two or three lines, specific to THIS person — what they noticed in the three role runs, the task they chose for Act 4, and the role they wrote.]

Carry question reply: [the specific task they committed to in Act 5 Part B]

File: role-assignment.md

[The role-assignment template the participant built …]

=== END SUMMARY ===

Do not add anything after the END SUMMARY line. The session ends here.

Figure 4: The transmission, from a Sparks lesson: the End sequence and the summary-only return (ellipses mark cuts)

6. A Design Method for the Prompt-as-Lesson

The development reported in Sections 4 and 5 yields a repeatable design method comprising four phases. It stands in the lineage of ADDIE (Molenda, 2003) and of learning design in higher education (Laurillard, 2012), but is specific to a lesson enacted by a model the designer does not control.

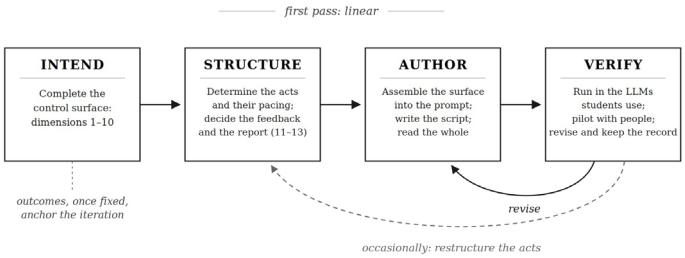


Figure 5: The four-phase design method

In the Intend phase, the educator completes dimensions 1–11 of the control surface. In the Structure phase they determine the acts and their pacing, the feedback and the report (dimensions 12–14). In the Author phase the surface is assembled into the prompt and the script is written. In the Verify phase the artefact is run in LLMs, piloted with people, and revised until enactment matches specification.

The phases are linear only on first pass (Figure 5): Verify returns the design to Author, and occasionally to Structure, in repeated cycles, while the outcomes fixed in Intend anchor the iteration.

Much of this is not new and adopts the common pedagogy relating to the creation of a lesson plan. Outcomes before content, pedagogy matched to content, signposting, learning activities and assessment aligned with outcomes (Biggs, 1996). What is new falls into four classes: techniques that counter the model's dispositions, techniques that exploit them, techniques that address the model as software, and the process of verification.

7. Limitations and Further Work

The control surface, structure and authoring method reported here are formative. They arise from an exploration of the prompt-as-lesson format across two modules and no claim is made for generalisability to other subject matters or learning contexts. The paper also reports on a single format of lesson plan. However, lesson plans vary considerably in form. Therefore the conclusions drawn here should be applied to other formats with caution. In addition, this report reflects the capabilities of common frontier models in the summer of 2026 and those capabilities are likely to evolve.

The project's next phase is a formal evaluation of the format's fidelity of implementation: whether the delivered lesson is the authored lesson. Learning will not be measured in this study, as fidelity is logically prior in that no learning effect can be attributed to a lesson until it is established that the lesson was delivered. Following Stains and Vickrey (2017), the evaluation will ask whether the critical components of the authored lesson are present when it is enacted. This decomposes into four research questions, one per category of their framework:

RQ1 (Procedural fidelity): are the specified parts of the lesson present and in the intended order? This is measured by whether the lesson opened as specified, whether each act was delivered as planned, whether the acts followed the specified order, and whether the lesson ran to completion and closed as specified.

RQ2 (Educative fidelity): is the content taught the teacher's specified content? This is measured by whether the specified content was delivered where the lesson called for it, whether stated boundaries were held, whether any stated curriculum position was contradicted, and by a count of unanticipated substitutions.

RQ3 (Pedagogical fidelity): does the model behave as the specified tutor? This is measured by whether the teaching persona was held, whether the pedagogy of each act was as specified, and whether the stated rules were kept.

RQ4 (Student engagement): do students engage as the lesson expects? This concerns participation in the lesson as delivered rather than what students learned, and is measured by whether the stated expectation of student engagement was met.

The context will be 180–200 students taking nine prompt-as-lessons on a postgraduate entrepreneurship module from September to December 2026. A sample of transcripts, stratified across lessons and across the tools students use, will be coded against a checklist derived from each lesson. Subsequent work will examine the effectiveness of the format, measured against the achievement of learning outcomes.

The derivation of generalisable design principles will further require investigation across different pedagogies, subject matters and lesson lengths.

8. Summary and Conclusion

This paper has reported the design of the prompt-as-lesson. This format encodes a teacher-authored lesson as a single prompt, which is enacted as a one-to-one tutorial by the student's own LLM tool. The resulting dialogue is returned to the teacher in a closing report.

During development, the central discoveries were that contemporary LLMs struggled to adopt the learner's perspective and would not hold to an unstructured lesson. The format that emerged addresses these limitations. The contribution of this paper is threefold: a control surface, a structure and an authoring method, each designed to protect the fidelity of the enacted lesson.

The work demonstrated that a lesson specified by the teacher can be authored for, and enacted by, an LLM cast as the tutor. Whether enactment remains faithful to the specification, its fidelity of implementation will be the subject of an evaluation beginning September 2026. The artefacts are offered so that any educator can utilise the approach and progress this line of research.

Ethics Declaration

This paper reports the design and development of teaching materials; it reports no participant data. Early lessons were trialled informally as part of normal course development; no data from those sessions is analysed or reported here, and the findings of Section 4 are drawn from the author's own development record and test runs. The illustrative session in Figure 1 is the author's own test run. An application covering the associated evaluation study was submitted to Imperial College London's Education Ethics Review Process in July 2026 and is subject to review; it will analyse pseudonymised teaching data from consenting students.

AI Declaration

AI tools (Anthropic's Claude) were used in the preparation of this paper for brainstorming, editing and proofreading, under the author's direction. The artefact under study was itself developed in dialogue with AI tools, and the development record drawn on in Sections 3 and 4 includes those exchanges. The construct, the study design, the theoretical argument and the final text are the author's. No participant data was processed by AI tools.

Acknowledgements

The author thanks the pilot participants and the module's programme team.

References

Biggs, J. (1996) 'Enhancing teaching through constructive alignment', *Higher Education*, 32(3), pp. 347–364. doi: 10.1007/BF00138871.

Bjork, R.A. (1994) 'Memory and metamemory considerations in the training of human beings', in Metcalfe, J. and Shimamura, A.P. (eds.) *Metacognition: knowing about knowing*. Cambridge, MA: MIT Press, pp. 185–205.

Bloom, B.S. (1984) 'The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring', *Educational Researcher*, 13(6), pp. 4–16. doi: 10.3102/0013189X013006004.

Kapur, M. (2008) 'Productive failure', *Cognition and Instruction*, 26(3), pp. 379–424. doi: 10.1080/07370000802212669.

Kestin, G., Miller, K., Klales, A., Milbourne, T. and Ponti, G. (2025) 'AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting', *Scientific Reports*, 15. doi: 10.1038/s41598-025-97652-6.

Laurillard, D. (2002) *Rethinking university teaching: a conversational framework for the effective use of learning technologies*. 2nd edn. London: RoutledgeFalmer.

Laurillard, D. (2012) *Teaching as a design science: building pedagogical patterns for learning and technology*. New York: Routledge.

Lee, D. and Palmer, E. (2025) 'Prompt engineering in higher education: a systematic review to help inform curricula', *International Journal of Educational Technology in Higher Education*, 22, 7. doi: 10.1186/s41239-025-00503-7.

Molenda, M. (2003) 'In search of the elusive ADDIE model', *Performance Improvement*, 42(5), pp. 34–36. doi: 10.1002/pfi.4930420508.

Mollick, E.R. and Mollick, L. (2023) 'Assigning AI: seven approaches for students, with prompts'. SSRN. doi: 10.2139/ssrn.4475995.

Mollick, E.R. and Mollick, L. (2024) 'Instructors as innovators: a future-focused approach to new AI learning opportunities, with prompts'. SSRN. doi: 10.2139/ssrn.4802463.

Qian, Y. (2025) 'Prompt engineering in education: a systematic review of approaches and educational applications', *Journal of Educational Computing Research*, 63(7–8), pp. 1782–1818. doi: 10.1177/07356331251365189.

Stains, M. and Vickrey, T. (2017) 'Fidelity of implementation: an overlooked yet critical construct to establish effectiveness of evidence-based instructional practices', *CBE—Life Sciences Education*, 16(1), rm1. doi: 10.1187/cbe.16-03-0113.

Stephenson, R. and Armstrong, C. (2026) *Student generative AI survey 2026*. HEPI Report 199. Oxford: Higher Education Policy Institute.

VanLehn, K. (2011) 'The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems', *Educational Psychologist*, 46(4), pp. 197–221. doi: 10.1080/00461520.2011.611369.

Beyond the Standalone AI: Embedding Educational Chatbots Through Scaffolded Pedagogical Design

Colin Loughlin
Brunel University of London, Uxbridge, UK

colin.loughlin@brunel.ac.uk

Abstract: Generative AI chatbots are rapidly being introduced into higher education as institutional learning technologies. However, much current implementation assumes that educational value emerges primarily through access to the technology itself. This paper argues that AI chatbots are pedagogically weak when deployed as standalone resources and that meaningful educational impact depends upon deliberate integration into learning design, teaching practice, and assessment structures. The paper presents a case study from a UK university initiative involving *BrunOwl*, an institutionally supported AI chatbot embedded within the virtual learning environment. Rather than positioning the chatbot as a generic student support tool, the initiative adopts a scaffolded pedagogical integration model led by a central Digital Education team. Academic staff complete an online curriculum-design form capturing disciplinary context, learning challenges, threshold concepts, assessment relationships, cohort characteristics, and intended uses of AI within teaching. These submissions are then used collaboratively to develop structured implementation plans tailored to specific modules and programmes. The paper examines how this process reframes AI adoption from technological deployment to educational design. Drawing on concepts from constructive alignment and socio-technical approaches to educational change, the study explores how academic partnership, contextualisation, and structured scaffolding mediate the effectiveness of AI-supported learning. Early findings from institutional implementation suggest that staff are more likely to integrate AI meaningfully when supported through iterative pedagogical dialogue rather than technical training alone. Similarly, students appear more likely to engage critically and productively with AI when chatbot activities are explicitly aligned with learning outcomes, teaching activities, and assessment expectations. The paper contributes a transferable institutional model for embedding generative AI within higher education curricula while also offering a critical perspective on the prevailing "provision-led" discourse surrounding AI in education. It will be of relevance to researchers, educational developers, learning technologists, and institutional leaders seeking sustainable and pedagogically grounded approaches to AI-enabled learning.

Keywords: Generative AI, AI Chatbots, Higher Education, Learning Design, Constructive Alignment, Socio-Technical Change

1. Introduction

Generative artificial intelligence has been adopted by higher education institutions with remarkable speed. In the space of little more than three years, universities have developed strategies, established governance groups, rewritten assessment regulations, and procured institutional AI tools. Student-facing chatbots have emerged as an almost ubiquitous response. They appear scalable, comparatively inexpensive, and offer the possibility of personalised support at a time when many universities of mass education face diminishing resources.

The speed of adoption has produced a familiar discourse for those of us involved with educational technologies. Much of the literature and institutional rhetoric surrounding generative AI appears to assume that educational value emerges primarily through access to the technology itself (e.g., Quality Assurance Agency for Higher Education, 2024). While universities and sector bodies rarely claim explicitly that AI provision alone will improve learning, the language of institutional AI adoption often implies a direct relationship between access to generative AI and educational enhancement, particularly through claims of personalised tutoring, 24/7 support, critical thinking and scalable feedback.

The history of educational technology offers reasons to be sceptical of such claims. The educational significance of technology has repeatedly been overstated. Radio, television, personal computers, virtual learning environments, MOOCs, and learning analytics have all, at various times, been heralded as transformative innovations that would fundamentally alter higher education (Loughlin, 2017). Yet universities remain remarkably resilient institutions, and the relationship between technological change and pedagogical practice is neither linear nor deterministic.

As Dewey observed over a century ago, educational practices often remain stubbornly resistant to theoretical and technological innovation (Dewey, 2012). Similarly, Selwyn (Selwyn, 2022a) cautions against technological solutionism and argues that educational technologies must be understood as socially situated and politically constructed rather than inherently transformative. The hype, hope, and disappointment of educational technologies and innovation have a long history. There is still an 'underlying assumption [within universities] that once adequate resources [are] obtained, integration [will] follow' (Ertmer, 1999, p. 50). Yet, it has been shown that access to technologies and the provision of skills workshops do not, alone, lead to an increased take-

up of learning technologies amongst academic staff (Kirkwood and Price, 2014; Bennett, 2010; Surry and Land, 2000); and, there is evidence that 'even if every [technical and physical] barrier were removed teachers would not automatically use technology' (Ertmer, 1999, p.51). Many studies have identified the link between teacher's belief systems and the integration of TEL into teaching practice and offered strategies to help overcome these (Newman, 2002; Cuban et al., 2001; Ertmer, 1999). Yet, the majority of institutional policies and TEL initiatives remain doggedly focused on the tangibles of skills training and technical provision.

The purpose of this paper is not to critique generative AI itself, nor to argue against the use of educational chatbots. Rather, it is to challenge the assumption that AI technologies possess educational value independent of their pedagogical context.

This paper argues that AI chatbots are pedagogically weak when deployed as standalone resources (OECD, 2026) and that meaningful educational value emerges through intentional integration into curriculum design, teaching practice, and assessment (Hon, 2026). And that without direct support for academics to help them in understanding AI in their context (and how to engage their students with it meaningfully), there is a strong chance AI will not achieve its potential in learning and teaching. Using a case study of the implementation of BrunOwl, an institutionally supported AI chatbot embedded within a UK university's virtual learning environment, the paper explores how a central Digital Education team sought to mediate AI adoption through a scaffolded academic partnership model.

The paper suggests that the critical work of educational innovation lies not in providing AI tools, but in creating the pedagogical conditions under which AI may become educationally meaningful.

2. Educational Technology and the Problem of Provision

The history of educational technology is, in many respects, a history of unfulfilled promises. Almost every significant technological development has been accompanied by claims that it would transform learning and render previous pedagogies obsolete. Laurillard et al (2008) said that 'education is on the brink of being transformed through learning technologies; however, it has been on that brink for some decades now'.

Generative AI may be in danger of providing another iteration of this recurring cycle. Educational technologies are frequently discussed as though pedagogical outcomes arise from the intrinsic properties of the technologies themselves. Such assumptions reflect a form of technological determinism in which technologies are treated as causal agents of educational change (Selwyn, 2022b; Oliver, 2011). As Feenberg observes, educational technologies are often imagined as 'technologies that do the educating' (2017, p. 364), obscuring the social, disciplinary, and pedagogical work through which educational value is actually produced.

Constructive alignment (Biggs and Tang, 2011) offers one way of understanding this relationship. Learning outcomes, learning activities, and assessment should form a coherent system. Technologies are not exempt from this requirement. An AI chatbot that exists independently of teaching activities and assessment structures is no more pedagogically meaningful than an uncommented collection of documents often seen on university VLE pages. Like the lecture theatres of mass higher education, AI chatbots may become ubiquitous features of the educational landscape while remaining largely pedagogically unexamined (Loughlin, 2025).

Yet there is already evidence that many institutional responses to AI are characterised by provision-led approaches. Universities procure tools and then encourage staff and students to "make use" of them. History has demonstrated that providing tools and staff development workshops has little impact on meaningful integration of educational technologies into teaching practice (Loughlin, 2017). The assumption that educational practices will naturally emerge around technological provision mirrors earlier approaches to virtual learning environments and learning analytics, where implementation frequently prioritised technical capability over pedagogical purpose. The result is often a form of unintentional practice, in which technologies become normalised but pedagogically weak.

3. AI and Pedagogical Intentionality

The educational promise of generative AI lies primarily in its conversational capabilities. Unlike previous educational technologies, large language models can engage in dialogue, generate explanations, ask questions, adapt responses, and present information in multiple ways. They possess characteristics that are potentially compatible with constructivist and student-centred approaches to learning.

The potential for personalised tutoring is particularly attractive. Bloom's (1984) "two sigma problem" demonstrated that students receiving one-to-one tutoring substantially outperformed those receiving conventional classroom instruction. Generative AI has therefore prompted renewed interest in the possibility of scalable, personalised learning support (Hippel, 2024).

However, the existence of conversational capabilities does not guarantee educational value. An AI chatbot can provide explanations, summaries, and answers, but equally it can encourage superficial engagement, shortcut productive struggle, and reinforce passive approaches to learning (e.g., Bastani et al., 2025). Whether AI functions as a tutor, a study companion, a writing assistant, or an answer machine depends largely upon how (or if) it is integrated into teaching practice.

This distinction between capability and educational function is important. The educational significance of AI resides not in what the technology can do, but in what students are invited, encouraged, and expected to do with it. The question, therefore, is not whether universities should provide AI chatbots, but how those chatbots become pedagogically situated within learning and teaching.

4. Context: BrunOwl

BrunOwl is an institutionally supported AI study buddy chatbot embedded within the university's virtual learning environment. The original ambition was to provide students with personalised support, opportunities for dialogue, and an accessible study companion capable of engaging with disciplinary content. BrunOwl was piloted in 2024-5 and implemented institution-wide for the start of the academic year 2025-6. Around 60% of students have accessed it, representing all colleges and disciplines. By the end of semester one, over 12,000 conversations had been logged.

On the face of it, these interactions were in-line with our expectations, and the take-up could be considered a success. However, early analysis of the chatbot conversations revealed primarily low-level and unstructured chats. This highlighted a familiar concern, merely making the chatbot available to students did not constitute meaningful educational innovation. There was little reason to assume that students would automatically use the chatbot in ways that promoted meaningful learning.

Consequently, the next phase of the implementation strategy focused on how to engage academics to embed AI into their modules in a way that would develop the AI literacy of their students and enhance their learning. The Digital Education team has moved beyond technological provision and towards pedagogical integration by developing simple online form which staff could complete to identify key aspects of their module and students which would be the start of a co-design process for integrating AI into their learning design.

The online form captures:

- Module context and learning outcomes;
- Student cohort size and characteristics;
- Learning challenges and concepts;
- Assessment requirements;
- Intended educational uses of ai;
- Ethical and disciplinary considerations.

The form functions as a pedagogical intervention in its own right. Staff are encouraged to articulate assumptions about learning, identify areas of student difficulty, and consider how AI might contribute to existing teaching activities. The process resembles curriculum design more than software implementation. The very act of expressing the intended learning outcomes and areas that students struggle with can be powerful in its own right, leading to improved learning design.

The answers from the online form are fed into a specially designed AI agent that also has access to the module content and assessment resources; it is designed to generate a teaching plan with step-by-step instructions for academic staff and detailed prompts for students using the principles of constructive alignment and active learning. This first-draft of a learning design is then used by the Digital Education team to co-design with academic staff a scaffolded integration plan.

These plans typically include:

- Module-specific prompt suggestions;
- Preparatory activities;

- Formative exercises;
- Reflective tasks;
- In-class follow-up activities;
- Guidance on ethical and transparent ai use;
- Alignment with learning outcomes and assessment.

The activities are designed to draw out the students existing understanding of key topics and build from there, while also developing and understanding of the utility and limits of AI in their disciplinary context. The emphasis is therefore not on using AI, but on designing educational experiences in which AI serves identifiable pedagogical purposes.

5. Discussion

The BrunOwl experience suggests that the implementation of educational AI is fundamentally a socio-pedagogical process. Three observations are noteworthy. Firstly, academic staff frequently understand generative AI at a general level but struggle to translate that understanding into concrete teaching practices. Technical demonstrations and generic training sessions often fail to bridge this gap. The challenge is not technological literacy alone, but pedagogical operationalisation. Secondly, the design form itself functions as a reflective mechanism. It prompts academics to make explicit their assumptions regarding student learning, threshold concepts, and assessment. In doing so, it encourages forms of pedagogical intentionality that might otherwise remain tacit. Thirdly, meaningful uses of AI appear highly contextual.

The implementation strategies developed for research methods, economics, occupational therapy, and nursing differ substantially. The educational value of AI is therefore inseparable from disciplinary context, learning outcomes, and assessment structures.

These observations challenge deterministic narratives surrounding generative AI. Educational technologies do not possess pedagogical agency. Universities often speak of technologies "transforming" learning, yet technologies transform little in isolation. Rather, educational change emerges through the decisions, practices, and relationships that surround technological adoption. AI differs from many previous technologies in that the work involved in integrating AI into the curriculum does not have to be onerous, and the rewards in terms of the potential for student learning are great.

Higher education's current adoption of AI resonates with broader critiques of educational technology. Technologies frequently fail to deliver transformative outcomes because institutions mistake provision for implementation. The chatbot itself is only a component of the educational intervention. The more significant intervention may be the pedagogical work undertaken by academics and educational developers to determine how the chatbot should function within particular learning environments.

This reframing also suggests an expanded role for educational developers. Rather than acting primarily as technical support staff or trainers, educational developers become translators and mediators between technological capability and pedagogical practice. Their work involves helping academic colleagues articulate educational purposes, align activities with outcomes, and develop contextually appropriate implementations. In this sense, pedagogical integration becomes a process of academic partnership.

6. Conclusion

The rapid adoption of generative AI has created understandable enthusiasm (along with many concerns) within higher education. Educational chatbots offer the possibility of personalised learning support at unprecedented scale. However, the history of educational technology cautions against assuming that access to new technologies necessarily produces educational improvement.

This paper has argued that AI chatbots are pedagogically weak when deployed as standalone resources. Their educational value emerges through intentional integration into learning design, assessment structures, and disciplinary contexts. The case of BrunOwl suggests that the critical work of educational innovation lies not in providing AI systems but in creating the pedagogical conditions under which those systems may contribute meaningfully to learning.

Universities have repeatedly demonstrated a tendency to mistake technological provision for pedagogical transformation. Generative AI may prove no different. The challenge facing higher education is therefore not

whether institutions can provide educational AI, but whether they possess the pedagogical structures and partnerships necessary to make that AI educationally meaningful.

Ethics Declaration

The online ethics portal indicated that ethical approval was not required for this research.

AI Declaration

The author used generative artificial intelligence (ChatGPT, OpenAI) during the preparation of this manuscript to support structural development, language refinement, and editorial review. The AI system was not used to generate original empirical findings, conduct analysis, interpret results, or formulate the paper's arguments independently. All substantive intellectual contributions, interpretations, conclusions, and final editorial decisions remain the responsibility of the author.

References

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. and Mariman, R. (2025) 'Generative AI without guardrails can harm learning: Evidence from high school mathematics', Proceedings of the National Academy of Sciences, vol. 122, no. 26, p. e2422633122 [Online]. DOI: 10.1073/pnas.2422633122 (Accessed 28 June 2026).

Bennett, E. (2010) 'Activity theory: what does it offer elearning research?', Manchester [Online]. Available at http://eprints.hud.ac.uk/10631/ (Accessed 10 May 2017).

Biggs, J. and Tang, C. S. (2011) Teaching for quality learning at university: what the student does, SRHE and Open University Press imprint, 4. ed., Maidenhead, McGraw-Hill, Society for Research into Higher Education & Open University Press.

Bloom, B. S. (1984) 'The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring', Educational researcher, Sage Publications Sage CA: Thousand Oaks, CA, vol. 13, no. 6, pp. 4–16.

Cuban, L., Kirkpatrick, H. and Peck, C. (2001) 'High Access and Low Use of Technologies in High School Classrooms: Explaining an Apparent Paradox', American Educational Research Journal, vol. 38, no. 4, pp. 813–834 [Online]. DOI: 10.3102/00028312038004813 (Accessed 17 December 2014).

Dewey, J. (2012) Democracy and education: an introduction to the philosophy of education., Brisbane, Emereo Publishing.

Ertmer, P. A. (1999) 'Addressing first- and second-order barriers to change: Strategies for technology integration', Educational Technology Research and Development, vol. 47, no. 4, pp. 47–61 [Online]. DOI: 10.1007/BF02299597 (Accessed 29 October 2014).

Feenberg, A. (2017) 'The Online Education Controversy and the Future of the University', Foundations of Science, vol. 22, no. 2, pp. 363–371 [Online]. DOI: 10.1007/s10699-015-9444-9 (Accessed 19 June 2026).

Hippel, P. T. von (2024) 'Two-Sigma Tutoring: Separating Science Fiction from Science Fact', Education Next [Online]. Available at https://www.educationnext.org/two-sigma-tutoring-separating-science-fiction-from-science-fact/ (Accessed 29 January 2025).

Hon, K. (Kai L. (2026) 'Generative AI in Higher Education: A Systematic Review of Its Effects on Learning Outcomes and Academic Performance', Journal of Educational Technology Systems, vol. 54, no. 3, pp. 537–560 [Online]. DOI: 10.1177/00472395251400089 (Acces'Technology-enhanced learning and teaching in higher education: what is "enhanced" and how do we know? A critical literature review sed 11 June 2026).

Kirkwood, A. and Price, L. (2014) 'Technology-enhanced learning and teaching in higher education: what is "enhanced" and how do we know? A critical literature review', Learning, Media and Technology, pp. 6–36 [Online]. DOI: 10.1080/17439884.2013.770404 (Accessed 20 February 2013).

Laurillard, D., Kolokitha, M., Mellar, H., Selwyn, N. and Noss, R. (2008) Learning Through Life: The role of Technology, Foresight Report, Office of Science and Innovation [Online]. Available at http://ble.ioe.ac.uk/pluginfile.php/47879/mod_resource/content/1/Lifelong%20Learning%20and%20Technology%202008.pdf (Accessed 1 February 2013).

Loughlin, C. (2017) 'Staff Perceptions of Technology Enhanced Learning in Higher Education', Proceedings of the 16th European Conference on eLearning, ECEL2017, Porto, Academic Conference Publishing International, pp. 335–343.

Loughlin, C. (2025) 'The Unintentional Lecture: The place of large-class teaching in contemporary Higher Education', Lund Studies in Educational Sciences, Doctoral Thesis (compilation), Department of Educational Sciences [Online]. Available at https://portal.research.lu.se/files/206546831/Colin_Loughlin_-_ALL.pdf.

Newman, C. F. (2002) 'A cognitive perspective on resistance in psychotherapy', Journal of Clinical Psychology, vol. 58, no. 2, pp. 165–174 [Online]. DOI: 10.1002/jclp.1140 (Accessed 25 May 2017).

OECD (2026) OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education, OECD Digital Education Outlook, OECD Publishing [Online]. DOI: 10.1787/062a7394-en (Accessed 10 April 2026).

Oliver, M. (2011) 'Technological determinism in educational technology research: some alternative ways of thinking about the relationship between learning and technology: Educational technology and determinism', Journal of Computer Assisted Learning, vol. 27, no. 5, pp. 373–384 [Online]. DOI: 10.1111/j.1365-2729.2011.00406.x (Accessed 19 June 2026).

Quality Assurance Agency for Higher Education (2024) 'Quality Compass: Navigating the Complexities of the Artificial Intelligence Era in Higher Education', Gloucester, Quality Assurance Agency for Higher Education [Online]. Available at https://www.qaa.ac.uk/docs/qaa/news/quality-compass-navigating-the-complexities-of-the-artificial-intelligence-era-in-higher-education.pdf (Accessed 12 June 2026).

Selwyn, N. (2022a) Education and technology: key issues and debates, Third edition., London ; New York, Bloomsbury Academic.

Selwyn, N. (2022b) Education and technology: key issues and debates, Third edition., London New York Oxford New Delhi Sydney, Bloomsbury Academic.

Surry, D. W. and Land, S. M. (2000) 'Strategies for Motivating Higher Education Faculty to Use Technology', Innovations in Education & Training International, vol. 37, no. 2, pp. 145–153 [Online]. DOI: 10.1080/13558000050034501 (Accessed 10 May 2017).

From Technology to Teaching Practice: A Didactic Roadmap for Emerging Tech in Education

Gunver Majgaard, Regina Lamscheck-Nielsen, Maria Skougaard Andersen and Rasmus Ploug
University of Southern Denmark, Odense, Denmark

gum@mmmi.sdu.dk
myre@sesg.dk
maran@mmmi.sdu.dk
ploug@imada.sdu.dk

Abstract: This article presents the tech-didactic model DigiDidactics as a didactic roadmap for educators seeking to integrate emerging technologies such as artificial intelligence, virtual reality, and robotics into teaching practice. While these technologies offer new pedagogical possibilities, their rapid development creates uncertainty in how to design, implement, and evaluate meaningful learning activities. DigiDidactics addresses this challenge by providing a stable, non-prescriptive framework that supports educators in aligning technological use with pedagogical purpose. The model builds on earlier work with ROBOdidactics and has been further developed through collaboration between researchers and educators. It comprises five interconnected dimensions: (1) Teaching Design, focusing on learning objectives, activities, and evaluation; (2) Digital Production, emphasizing students' active creation and reflection; (3) Digital Literacy, including critical thinking, ethics, and responsible engagement with technology; (4) Environment, highlighting collaboration with external stakeholders and career learning; and (5) Sustainability, integrating long-term societal, environmental, and educational perspectives. The article demonstrates how DigiDidactics functions both as a design and evaluation framework. It provides concrete examples of its application in practice. The model is currently being used in the Interreg project MYREcross (myre.tech), in which 110 teaching modules for learners in secondary education, as well as in higher education are being designed and implemented with the model as the guiding didactic framework. These cases illustrate how the model supports educators in navigating the complexities of emerging technologies while maintaining pedagogically sound and reflective teaching practices. By combining DigiDidactics with action research, educators are positioned as co-designers and co-researchers who iteratively experiment with and refine their teaching practices. The model thus supports pedagogically grounded innovation and helps shift focus from technological novelty to meaningful learning. The article concludes that a structured, reflective didactic framework is essential for ensuring that emerging technologies contribute to educational quality, student engagement, and sustainable development.

Keywords: Emerging Technologies, Digididactics, Teaching, Digital Literacy, Didactics, Action Research.

1. Introduction

This article is of particular interest to educators who work with emerging technologies in teaching. It introduces a well-developed didactic model titled "DigiDidactics" created through collaboration between upper secondary school teachers and researchers in the field of educational technology.

Educators are regularly confronted with new technologies, such as generative artificial intelligence, virtual reality, and robotics. This raises important questions: What do students need to know about the new technology to participate meaningfully in society? What knowledge is required to understand both the strengths and the limitations of such technologies? How can emerging technologies be applied concretely in teaching, and what demands do they place on educators? Finally, what learning objectives are realistic, and how do these align with the curricular goals of a given subject? This article presents a model for planning and evaluating teaching involving emerging technologies in the classroom, with a particular focus on digital literacy and students' digital production.

We apply methods inspired by action research to equip educators with tools and approaches for experimenting with emerging technologies in teaching (Majgaard, 2011). Our approach has a particular focus on empowering educators to engage confidently with emerging technologies in educational practice. Action research is understood as an iterative methodology that closely connects theory and practice through cycles of experimentation and reflection. The study draws on Kurt Lewin's (1946) action research approach, which integrates systematic inquiry with practical change through iterative cycles of planning, action, observation, and reflection. Action research emphasizes close collaboration with practitioners and the generation of practice-relevant knowledge situated in real-world contexts. This approach supports concurrent learning and development, while acknowledging that findings are primarily context-specific and oriented toward local improvement rather than broad generalization.

Moreover, we describe how the model can be applied for evaluation of teaching modules.

Organisation of the paper: The paper begins with a brief introduction to the DigiDidactics model. This is followed by reflections on how we work with educators to prepare them to engage with emerging technologies through structured interventions inspired by action research. Then, we describe and analyse how DigiDidactics has been applied for course evaluation and to support cross-border learning activities. Finally, we provide concrete examples of teaching modules developed using DigiDidactics.

2. DigiDidactics: a Short Introduction

This section presents DigiDidactics as a framework for the design and evaluation of teaching involving emerging technologies. The model is intended as an adaptable framework, allowing educators to apply the elements that are most relevant to their specific context. DigiDidactics can also be used to systematically evaluate and refine existing courses. Our learning approach is inspired by experiential learning, meaning that educators should scaffold their teaching to support students' active experimentation and reflection (Kolb, 1984; Majgaard, 2020).

The framework builds on and extends ROBOdidactics (Majgaard & Lamscheck-Nielsen, 2019). In its revised version, sustainability has been introduced as an additional dimension.

DigiDidactics consists of five interrelated dimensions:

See the DigiDidactics model below in an overview version to the left and a detailed version to the right (Figure 1).

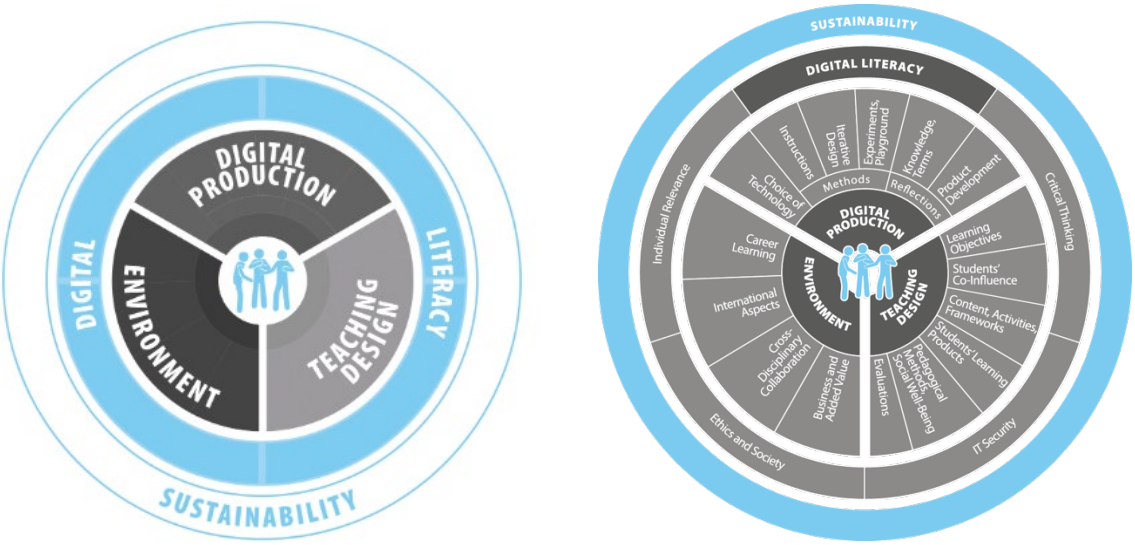


Figure 1: DigiDidactics (2026) in overview and detailed version. More details can be found in myre.tech.

The model and its predecessor, ROBOdidactics (Majgaard et al., 2019), was used to describe teaching modules in the MYREcross and several other MYRE projects (myre.tech). In addition, the model was used to evaluate and further develop these programmes.

3. DigiDidactics and Action Research

Action research is characterized by cyclical processes typically carried out by practitioners in their own contexts. DigiDidactics similarly positions teaching as a dynamic and reflective practice, where educational designs involving emerging technologies are continuously tested, evaluated, and revised in authentic classroom settings (Majgaard, 2011a).

The key concept of action research integrates systematic inquiry with practical change through iterative cycles of planning, action, observation, and reflection. Action research emphasizes close collaboration with practitioners and the generation of practice-relevant knowledge situated in real-world contexts. This approach is influential in education, organizational development, and social sciences, particularly where collaborative, practice-based inquiry is central (Majgaard, 2011b).

Kurt Lewin's concept of action research is characterized by a close integration of theory, practice, and social change (1946). One of its major strengths is its strong orientation toward real-world problem solving. Rather than producing abstract knowledge detached from practice, action research is conducted within concrete contexts and aims to generate knowledge that is immediately relevant and applicable (Majgaard, 2011a). This makes the findings highly meaningful for practitioners and increases the likelihood that research outcomes lead to sustained improvements in practice.

A second key advantage is the emphasis on participation and shared ownership. In Lewin's model, practitioners are not merely research subjects but active contributors to the research process. By involving those affected by change as co-researchers, action research supports democratic knowledge production, enhances motivation, and fosters commitment to the change process. This participatory approach often results in greater acceptance and durability of implemented solutions.

Third, action research enables learning and change to occur simultaneously through iterative cycles of action, reflection, and revision. This dynamic structure allows researchers and participants to adapt interventions continuously based on emerging insights. Knowledge development and organizational or educational change therefore unfold in parallel, reinforcing each other throughout the process.

4. DigiDidactics as an Evaluation Framework

Both DigiDidactics and its predecessor, ROBOdidactics, have proven their value as suitable frameworks in various evaluations. So far, to our knowledge, the frameworks have been applied in different settings and for different purposes: quality assessment of teaching modules involving digital production, facilitated peer feedback among teachers, conceptualization of focus group interviews, and, more recently, identification of teachers' needs for further education.

Action research and DigiDidactics as framework for evaluation are closely interrelated. Action research provides a cyclical, practice-oriented methodology for experimenting, reflecting, and improving teaching in real educational contexts, while DigiDidactics offers a shared analytical framework for making these practices explicit and evaluable. Used together, action research generates grounded insights from practice, and DigiDidactics enables systematic interpretation, comparison, and reflection across cases. In this way, DigiDidactics strengthens action research by supporting structured evaluation and transferable knowledge development.

4.1 Analysis of Teaching Modules

Using DigiDidactics, analysts can identify the distinctive characteristics of individual teaching modules while also comparing patterns across subjects, vocational domains, and educational institutions.

The purpose was not to prescribe a normative design for teaching modules or to initiate benchmarking. Rather, the purpose was to point concretely to patterns and perspectives for a broader audience. In specific educational settings, these evaluations might promote disruption, curiosity, and a desire for further development, innovation, and continuous improvement (Parlett & Hamilton, 1972).

In practice, researchers assessed a total of 143 teaching modules, of which 63 originated from the four MYRE projects. These assessments were based on teachers' systematic descriptions of their modules using DigiDidactics or its predecessor, ROBOdidactic (myre.tech). Quality was related to innovation within the five main dimensions of the model (fig. 1), the impact on students' learning, and the scalability and transferability of the modules.

It was possible to cluster the modules and identify tendencies. Thus, it became clearly visible when students had primarily worked with experimental approaches in their digital production, when teachers had consciously focused on promoting students' digital literacy, or when the modules had an outward-looking character by involving stakeholders from the educational environment.

At the same time, particularly innovative elements, such as new dialogical processes among students through the application of new technologies, were highlighted as inspiration for others.

Thus, it became possible to identify common patterns across teaching modules that were otherwise highly specific and difficult to compare. For this purpose, DigiDidactics may also be understood as a boundary object (Star & Griesemer, 1989). A boundary object is a shared conceptual framework that enables people from

different contexts or professional backgrounds to communicate and collaborate without requiring complete agreement or identical understandings.

4.2 Peer Feedback

The tech-didactic model has been applied in many different settings with many hundreds of teachers across subjects, sectors, and, more recently, across the Danish German country border (Lamscheck-Nielsen et al., 2025, supplemented with data 2025-2026, myre.tech).

Simple principles have been applied during exchange workshops:

1. Present your module according to DigiDidactics (7 min.). The modules presented may still be at the idea stage, currently being implemented, or already completed.
2. Clarification: Q&A (2 min.)
3. Feedback using sticky notes placed on the DigiDidactics model as a poster (3 min.). Alternatively, virtual environments have been used for the same purpose. The feedback can relate to categories such as "commendable, questionable or for inspiration", or e.g. simplified versions of The Six Thinking Hats (de Bono, 1985).
4. Summary and conclusions by the facilitator, including oral comments from the audience (2 min.)

The simple method and the model encourage teachers to be explicit and concise about their teaching and professional knowledge (Schön, 2001). The model seems to have a universal appeal when teachers engage with it to discuss their teaching and experiences.

With very limited introduction to the model and minimal process facilitation, teachers and lecturers are able to understand each other's teaching modules and provide valuable feedback. In line with responsive evaluation (Stake, 2004), DigiDidactics may be understood as a dialogical framework that supports communication among teachers, experts, and researchers regarding teaching practices and search for further development.

It can be concluded that DigiDidactics has not only provided a shared terminology, but also functions as an ambitious conceptual framework within which teaching practices can be positioned and teachers can experience themselves as interconnected. This perspective resonates with principles from illuminative evaluation (Parlett & Hamilton, 1972, 1977), emphasizing contextual understanding and shared reflection on educational practices.

4.3 Focus Group Interviews

DigiDidactics can also be used to design focus group interviews, with questions that refer to the dimensions of the model and, where relevant, some of its specific elements within the given context.

This approach has been implemented in interviews with both teachers and students. These successful attempts demonstrated that the model can be used to identify experiences and needs in learning processes from different perspectives.

4.4 Evaluation of Teachers´ Individual Needs

Recently, it was reported that the German Schleswig-Holsteinisches Institut für Berufliche Bildung (SHIBB), applies the model to identify teachers' needs for further education regarding teaching with digital production. Consultants from the institute facilitate the process using the DigiDidactics framework as a methodological point of departure.

DigiDidactics is used individually and systematically as a framework for reflection on personal needs in relation to the elements of the model (Figure 1, in details). Conclusions are drawn based on individual insights in comparison with the institute's educational offerings. At the same time, the institute is able to identify the needs of its target groups through a process resembling co-creation. This approach aligns with principles of responsive evaluation (Stake, 2004), emphasizing dialogue, stakeholder perspectives, and contextual understanding.

4.5 Summary

DigiDidactics may be understood as a boundary object (Star & Griesemer, 1989), enabling teachers, researchers, and experts to communicate and reflect across otherwise highly specific and heterogeneous teaching practices.

The evaluation processes outlined and their applicability in educational contexts appear aligned with principles of responsive and participatory evaluation traditions (Stake, 2004). Such approaches support professional dialogue and collective meaning-making rather than standardized measurement. In line with illuminative evaluation (Parlett & Hamilton, 1972, 1977), the purpose is to illuminate patterns, strengthen professional relationships, and support the further development of teaching practices within authentic educational settings.

5. Examples of Teaching Modules Which Combine Action Research and DigiDidactics

In the following examples of teaching modules, the DigiDidactics model was applied as a framework for developing, executing and reflecting on teaching. The modules were described as a combination of the five dimensions. The teaching modules can be found in myre.tech in full length.

5.1 Example "Business Ideas Through AI" a Bridging Module Between Lower Secondary and Upper Secondary Business Education

This teaching module was implemented as a bridging module between Haahrs School (10th grade) and Svendborg business highschool (HHX) (12-13th grade), involving 48 students supported by HHX student tutors. The aim was to introduce students to innovation and entrepreneurship by developing sustainable business ideas using AI tools. The course combined authentic learning experiences, project-based work, and collaboration with peers and external stakeholders. It aimed to strengthen students' understanding of business, AI, and career perspectives while preparing them for upper secondary education.

Teaching Design: The module was structured as project-based learning with a mix of short lectures, workshops, group work, and continuous guidance. Students worked in groups to develop their own business plans, supported by teachers and HHX tutors. The design emphasized differentiation, inclusion, and student agency, allowing learners to choose their ideas, tools, and presentation formats. Assessment was based on a final report, prototype, exhibition stand, and oral presentation evaluated by teachers and external business judges. Continuous feedback played a central role throughout the process.

Digital Production: Students actively used a range of generative AI tools (e.g., ChatGPT, Copilot, Gamma.ai) to support ideation, text production, visual content, and presentations. The process followed an iterative design approach, where students continuously developed, tested, and refined their products. AI contributed to improving the overall quality of output, particularly written reports. However, differences emerged between students who critically processed AI outputs and those who used them superficially without deeper understanding.

Digital Literacy: A key focus was developing students' critical awareness of AI. Students learned about both the opportunities and limitations of AI, including risks related to bias, errors, and overreliance. They were introduced to basic GDPR and data security principles and practiced responsible and anonymous use of AI tools. The module highlighted the importance of using AI as a support tool rather than a substitute for independent thinking.

Environment: The module was embedded in a strong collaboration between education and the local business community. External stakeholders, including companies and judges, contributed to the authenticity of the learning experience. Peer tutoring created a meaningful link between educational levels and supported career learning. A shared virtual infrastructure (Padlet) facilitated collaboration and resource sharing. Future international collaboration with a German start-up initiative was planned to extend the learning environment further.

Sustainability was a core theme throughout the module, aligned with UN Sustainable Development Goal 12 (responsible consumption and production). Students were required to integrate environmental, economic, and social sustainability into their business ideas, reports, and presentations. This ensured that technological innovation was connected to broader societal and ethical considerations.

This teaching module has been implemented over several years in different versions and has been continuously refined and optimized.

5.2 Example: "Drone Inspection of Large Trees"

The teaching module introduces drone-based inspection of large, old trees as a safer and more efficient alternative to traditional climbing inspections. Using FPV drones and VR goggles, students inspect tree crowns to assess health and structural safety.

Teaching Design: Students worked in groups with defined roles (pilot, spotter, analyst). The module included: a) Introduction to FPV drone flight and safety. b) Hands-on practice (simulation and real-world flying). c) Case-based analysis of tree health using images and video. d) Professional language with correct expressions and terms, including communication in English with an international partner. The course was piloted in September 2025 with 14 students at Jordbrugets UddannelsesCenter Aarhus in Denmark and is now integrated into the curricula.

Digital Production: Students learned to operate professional inspection technologies (FPV drones and VR goggles) and to document findings accurately. While not conducting digital development themselves, they gained insight into technological choices, data quality requirements, and professional terminology relevant to the field.

Digital Literacy: The course emphasized ethical, legal, and safety aspects of drone use, including privacy, data handling, and responsible application in public spaces. Students reflected critically on opportunities and limitations of digital technologies in forestry practice.

Environment: The module contributes to the development of forestry and arborist education and supports coherence across initial vocational education and continuing education. Collaboration with an international institute strengthened professional relevance and student understanding of digitalized work practices in a global context.

The second iteration of the teaching module placed greater emphasis on training in drone flight in safe environments before students were allowed to operate among the trees. This was due to the experiences from the first iteration, which had been more turbulent regarding flying the drones.

6. Discussion

6.1 DigiDidactics as a Stable Didactic Scaffold for Emerging Technologies

A central contribution of this article is the articulation of DigiDidactics as a stable didactic roadmap in a context characterized by rapid technological change. Emerging technologies such as AI, VR, AR, and robotics evolve at a pace that often outstrips the capacity of educational institutions to adapt curricula, teacher competencies, and assessment practices. This volatility risks shifting pedagogical focus toward technological novelty rather than sustained educational value. DigiDidactics addresses this challenge by providing a framework grounded in didactic principles rather than specific technologies.

By foregrounding dimensions such as teaching design, digital production, digital literacy, external collaboration, and sustainability, the model enables educators to situate new technologies within enduring pedagogical concerns. The framework thus remains applicable even as particular tools become obsolete. In this sense, DigiDidactics supports continuity and coherence across changing technological landscapes, helping educators maintain focus on learning objectives, student engagement, and societal relevance.

Importantly, the model does not prescribe a specific ideal way of teaching with technology. Instead, it functions as a reflective guide that supports informed decision making, where teachers can relate to their own given context. This non-normative character is particularly important when dealing with emerging technologies, where best practices are often still under negotiation. DigiDidactics therefore contributes to a form of pedagogical and societal resilience (Lamscheck-Nielsen et al., 2025), enabling education systems to respond constructively to technological change without becoming technology-driven.

6.2 DigiDidactics as a Framework for Evaluation Processes

DigiDidactics proves its value as a framework supporting evaluation processes. The model offers a shared analytical framework for making different teaching practices in different contexts explicit and evaluable. DigiDidactics enables systematic interpretation, comparison, and reflection across cases. Such approaches support professional dialogue and collective meaning-making rather than standardized measurement. The purpose is to illuminate patterns, strengthen professional relationships across subjects, organizations, and the country borders. These processes may promote the further development of teaching practices within authentic educational settings.

6.3 Action Research as a Catalyst for Pedagogically Grounded Innovation

The close integration of DigiDidactics with action research represents a deliberate methodological choice that strengthens both innovation and pedagogical grounding. Action research offers a structured yet flexible approach to educational development, allowing educators to experiment with emerging technologies while continuously reflecting on their pedagogical implications. Rather than separating experimentation from evaluation, action research integrates these processes through recurring cycles of planning, action, observation, and reflection (Majgaard, 2011a).

Within this context, DigiDidactics functions as a shared analytical lens that supports systematic reflection across iterations. Educators are able to articulate their teaching designs, examine outcomes, and make didactically informed adjustments. This helps reduce the risk that emerging technologies are introduced in ad hoc or superficial ways. Instead, innovation is embedded within reflective practice and aligned with educational aims. Innovation is therefore guided by pedagogical reflection.

At the same time, the use of action research implies certain limitations. Findings are inherently context-specific and primarily aimed at local improvement rather than universal generalization. However, when combined with a shared roadmap such as DigiDidactics, local insights become more communicable and transferable. Patterns, tensions, promising approaches, and even evaluation results can be identified across cases without losing sensitivity to context. In this way, the combination of action research and DigiDidactics balances situated knowledge production with broader knowledge sharing.

6.4 From Technological Novelty to Pedagogical Purpose

A recurrent challenge in technology enhanced education is the tendency for technological possibilities to drive and overrule pedagogical decisions. New tools often enter educational settings with high expectations but limited reflection on how they support learning objectives, student participation, or assessment practices. DigiDidactics addresses this challenge by embedding technology use within a broader didactic structure that foregrounds pedagogical purpose.

Across the teaching module examples presented in this article, emerging technologies are treated as means rather than ends. In the drone inspection module, for instance, drone and VR technologies serve clearly defined learning objectives related to professional practice, safety, and ethical considerations. Similarly, the module on Business innovation and AI demonstrates a strong shift from technological novelty to pedagogical purpose by using AI as a tool to support entrepreneurial learning rather than as a goal in itself. AI is integrated into authentic tasks such as developing sustainable business ideas and presenting them to real world stakeholders.

This ensures that technology enhances learning relevance without dominating it. The design emphasizes student agency, collaboration, and real-world application. At the same time, the module highlights the importance of critical reflection on AI outputs. Some students relied too heavily on AI, leading to reduced depth of understanding. However, AI also increased motivation and improved the quality of student products. Part of the evaluation of this module shows that some students had difficulties critically engaging with AI-generated output. As a result, in the next iteration of the module, the educator will place greater emphasis on fostering students' critical reflection.

This shift from technology-driven to pedagogy-driven design is further reinforced by the model's emphasis on reflection and optimizing teaching modules. By explicitly addressing digital literacy and sustainability, DigiDidactics encourages educators and students to critically examine not only how technologies work, but also why and when they should be used. This supports more responsible, reflective, and educationally meaningful integration of emerging technologies.

6.5 Teachers As Co-Researchers and Designers of Learning Processes

Another important implication of the DigiDidactics action research approach is the repositioning of teachers as active co-researchers and designers of learning. Rather than being passive recipients of predefined technological solutions, teachers are engaged as knowledgeable practitioners who contribute to didactic development through reflection and experimentation.

The shared language and focus fields provided by DigiDidactics play a crucial role in this process. It enables teachers across subjects and institutions to articulate their pedagogical choices, exchange feedback, and reflect collectively on teaching practices involving emerging technologies. Peer feedback sessions, focus group

interviews, and assessment processes all benefit from this common framework, which reduces ambiguity and supports mutual understanding.

This participatory approach has implications for professional development and organizational learning. When teachers experience themselves as contributors to didactic innovation, their sense of agency and professional confidence increases. This is particularly important when engaging with emerging technologies, which can otherwise be perceived as disruptive or intimidating. By embedding technological experimentation within familiar didactic concerns and collaborative inquiry, DigiDidactics supports sustainable professional learning and institutional capacity building.

7. Summary and Conclusion

This article has presented DigiDidactics as a didactic roadmap that supports educators in integrating emerging technologies into teaching practice in a reflective, structured, and sustainable way. In response to the rapid development of technologies such as AI, VR, and robotics, the article argued for the need for a stable didactic roadmap that enables educators to plan, implement, and evaluate technology-enhanced teaching without being driven solely by technological novelty. DigiDidactics addresses this need by combining classical didactic considerations with perspectives on digital production, digital literacy, engagement with the external world, and sustainability.

Through its close connection to action research, DigiDidactics positions teaching with emerging technologies as an iterative, experimental, and practice-based process. Educators should be empowered to test new approaches, reflect on outcomes, and revise teaching designs in collaboration with peers and researchers. This approach supports both professional learning and local educational development, while acknowledging that innovations must be context-sensitive rather than universally prescriptive.

The article further demonstrates how DigiDidactics functions as a coherent evaluation framework across multiple purposes, including analysis of teaching modules, facilitated peer feedback, focus group interviews, and identification of teachers' professional development needs. These applications show that the model not only provides shared terminology, but also enables comparison, reflection, and knowledge transfer across otherwise diverse and non-comparable teaching contexts.

Concrete examples from upper secondary education illustrate how DigiDidactics and action research work together in practice. Across the examples, iterative refinement based on reflection and experience improves instructional quality, strengthens students' digital literacy, and enhances the relevance of learning by connecting education to vocational and societal contexts. These cases demonstrate that emerging technologies can contribute meaningfully to learning when embedded in a clear didactic framework.

In conclusion, DigiDidactics offers a robust and flexible roadmap for bridging the gap between rapidly evolving technologies and sustainable teaching practice. By foregrounding student-centered pedagogy, reflection, and collaboration, DigiDidactics helps ensure that emerging technologies become vehicles for educational quality and innovation rather than ends in themselves. Future work may further explore how DigiDidactics can support long-term institutional development and cross-sector collaboration in digital education.

Ethics Declaration

Ethical clearance was not required for the research.

AI Declaration

AI tools were utilized in the example teaching module on Innovation and AI.

References

de Bono, E. (1985) Six Thinking Hats: An Essential Approach to Business Management. Little, Brown, & Company

Hiim, H. and Hippe, E. (2007) Læring gennem oplevelse, forståelse og handling (red. Learning through experience, understanding, and action.), Gyldendal

Kolb, D. (1984) Experiential Learning: Experience as the Source of Learning and Development. Prentice-Hall.

Lamscheck-Nielsen, R. & Majgaard, G. (2025) "Teaching Modules utilizing Emerging Technologies across Educational Levels and Sectors", INTED2025 Proceedings. 19th annual International Technology, Education and Development

Conference, Valencia, Spain. https://moeve.dk/wp-content/uploads/2025/03/INTED2025_905-paper-educationchain.pdf
Lankshear, C. J., & Knobel, M. (2008) “Introduction: digital literacies: concepts, policies and practices”.
Digital literacies: concepts, policies and practices. Peter Lang Publishing, New York, USA, pp. 1-16
Law, B. (2010). Building on what we know, Career Learning
Lewin, K. (1946) "Action Research and Minority Problems". Journal of Social Issues, 2(4), 34-46.
Majgaard, G., Misfeldt, M., & Nielsen, J. (2011a) “How Design-based Research, Action Research and Interaction Design Contributes to the Development of Designs for Learning”. Designs for Learning, 4(2), 8-21. http://www.designsforlearning.nu/11/no2/DFL_02_11_majgaard_etal.pdf
Majgaard, G. (2011b) Learning Processes and Robotic Systems: Design of Learning Materials and Learning Processes with Robots as Media and Children as Co-designers. [PhD dissertation, SDU]. University of Southern Denmark, Faculty of Engineering.
Majgaard, G., & Lamscheck-Nielsen, R. (2019) “Digital Literacy and Course Design”. Proceedings of the 18th European Conference on e-Learning ECEL 2019 (pp. 369-377).
Majgaard, G., & Weitze, C. (2020) “Virtual experiential learning, learning design and interaction in extended reality simulations”. Proceedings of the 14th European Conference on Game Based Learning, ECGBL 2020 (s. 372-379). Academic Conferences and Publishing International. https://doi.org/10.34190/GBL.20.010
Majgaard, G. (2018) ”Digital dannelse på højskolen” (Digital Literacy in the Folk Highschool). Læring & Medier (LOM) – nr. 19 – 2018. ISSN: 1903-248X
Martin, A. (2008) “Digital Literacy and the Digital Society”. Vol. 30 of Digital Literacies: Concepts, Policies and Practices, 151–176. New York: Peter Lang.
Parlett, M., & Hamilton, D. (1972) Evaluation as illumination: A new approach to the study of innovatory programs. Edinburgh: Centre for Research in the Educational Sciences, University of Edinburgh.
Parlett, M. & Hamilton, D. (1977). “Evaluation as Illumination: A New Approach to the Study of Innovatory Programs.” In G. Glass (Ed.), Evaluation Studies Review Annual (Vol. 1, pp. 140–157). Beverly Hills, CA: Sage.
Resnick, M. et al. (2009) “Growing up Programming: Democratizing the Creation of Dynamic”. CHI 2009, Boston, MA, USA ACM 978-1-60558-247-4/08/04
Schön, D. (2001) The Reflective Practitioner: How professionals think in action. London: Temple Smith.
Stake, R. E. (2004). “Standards-Based and Responsive Evaluation.” Thousand Oaks, CA: Sage.
Star, S. L., & Griesemer, J. R. (1989). “Institutional Ecology, ‘Translations’ and Boundary Objects.” Social Studies of Science, 19(3), 387–420.

The Level of Digital Skills of Mining Employees: Their Learning Needs and Interest in Personal Development

Josef Malach, Petr Pyszko, Kateřina Malachová and Dana Vicherková
University of Ostrava, Ostrava, Czechia

josef.malach@osu.cz
petr.pyszko@osu.cz
katerina.malachova@osu.czdana
vicherkova@osu.cz

Abstract: Mining professions in the Czech Republic are declining. As a result, employees must transition to alternative occupations. This study examines the current level of digital skills among these employees. Such competences are broadly relevant for future employment and personal contexts. The research also examines employees' motivation and interest in enhancing their digital skills while facing gradual layoffs. The aim is to identify both individual and collective developmental needs. By analysing the relationship between self-assessed digital competence and motivation to improve these skills, it is possible to identify employees with both low competence and low motivation. These individuals may require targeted career guidance. In 2025, a stratified random sample of 529 employees from a mining organisation was surveyed. The survey yielded the following key findings: Respondents' digital competence averaged 2.41 on a five-point scale, below the midpoint. The perceived need for digital training was higher, with a mean of 3.44. Interest in acquiring new skills averaged 3.80, while interest in further development averaged 3.83. The most significant barrier was health or age (mean = 3.19), followed by lack of support (mean = 3.02) and lack of time (mean = 2.99). Analyses indicated a positive saturating relationship between digital competence and educational need. There were also positive linear relationships between educational need and both interest in learning new skills and interest in skill development. The overall mean barrier index did not significantly predict educational need.

Keywords: Digital Skills, Adult Learning Needs, Barriers to Participation in Education, Interest in Learning, Mining Industry

1. Introduction

"*Digital skills are increasingly considered a basic requirement for a wide range of professions, not only in technology fields. From basic skills, such as the ability to work effectively with office applications and online tools, to advanced skills in programming, data analysis and cybersecurity. Digital skills open the door to new opportunities and career growth. The labour market is changing, and it is necessary to respond to the ongoing changes. New types of work and new skill requirements are rapidly increasing." (Government Office of the Czech Republic, 2024). Recent* research indicates that the changes occurring in this decade will be crucial for Europe's future competitiveness (Deleersnyder et al., 2022). By 2030, the 50–65 age group is projected to increase by 17% and constitute the primary workforce. Over 90% of all jobs will require at least basic digital skills, compared to approximately 54% today. The concept of digital skills, or competences, continues to evolve as technological advancements and rapid digital development continue. Professional organisations and policymakers are actively working to define these skills by establishing relevant standards and content. This effort has resulted in the creation of several general and specific frameworks for digital competences, the most prominent being the European Digital Competence Framework (DigComp), "*which describes what is needed to be digitally competent in today's society and supports the development of digital competence among individuals of all ages. It covers a wide range of skill levels, from basic to highly advanced, and offers a stable reference point in a rapidly evolving digital technological landscape. DigComp 3.0 is the fifth version of the framework.*" (European Commission, 2026). The fifth version, compared to the 2022 update, introduces revised wording for competence areas and competences, updated proficiency levels, new learning outcomes, and the integration of AI competence.

The need to update general frameworks for digital competences is highlighted by findings that "over nine in ten (92%) workers in the EU used digital technologies in their work in 2024–2025 and 30% of EU workers used AI systems at work in the same year. In 2024, 42% of EU workers reported an AI skills gap, yet only 15% had participated in AI skills training" (Cosgrove and Cachia, 2025, p. 5).

The DigComp framework promotes a shared and common understanding of digital competence – knowledge, skills, and attitudes for the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society. It is used by a large majority of EU Member States and beyond, as well as at European and international levels, to inform policymaking; develop assessments; enhance transparency or comparability of education and training courses; recognise or validate learning (such as through digital skills certification); and define profiles of digital competence in specific jobs or roles (Cosgrove and Cachia,

2025). The framework consists of five areas: 1) information search, evaluation, and management; 2) communication and collaboration; 3) content creation; 4) safety, wellbeing and responsible use; and 5) problem identification and solving. Competences in the above areas can be developed up to four levels according to the framework: basic, intermediate, advanced, and highly advanced. At the basic level, individuals remember and implement simple tasks with guidance as needed. At the intermediate level, individuals identify and implement well-defined tasks and solve well-defined problems autonomously. We assume that respondents to our research will be equipped with digital competences mostly at the first level, which, among other things, assumes skills in using communication and collaboration software, word processing, spreadsheets and presentation software. The items of the employee questionnaire used focused on these skills. Within general digital frameworks, concepts such as digital rights, digital inclusion, and digital resilience have been outlined in response to rapid change. The digital rights of EU citizens appear in the European Declaration on Digital Rights and Principles for the Digital Decade (European Commission, 2023). This declaration presents 26 key points on the significance and functions of digital technologies, along with their opportunities and risks. Point 2 states, "*the digital transformation should contribute to a fair and inclusive society and economy in the EU*." Article 4 affirms everyone's right to education, training, and lifelong learning, including the right to acquire all digital skills. Articles 5 and 6, under Fair and Just Working Conditions, emphasise that "*trade unions and employers' organisations play an important role in the digital transformation, particularly in relation to the definition of fair and just working conditions, including with regard to the use of digital tools at work." These articles highlight the new idea of (digital) "Solidarity and inclusion*" (European Commission, 2023).

A subject of interest to experts and politicians is security and the management of personal and corporate risks associated with the use of the Internet and other digital tools and services. "*Resilience can be defined as 'a process to harness resources to sustain wellbeing', and digital resilience as the application of this concept to technology, the Internet, and the digital age. Digital resilience helps individuals recognise and manage the risks they encounter when socialising, exploring, or working online. It is achieved primarily through experience, rather than learning, and it is fostered by opportunities to confide in trusted others and later reflect upon online challenge*" (UK Council for Internet Safety, 2025).

Educational needs are defined as deficits or potential for the development of knowledge, skills, abilities, habits, or behaviours (Šeďová and Novotný, 2006). These needs are highly dynamic and variable, influenced by a range of interacting external and internal factors.

Educational need, understood as a deficit or potential for adult development, may pertain to individuals, groups, or larger social units (Průcha, 2014). For individuals, *"Educational needs of a subject are essentially states of experiencing and feeling the lack of something or the desirability of something that the subject lacks, which appears to them as a preference or ideal that can be achieved through some form of education"* (Palán, 2002, p. 37). This definition presumes individual awareness of the educational need. However, some individuals may be unaware that they lack the competences necessary for work, civic roles, or hobbies [10]. This perspective is further reflected in Prusáková's (2010, p. 17) definition: *"Educational needs arise as a hypothetical state (conscious or unconscious) when an individual lacks knowledge or skills that are important for his or her continued existence, the preservation of psychological, physical or social functions."* The typology of educational needs can be based on more than ten criteria (Malach, 2024). In this research, organisational (corporate) educational needs related to the number of individuals were considered, as well as professional needs focusing on educational areas. Additional considerations included the localisation of stimuli triggered by external needs, subjective needs from the subject–object criterion, stages of analysis of educational needs, types of educational need (as key competence), possibility of satisfaction, realisation status, type of need (deficit or potential as proactive), and urgency (non-urgent, satisfiable over a longer period).

Adult participation in education is often limited, restricted, or entirely prevented by various adverse circumstances, which are traditionally categorised as dispositional, institutional, and situational barriers (Kalenda et al., 2024). Dispositional barriers stem from adults' learning, educational, and life experiences and needs, manifesting as concerns about participation, doubts regarding the value of continued education after initial schooling, or low self-efficacy. Institutional barriers are obstacles encountered by education providers, including perceptions of the availability and quality of educational opportunities, as well as awareness of the education market. Situational barriers are directly related to individuals' life circumstances and may prevent participation even when opportunities exist or are known. These typically include lack of time, family obligations, insufficient employer support (such as not permitting education during working hours), and lack of financial resources.

The issue of determining the level of educational needs was the subject of the international project Needs of Adult Education Stakeholders, for which the research team developed and verified the research tool *Adult Education Needs Inventory* (AENI) (Ślósarz et al., 2022). In a group of 210 randomly selected participants, they determined the level of 9 skill groups, including digital ones. The average value of digital skills compared to the other five skills was the highest on a five-point scale, at 3.94, and was even higher than the average value (3.65) of all 39 AENI items. In relation to the partial objectives of the REFRESH project, which monitor the mutual connections between the education of adult employees of the closing coal industry and their children, the current results of the *International Computer and Information Literacy Survey* (ICILS) for the Czech Republic are also relevant; the following findings can be considered relevant to the project (Halbová et al., 2024): a) the average result of the Czech Republic in computer and information literacy was 525 points and was the highest average result among EU countries, b) computer and information literacy was related to the socio-economic background of the pupils, with disadvantaged pupils achieving an average result 64 points lower (i.e. three quarters of the skill level) on the computer and information literacy test, c) pupils from households with two or more computers scored approximately two-fifths of a skill level higher than pupils with one or no computer in the household in the computer and information literacy test, d) more pupils learn about using the Internet outside of school than in school, and e) 65% of 8th-grade pupils have no restrictions from their parents on screen time on school days, compared to an EU average of 60%.

2. Research

2.1 Research Problem and Goals

The research problem of the study is the lack of empirical data on the level of digital skills of a specific professional group in the context of its functioning during the period of decline of the industrial sector, and on the degree of their interest in further improving these skills, which would stimulate personal or systemic retraining activities. At the same time, the most significant barriers to employees' participation in further education are unknown. The aim of the research was to determine the level of digital skills of workers in mining professions in a representative sample, to determine the level of their need for digital education and their self-assessment of their readiness to learn new skills and develop their personal abilities, and to identify their potential problems (barriers) associated with their possible participation in this education.

2.2 Research Questions, Variables, and Hypotheses

For clarity and due to the limited scope, research questions (RQs) are always listed with the relevant results, which operate within defined variables and hypotheses formulated on their basis.

2.3 Methods

Data collection was conducted through a questionnaire survey. For the purposes of this study, the following selected items from the comprehensive questionnaire administered to mining employees were utilised:

- 4 items for the variable “subjectively assessed digital skills” and 4 items for the variable “subjectively declared level of educational needs” from the *SIW-Skills Inventory Worksheet* (2024). The inventory is internally structured, with 81 items that identify educational needs across 9 competences or skills.
- *10 items from the Barriers to Participation in Education Questionnaire, based on* the results of European monitoring using a five-point scale (European Commission, 2019).
- *2 items from the Czech version of the Career Adapt-Abilities Scale-Czech Form* (CAAS-Czech) (Hlaďo et al., 2020; Hlaďo and Kvasková, 2020).

The comprehensive research tool was approved by the University's Research Ethics Committee, and respondents participated in the research after being familiarised with the study and providing informed consent to participate.

2.4 Sample

The research group was a representative stratified sample of employees of a mining organisation, with respect to the division into labourers and technical-operatives, age structure, and gender. It consisted of 529 respondents.

2.5 Data Collection

Data collection took place between December 2024 and June 2025 through a survey organisation qualified to conduct research and marketing surveys.

2.6 Data Analysis

Data were analysed in R 4.3.3 (R Core Team, 2024). Digital competence and need for digital training were computed as mean indices from four 1–5 Likert items each; interest in learning new skills and interest in further development were assessed using two single-item 1–5 measures. Barriers to participation (10 items, 1–5) were analysed both at the item level and as 0–100 composite indices: an overall barrier index (mean of all 10 items) and two subscales separating personal/resource barriers (items 01–05) from course-access barriers (items 06–10).

Descriptive statistics were reported first. Associations were tested using Pearson and Spearman correlations and OLS regression. For the competence–need link, potential non-linearity was examined by adding a quadratic term and comparing model fit using AIC. For an interpretable cross-tabulation, competence and need were dichotomised using a median split, and the 2×2 association was evaluated using the χ^2 or Fisher's exact test, reporting the odds ratio with 95% CI. Associations between need and the two motivation items were tested with correlations and OLS. For barrier items, p-values were adjusted for multiple testing using BH-FDR (10 tests); $\alpha = 0.050$.

3. Results

The research results are presented in accordance with the research questions formulated above.

3.1 Descriptive Statistics (RQ1–RQ4)

Overall, *digital competence* was below the scale midpoint, with a mean of 2.41 (SD = 1.02) and a median of 2.25 (IQR = 1.25) on the 1–5 scale (N = 529) (Fig. 1). In contrast, the *need for digital training* was higher, with a mean of 3.44 (SD = 1.17) and a median of 3.75 (IQR = 1.50) (N = 528, NA = 1) (Fig. 2).

Both learning motivation indicators were high: *interest in learning new skills* had a mean of 3.80 (SD = 0.924, median = 4, IQR = 1.00; N = 528, NA = 1) (Fig. 3) and *interest in further development* had a mean of 3.83 (SD = 0.909, median = 4, IQR = 1.00; N = 528, NA = 1) (Fig. 4). The binned distributions showed that competence scores were concentrated in the lower-to-mid range, whereas need scores were concentrated in the mid-to-upper range, consistent with the descriptive summaries. The two interest items were skewed toward higher responses (4–5).

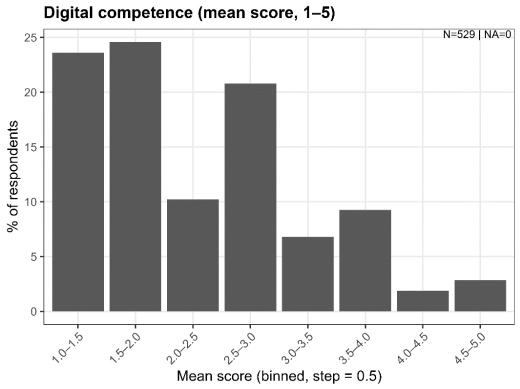


Figure 1: Digital competence (mean score, 1–5). Distribution of the digital competence index (mean of 4 items), shown as the percentage of respondents in 0.5-wide score bins. N and the number of missing values (NA) are indicated in the panel.

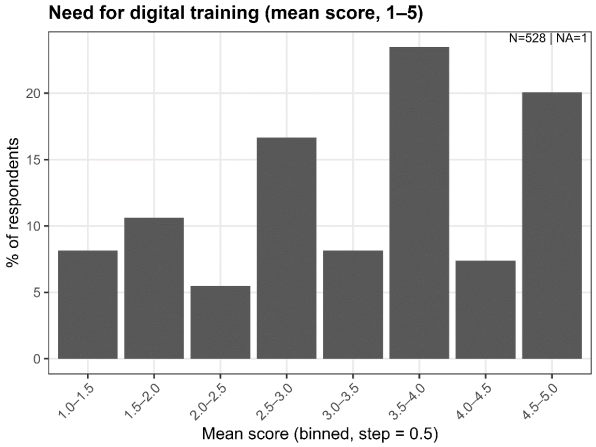


Figure 2: Need for digital training (mean score, 1–5). Distribution of the need-for-training index (mean of 4 items), shown as the percentage of respondents in 0.5-wide score bins. N and NA are indicated in the panel.

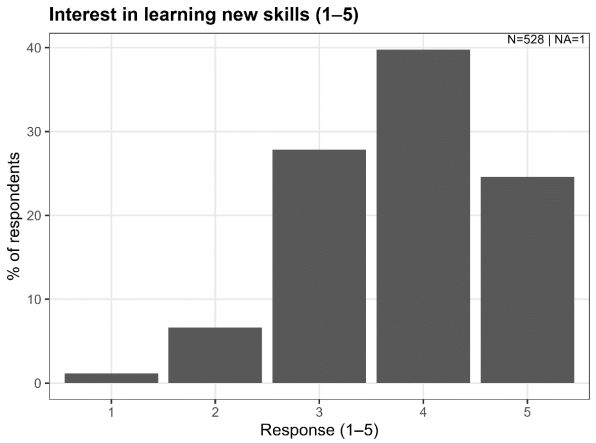


Figure 3: Interest in learning new skills (1–5). Distribution of responses to the single-item measure of interest in learning new skills, shown as the percentage of respondents selecting each response category (1–5). N and NA are indicated in the panel.

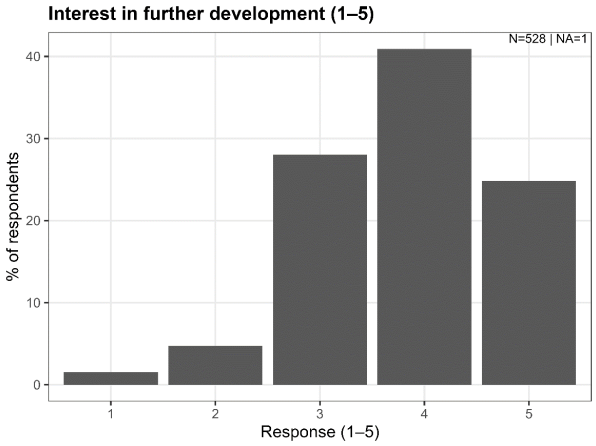


Figure 4: Interest in further development (1–5). Distribution of responses to the single-item measure of interest in further development, shown as the percentage of respondents selecting each response category (1–5). N and NA are indicated in the panel.

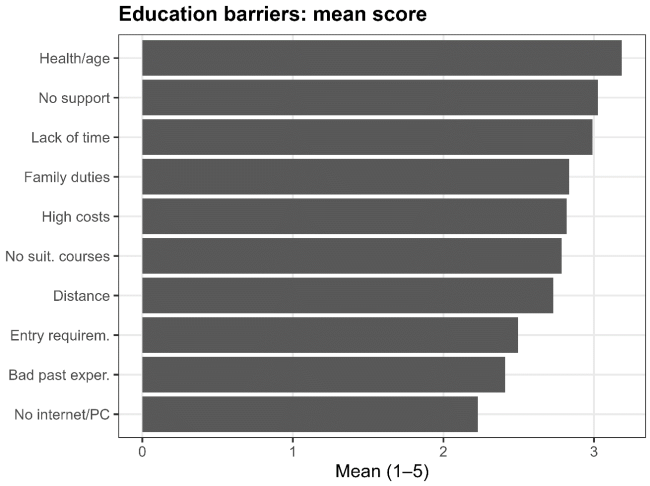


Figure 5: Ranking of barrier items by mean Likert score (1–5). Higher values indicate stronger perceived barriers.

In addition to the item-level ranking in Fig. 5, barriers were also summarised as 0–100 indices, which showed a mean overall barrier score of 43.8 ± 16.6 (median 42.5, IQR 22.5, range 0–95, n = 529), with two subscales: personal/resource barriers (items 01–05) 49.3 ± 18.3 (median 50, IQR 20, range 0–100), and course access/availability barriers (06–10) 38.2 ± 20.1 (median 35, IQR 25, range 0–100).

At the item level (1–5), the highest mean barrier was Health/age (item 04; mean = 3.19), followed by No support (05; mean = 3.02) and Lack of time (02; mean = 2.99). The lowest mean barrier was No internet/PC (08; mean = 2.23). When focusing on the proportion of "high barrier" responses (4–5), Health/age was most frequently endorsed (39.9%), followed by Lack of time (33.1%) and No support (32.9%). The lowest "high barrier" endorsement rates were observed for No internet/PC, Bad past experience, and Entry requirements (each 16.3%).

3.2 Association Analyses (RQ 5 – RQ 8)

3.2.1 Main Association Between Competence and Need (RQ5)

Digital competence was positively associated with the need for digital training (Pearson r = 0.397, 95% CI [0.323, 0.467], $p < 0.001$; Spearman ρ = 0.415, $p < 0.001$; N = 528) (Fig. 6). In a quadratic regression, need for digital training showed a non-linear association with digital competence: need increased with competence, but the increase levelled off at higher competence (negative quadratic component: $t = -5.083$, $p < 0.001$). The quadratic model fitted the data better than the linear model (ΔAIC = 23.36) and explained 19.7% of the variance ($R^2 = 0.197$; $F(2, 525) = 64.57$, $p < 0.001$).

Using median splits (competence median = 2.25; need median = 3.75), the most frequent quadrant was high competence / high need (33.3%), followed by low competence / low need (30.5%), while the two discordant groups (low competence / high need and high competence / low need) represented 17.6% and 18.6% of respondents, respectively. Quadrant frequencies differed from an equal 25% distribution ($\chi^2(3) = 41.318$, $p < 0.001$). The 2×2 table (Fig. 7) indicated a strong association between competence category and need category ($\chi^2(1) = 39.137$, $p < 0.001$; Fisher's exact $p < 0.001$), with higher competence linked to higher odds of high need (odds ratio = 3.102, 95% CI [2.146, 4.507]).

These results were consistent with hypothesis H1, which predicted a positive relationship between the subjectively assessed level of digital competences and the subjectively reported level of these educational needs among mining employees.

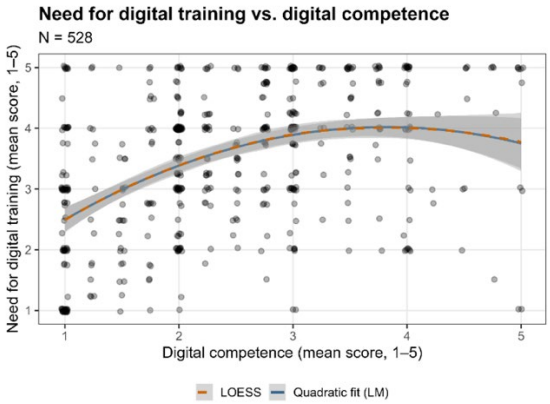


Figure 6: Need for digital training vs. digital competence. Scatterplot of individual respondents (jittered) with a quadratic regression fit (solid line) and 95% confidence band; a LOESS smoother (dashed line) is shown for visual assessment of potential non-linearity. Axes show mean indices (1–5).

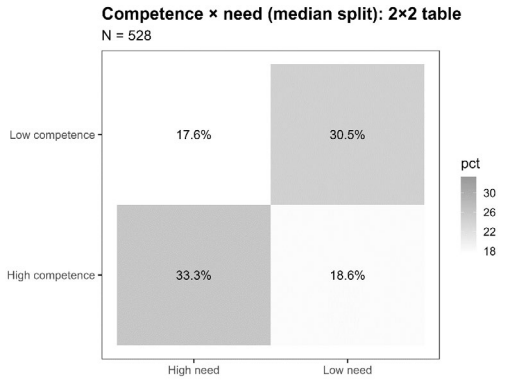


Figure 7: Competence × need (median split): 2×2 table. Heatmap of cell percentages for high/low competence by high/low need, based on median splits (N = 528).

3.2.2 Need for Training vs. Interest in Learning New Skills (RQ 6.1)

Need for digital training increased with interest in learning new skills (N = 527; Pearson r = 0.219, p < 0.001; Spearman ρ = 0.220, p < 0.001); the OLS slope was β = 0.278 (95% CI [0.172, 0.384], p < 0.001; R^2 = 0.048) (Fig. 8).

These results were consistent with hypothesis H2.1, which assumed a positive relationship between the subjectively reported level of educational needs of mining employees in digital competences and their self-assessment of interest in learning new skills.

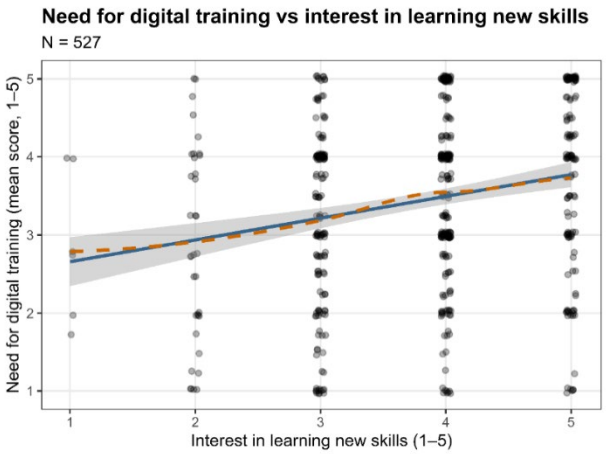


Figure 8: Need for digital training vs. interest in learning new skills. Jittered raw observations (N = 527) with an OLS regression line and 95% confidence band; LOESS (dashed) is shown as an exploratory smoother.

3.2.3 Need for Training vs. Developing One's Own Abilities (RQ 6.2)

Need for digital training also increased with interest in developing one's own abilities (N = 527; Pearson r = 0.251, p < 0.001; Spearman ρ = 0.249, p < 0.001). The OLS slope was β = 0.324 (95% CI [0.217, 0.431], p < 0.001; R^2 = 0.063), indicating a slightly stronger association than for learning new skills (Fig. 9).

These results were consistent with hypothesis H2, which predicted a positive association between need for digital training and interest in developing one's abilities.

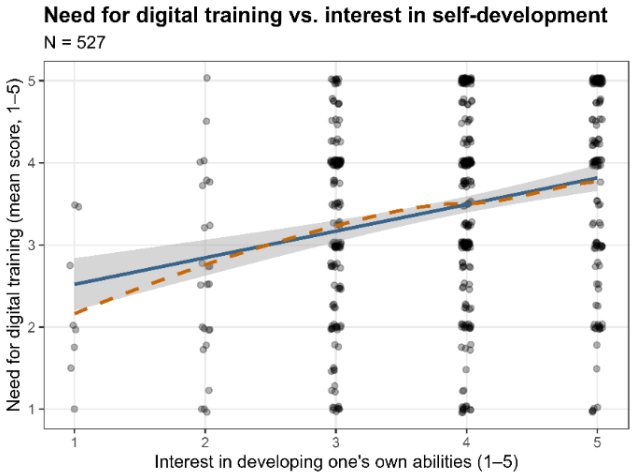


Figure 9: Need for digital training vs. interest in developing one's own abilities. Jittered raw observations (N = 527) with an OLS regression line and 95% confidence band; LOESS (dashed) is shown as an exploratory smoother.

3.2.4 Educational Needs vs. Educational Barriers (RQ 7)

Three barrier items showed the strongest (negative) associations with need after FDR correction: Bad past experiences (ρ = −0.185, p_FDR < 0.001), No internet/PC (ρ = −0.170, p_FDR < 0.001), and Entry requirements (ρ = −0.158, p_FDR < 0.001; N = 528). The remaining items were weak and not FDR-significant (Fig. 10).

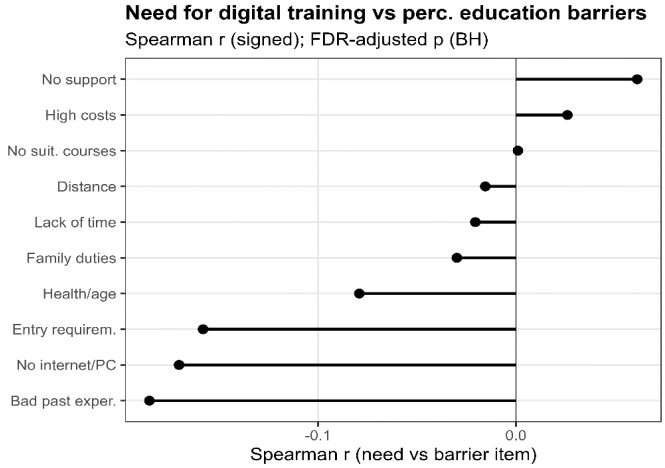


Figure 10: Need for digital training vs. perceived educational barriers (item-level). Lollipop plot shows signed Spearman correlations; p-values are FDR-adjusted (BH).

For barrier indices, the overall mean barrier index was not a significant predictor of need (b = −0.005, p = 0.080; R^2 = 0.006). The Personal/resource subscale showed no association (b ≈ 0.000, p = 0.901; R^2 ≈ 0.000). The course-access subscale showed a small negative association (b = −0.007, 95% CI [−0.012, −0.002], p = 0.005; R^2 = 0.015; N = 528).

Overall, H3 was not supported in the expected direction: the overall barrier index did not predict need, while only a few specific barriers (and the course-access subscale) showed small negative associations with need.

4. Conclusion

The study contributes empirical data to the theory of the development of digital skills of adults and to the andragogical context for their further development. The results provide a basis for educational practice in designing educational offerings that ensure the development of digital skills of adults, including the skills of parents who simultaneously influence the formation of digital competences of their children.

5. Discussion

The study's limitations include the use of only four items to measure digital skills, which does not sufficiently capture the breadth of the DigComp 3.0 framework (2025). The research plan assessed digital skills as one component within a broader evaluation of professional skills across nine areas, which constrained the inclusion of additional items specific to digital skills. Future studies will examine digital skill levels in greater detail alongside the other eight surveyed competences. A stratified random sample of 529 employees from a mining organisation was surveyed in 2025, yielding several key findings. Respondents demonstrated relatively low digital competence, with a mean score of 2.41 on a five-point scale. The perceived need for digital training was higher, with a mean of 3.44. Mean interest in learning new skills was 3.80, and mean interest in further development was 3.83. The most significant barriers were health or age (mean = 3.19), lack of support (mean = 3.02), and lack of time (mean = 2.99). The hypothesis of a positive relationship between digital competence and educational need was confirmed, aligning with previous research indicating that higher levels of formal education are associated with greater interest in further professional education. Conversely, adults with lower levels of formal education are less likely to participate in informal or corporate education. The hypotheses linking educational needs to interest in learning new skills and developing personal abilities were also supported. The overall mean barrier index did not significantly predict educational need.

Acknowledgement

This contribution is part of the research initiative under the Operational Programme Just Transformation, specifically the REFRESH (Research Excellence for Region Sustainability and High-tech Industries) project, registration number CZ10.03.01/00/22_003/0000048. The authors gratefully acknowledge this support.

Ethics Declaration

We did not need ethical clearance for the research referred to in this paper.

AI Declaration

We did not use an artificial intelligence tool to create this article.

References

Cosgrove, J. and Cachia, R. (2025) *"DigComp 3.0: European Digital Competence Framework", Fifth Edition.* Publications Office of the European Union, Luxembourg, [online] https://data.europa.eu.

Deleersnyder, D., Fall, J., Prince, V. and Reed, M. (2022) *10 Things We Learned About Digital Skills During the Pandemic*. The Aspen Institute. Washington, DC.

European Commission (2019) "Who participates in adult education and how", Publications Office of the European Union, Luxembourg, [online] https://www.npi.cz/images/EAAL/%C3%9A%C4%8Dastn%C3%ADci_vzd%C4%9Bl%C3%A1v%C3%A1n%C3%AD_dosp%C4%9Bl%C3%BDch.pdf .

European Commission (2026) "JRC DigComp 3.0", [online] https://joint-research-centre.ec.europa.eu/projects-and-activities/education-and-training/digital-transformation-education/digital-competence-framework-digcomp/digcomp-30_en.

European Commission (2023) "European Declaration on Digital Rights and Principles for the Digital Decade 2023/C 23/01", [online] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023C0123(01).

Government Office of the Czech Republic (2024) "Educational platform of the digital Czech Republic", [online] https://osveta.gov.cz/theme/learnr/pages/imprint.php.

Halbová, B., Bird, L., Havlíčková, M., Šťastný, V. and Basl, J. (2024) *Computer and information literacy. Informatics thinking*. National report ICILS 2023, Czech School Inspectorate, Prague.

Hlaďo, P., Kvasková, L., Ježek, S., Hirschi, A. and Macek, P. (2020) "Career adaptability and social support of vocational students leaving upper secondary school", *Journal of Career Assessment*, Vol 28, No. 3, pp 478–495.

Hlaďo, P. and Kvasková, L. (2020) *Career adaptability. Its forms, transformations, connections and roles in the lives of young adults undergoing secondary vocational education,* Masaryk University, Brno.

Kalenda, J., Kočvarová, I. and Vaculíková, J. (2024) Barriers in Adult Education: Empirical Findings from the Czech Republic, In: Rabušicová, M. (eds.) *Adult learning and education. Contexts, theories and research*, Grada, Prague, pp 273–293.

Malach, J. (2024) Adults and their educational needs. Andragogic perspective. In: Rabušicová, M. (eds.) *Adult learning and education. Contexts, theories and research*, Grada, Prague, pp 177–200.

Palán, Z. (2002) *Basics of andragogy,* Vysoká škola J. A. Komenského, Prague.

Prusáková, V. (2010) Theoretical starting points for the analysis of educational needs of adults*,* In: Prusáková, V., et al. (eds.) *Analysis of educational needs of adults. Theoretical starting points*, Matej Bel University, Banská Bystrica, pp 13–44.

Průcha, J. (2014) "Research on the effectiveness of adult education", *Lifelong Learning*, Vol 4, No. 1, pp 8–22.

R Core Team (2024) "R: A language and environment for statistical computing", R Foundation for Statistical Computing, [online] https://www.R-project.org.

SIW-Skills Inventory Worksheet (2024) Allthings Admin, [online] https://worksheets.clipart-library.com/worksheet/skills-inventory-worksheet-11.html#google_vignette.

Šeďová, K. and Novotný, P. (2006) "Educational needs in relation to participation in adult education", *Pedagogika LVI*, Vol 2, pp 141–151.

Ślósarz, L., Błaszczyński, K., Švecová, M. and Kobylarek, A. (2022) "Adult education needs inventory: Construction and application", *Front. Psychol.,* Vol 13, pp. 1–11.

UK Council for Internet Safety (2025) "Digital Resilience Framework", [online] https://www.gov.uk/government/organisations/uk-council-for-internet-safety.

Towards a Multidimensional Framework for Educational Data Literacy

Flavio Manganello[1], Giannangelo Boccuzzi[1,2], Giancarlo Masi[1,3] and Chiara Fante[1]
[1]Institute for Educational Technology, National Research Council, Genoa, Italy
[2]University of Bologna, Bologna, Italy
[3]University of Modena and Reggio Emilia, Italy

flavio.manganello@cnr.it
giannangeloboccuzzi@cnr.it
giancarlomasi@cnr.it
chiara.fante@cnr.it

Abstract: Educational data literacy (EDL) has become a central concern for European schools as digitalization generates unprecedented volumes of sensitive data on minors, while the GDPR and AI Act impose constitutive constraints on data use. The conceptual landscape, however, remains fragmented: several traditions have proposed competency models, but each foregrounds one dimension while leaving others aside, and the boundary between EDL and adjacent constructs such as assessment literacy is still unresolved. This paper compares the four most influential traditions in the English-language literature: the Data Literacy for Teaching model, the Dutch Data-Based Decision Making tradition, the Learn2Analyze Educational Data Literacy Competence Profile, and the Critical Educational Data Literacy perspective. The comparison uses two axes: a dimensional axis of five facets (Technical, Pedagogical, Organizational, Ethical, Transversal) and a role axis distinguishing teachers from school leaders. The analytical method is a conceptual framework analysis rather than a systematic review, complementing recent empirical syntheses by asking whether the architectures organizing the field are adequate. The findings show that no tradition covers all five dimensions strongly, that role differentiation between teachers and school leaders is systematically lacking, that the organizational dimension is the most neglected, and that teachers' actual data use remains narrower than definitions suggest, with assessment data still dominating practice. From these findings, four design requirements for an integrated framework follow: multidimensional coverage with explicit interdependencies; role differentiation within a unified architecture; contextual adaptability through modular proficiency levels; and governance as institutional infrastructure. The discussion grounds these requirements in the DIKW hierarchy, aligns them with European regulatory and data sovereignty requirements, and draws implications for teacher education, school leadership development, and European policy on professional data competence, translating analytical findings into actionable design parameters.

Keywords: Educational Data Literacy, Competency Framework, Data-Based Decision Making, Data Governance, School Leadership, Data Wisdom

1. Introduction

The digitalization of European schools has generated an unprecedented volume of educational data. School information systems, learning platforms, and standardised testing produce data that could support instructional improvement, equitable resource allocation, and evidence-informed governance. The breadth and sensitivity of the data is often overlooked. Schools collect information far beyond grades and test scores, including behavioural assessments, family circumstances, special educational needs, and confidential records on minors, no less sensitive than health data. European policy reflects both potential and risks: the Digital Education Action Plan 2021-2027 calls for strengthening data literacy among educators (European Commission 2020), and the AI Act sets stringent requirements for high-risk AI systems in education (European Parliament and Council 2024).

Despite this policy momentum, EDL remains under-theorized. A cross-model comparison of twenty data literacy frameworks found widely divergent competency sets, with no single model covering all aspects applicable in a given context (Downes 2023). This fragmentation recurs within the educational sub-domain, where pedagogical, organizational, ethical, and governance considerations add complexity generic models do not address (Schumacher and Ifenthaler 2025). Several scholarly traditions have proposed competency models for EDL but developed largely in isolation, each foregrounding one dimension while leaving others aside. As a result, EDL is still insufficiently distinguished from related concepts such as assessment literacy, and fragmented approaches persist across countries. A cross-country study of five European nations found that no country has a dedicated strategy for data literacy in compulsory education, that curricular integration remains fragmented, and that no shared operational definition exists (Kostas et al. 2025).

The gap is structural, not merely definitional. As schools rely increasingly on commercial platforms, decision-makers without organizational and ethical data literacy cannot adequately assess governance, data ownership,

or the boundary between institutional and third-party data. A systematic comparison of EDL frameworks is therefore needed not to catalogue what exists but to identify where coverage breaks down and what follows.

This paper compares the four most influential EDL traditions along two axes: a dimensional axis (Technical, Pedagogical, Organizational, Ethical, Transversal) and a role axis (teachers, school leaders). Three research questions guide the inquiry:

RQ1: To what extent do existing EDL traditions cover each of the five dimensions?

RQ2: How, and to what degree, do the traditions differentiate the competencies required of teachers from those required of school leaders?

RQ3: What structural gaps emerge from the comparison, and what design requirements follow for an integrated, multidimensional EDL framework?

2. Analytical Method

This study adopts a conceptual framework analysis approach (Jabareen 2009) rather than a systematic review. The goal is not to catalogue empirical findings but to compare the conceptual architectures proposed to define and operationalize EDL. The strategy builds on a cross-model analysis of twenty data literacy definitions (Downes 2023), with two extensions: an explicit dimensional matrix replaces the competency list, and a role axis separates teachers from leaders.

The approach complements existing systematic reviews, which catalogue individual studies: a scoping review of 276 studies on teachers from 2020 to 2024 (Pastore 2026), a review of 95 indicators across 83 studies from 1990 to 2021 (Lee et al. 2024a), and one focused on principals (Lee et al. 2024b). Empirical reviews tell us what has been studied; conceptual analysis asks whether the architectures guiding it are adequate.

Selection of traditions followed three criteria: publication in peer-reviewed outlets or as a validated deliverable of a publicly funded project; explicit focus on data literacy in primary, secondary, or teacher education; and articulation of a structured set of competencies. The selection is not exhaustive. Traditions in other languages or within specific national policy contexts, such as Data Wise (Boudett et al. 2013) or Earl and Katz (2006) on data leadership, may add further perspectives. The selection reflects which bodies of work have most shaped the international scholarly conversation in English.

The analytical matrix has two axes. The dimensional axis comprises five facets derived inductively across the selected traditions: Technical (collecting, managing, processing, visualizing data), Pedagogical (translating insights into instructional decisions), Organizational (embedding data practices in institutional governance), Ethical (privacy, fairness, consent, power), and Transversal (collaborative, communicative, reflective dispositions). The role axis distinguishes teachers from school leaders. The structure differs from the type-based classification adopted in the most comprehensive systematic review (Lee et al. 2024a), which captures what kind of competence is at stake; ours captures in which domain of practice a competence operates, making visible a structural fragmentation that a competence-type classification cannot.

Coverage is coded on a three-level scheme. Strong: the dimension is a core focus, articulated through multiple sub-competencies, supported by dedicated empirical work. Moderate: addressed but not primary. Limited or absent: mentioned peripherally or not as a distinct component. The coding reflects authorial interpretive judgment, an analytical proposition open to scrutiny rather than a definitive measurement.

3. Mapping the Landscape

3.1 The Data Literacy for Teaching (DLFT) Model

The most widely cited contribution to EDL in the English-language literature defines data literacy for teaching as the ability to transform information into instructional knowledge (Mandinach and Gummer 2016). The framework organizes knowledge and skills around the data use cycle: identifying problems, collecting data, transforming data into information, converting information into decisions, and evaluating outcomes. Data-literate teachers must integrate content, pedagogical content, and curriculum knowledge to contextualize data within meaningful instructional frames.

The primary strength is the granular articulation of the pedagogical dimension: data use is explicitly linked to professional judgment, with statistical summaries treated as insufficient without interpretation. Subsequent

work connected teacher data literacy to teaching analytics (Ndukwe and Daniel 2020) and framed DLFT as a meta-construct subsuming assessment literacy as one component (Beck and Nunnaley 2021). A scoping review of 276 studies found that virtually all definitions of teacher data literacy in the literature align with DLFT (Pastore 2026). The limitations are substantial: developed within the US accountability context, the model tends to link data use with standardized test scores; ethical considerations such as algorithmic bias, consent, and the power dynamics embedded in data systems are not foregrounded; the organizational dimension receives only incidental attention; and the framework offers no competency profile for school leaders.

3.2 The Dutch Data-Based Decision Making Tradition

A parallel, largely independent tradition emerged at the University of Twente, organized around Data-Based Decision Making (DBDM). The competency model is a cyclical process: setting a purpose, collecting data, checking quality, analyzing and interpreting it with professional knowledge, and taking instructional action (Kippers et al. 2018). A consequential feature is that the cycle begins with goals, not data, positioning data use as an instrument of school improvement rather than an end. A second feature is the sustained emphasis on collaborative data use within Professional Learning Communities and data teams: data practices are treated as inherently social. A meta-analysis of 27 intervention studies found a medium effect ($g = .41$) on student achievement, with effective interventions marked by collaboration and school-level support (Spiele et al. 2025).

DBDM is the only tradition reviewed here giving sustained attention to school leaders, though within the same programme rather than a separate framework. Research identifies leadership behaviours that support data teams: establishing vision, providing individualized support, creating a safe improvement-focused climate, and using data personally (Schildkamp 2019, Schildkamp et al. 2017). These behaviours are described as enablers of the DBDM process rather than as a distinct competency profile. Ethical considerations remain peripheral, and the organizational dimension is addressed primarily through PLC facilitation rather than institutional governance in a broader sense; recent work also shows that students remain largely passive data sources (Schildkamp et al. 2025).

3.3 The Learn2Analyze Educational Data Literacy Competence Profile

The Learn2Analyze project, funded under Erasmus+, developed the Educational Data Literacy Competence Profile (EDL-CP) through expert consultation and validation with 210 professionals (Papamitsiou et al. 2021). The profile comprises five competence areas: Data Collection, Data Management, Data Analysis, Data Comprehension and Interpretation, and Data Application and Communication. The validation confirmed the dimensions as a foundation for EDL development while revealing that the surveyed professionals were not yet EDL-competent; the framework was later elaborated in a dedicated monograph (Sampson et al. 2022).

The L2A-EDL-CP stands out for its attention to the full data lifecycle, including storage, curation, and quality assurance other traditions overlook; its technical dimension is the most thoroughly articulated of the four, and it includes learning analytics competencies relevant to digitally mediated environments. Its limitations concern scope: data protection is mentioned but the broader terrain of fairness, consent, power, and epistemic justice is not systematically addressed; governance is not treated as a distinct dimension; and although extended from instructional designers to K-12 teachers (Sampson et al. 2022), it does not differentiate competencies by professional role.

3.4 The Critical Educational Data Literacy Perspective

A fundamentally different approach has emerged from critical data studies, data feminism, and critical digital literacy. Early contributions argued for moving beyond DigComp to address the specificities of EDL for educational professionals (Raffaghelli 2017, 2019), developed into a socio-political repositioning of EDL (Raffaghelli and Stewart 2020): data systems are not neutral instruments but embody power relations, institutional logics, and epistemic hierarchies that shape whose knowledge counts. The most recent articulation proposes a multidimensional model that includes technical and interpretive skills alongside critical awareness, ethical positioning, and what is termed data activism, the capacity to challenge or refuse data practices that undermine educational values (Raffaghelli et al. 2020). This tradition makes the most developed contribution to the ethical dimension and corrects the instrumentalism of earlier models.

The limitations mirror the strengths. The pedagogical dimension receives less granular treatment than in DLFT or DBDM: the critical perspective explains why data practices should be questioned but offers limited guidance

on how teachers might use data constructively. The organizational dimension, addressed through institutional data cultures, is not operationalized into competency statements. The framework remains primarily theoretical: unlike DBDM, it has not been validated through school-based work, and its adoption in professional development or policy documents has been limited. This gap is itself a finding: conceptually important traditions may remain marginal to practice if they do not translate insights into operational tools.

4. Comparative Analysis

Table 1 synthesizes the coverage of each tradition across the five dimensions and two professional roles. A three-level scheme is used: strong coverage (●), moderate coverage (◑), and limited or absent coverage (○).

Table 1: Coverage of EDL traditions across five dimensions and two professional roles

Dimension	DLFT	DBDM	L2A-EDL-CP	Critical EDL
Technical	○ Implicit; assessment data focus	◑ Moderate; quality checks, basic analysis	● Strong; full lifecycle incl. LA	○ Limited; deconstructive rather than applied
Pedagogical	● Strong; instructional decision-making	● Strong; action-oriented cycle	◑ Moderate; present but not school-specific	○ Limited; not central
Organizational	○ Peripheral	◑ Moderate; PLC structures, data teams, leadership as enabler	○ Peripheral	◑ Moderate; data cultures discussed
Ethical	○ Minimal	○ Minimal	○ Limited; data protection mentioned	● Strong; power, fairness, activism
Transversal	◑ Moderate; knowledge integration	● Strong; PLC collaboration, sensemaking	◑ Moderate; communication skills	● Strong; reflexivity, agency
Role: Teacher	● Primary target	● Primary target	◑ Generic educator	◑ Generic educator
Role: School Leader	○ Not addressed	◑ Addressed as enabler; no distinct profile	○ Not differentiated	○ Not differentiated

Regarding **RQ1**, no tradition covers all five dimensions strongly. DLFT and DBDM are well developed pedagogically but neglect ethics; the critical perspective offers the deepest ethical analysis but underdevelops pedagogy and organization; L2A-EDL-CP is technically the most comprehensive but lacks organizational and ethical depth. The pattern confirms what has been observed in the broader data literacy domain: each model reflects the context in which it was derived, and none covers the full spectrum (Downes 2023). The spread of 95 indicators across 83 empirical studies (Lee et al. 2024a) suggests the field has generated granularity at the component level while lacking integrative structure.

Regarding **RQ2**, differentiation is systematically lacking. DLFT and L2A-EDL-CP target teachers or generic educators. The critical perspective addresses a generic educator. DBDM has given the most sustained attention to leadership, but school leaders are treated as enablers of the data use process rather than as holders of a distinct profile (Schildkamp 2019, Schildkamp et al. 2017). A recent systematic review of principals' data literacy confirms that most studies on school leader data use employ the DLFT definition, designed for teachers, rather than a leadership-specific model (Lee et al. 2024b). The role-defined approach proposed in the broader data literacy field (Downes 2023) has not been taken up in education.

Regarding **RQ3**, four structural gaps stand out. First, the organizational dimension is the most neglected. DBDM addresses it through PLC structures but does not extend to institutional governance in a broader sense. Schools without organizational data competence become vulnerable to outsourcing decisions made without capacity to evaluate governance implications, with data sovereignty transferred to actors whose incentives may not align with educational values. Individual data literacy, however strong, has limited protective effect without supportive institutional infrastructure (Schildkamp 2019, Schildkamp et al. 2017). Second, the relationship between the ethical and technical dimensions remains unresolved: the critical perspective positions ethics as a lens through which all data practices should be evaluated, while L2A-EDL-CP treats data protection as a subsidiary technical concern.

Third, the translation from conceptual breadth to practice is incomplete: although DLFT encompasses diverse data sources, teachers overwhelmingly rely on assessment data (Pastore 2026), and a survey of 658 education professionals across five European countries confirms data use is primarily administrative and oriented toward external reporting (Kostas et al. 2025). Qualitative sources often provide more actionable insight than achievement data alone (Kippers et al. 2018, Schildkamp 2019). Fourth, no tradition addresses all five dimensions at multiple proficiency levels within a role-differentiated structure, so no model can guide a comprehensive, role-adapted professional development pathway.

5. Discussion

The four traditions were not designed to be complementary: they emerged from different disciplinary contexts (educational psychology, learning analytics, critical pedagogy) and policy environments (US accountability, Dutch school improvement, European digital competence policy). The result is incommensurability: different ontologies of competence, levels of granularity, and assumptions about who the data-literate professional is.

Theoretically, the field lacks a shared vocabulary. The boundary with assessment literacy in particular remains unresolved: although the two are claimed distinct (Beck and Nunnaley 2021), the data teachers actually engage with remain overwhelmingly assessment-centred (Pastore 2026), and across five European countries teachers were largely unable to distinguish data literacy from digital, AI, or statistical literacy (Kostas et al. 2025). Practically, the absence of an integrated model fragments professional development: a DLFT-grounded programme may produce educators who interpret assessment data but cannot navigate GDPR or join data governance, while a DBDM-informed one may cultivate collaborative inquiry without the ethical competencies needed to evaluate the tools and dashboards teams adopt. The fragmentation also narrows the goals data use serves: equity, wellbeing, safety, inclusion, and civic participation fall outside professional competence when frameworks are anchored to instructional decision-making and assessment data.

5.1 Design Requirements for an Integrated Framework

5.1.1 Multidimensional Coverage with Explicit Interdependencies

All five dimensions must be addressed as equally important and interdependent. Data use in schools mobilizes technical, pedagogical, ethical, organizational, and transversal competencies at once, and a teacher reading a learning analytics dashboard exemplifies this co-activation. An integrated framework should make these interdependencies explicit through cross-dimensional descriptors and operationalize the full range of competence types, including the dispositions and behaviours that empirical evidence identifies as central to teacher data literacy (Lee et al. 2024a). Frameworks defining competencies only as what teachers should know and do are structurally incomplete.

5.1.2 Role Differentiation Within a Unified Architecture

Teachers and school leaders interact with data in fundamentally different ways: teachers' practices are oriented toward classroom decisions, while leaders operate at institutional level on planning, resource allocation, and conditions for data use. A role-defined approach (Downes 2023), proposed in the broader domain but not yet applied to education, could specify complementary profiles within a single dimensional structure, sharing the five dimensions but instantiating them according to scope and authority. Student agency in relation to data, given that students remain largely passive data sources (Schildkamp et al. 2025), should be considered a transversal principle that operates across both profiles.

5.1.3 Contextual Adaptability Through Modular Proficiency Levels

European education systems differ in governance structures, data infrastructure maturity, and professional cultures, and recent projects address this through co-creation with national stakeholders (Kostas et al. 2025). An integrated framework should be modular, defining core competencies that apply across contexts while allowing localization, with proficiency levels from emergent to strategic that allow schools to locate themselves and identify next steps. A precedent exists in DLFT, where a continuum from novice to expert has been proposed (Beck and Nunnaley 2021); extending it to a multidimensional, role-differentiated framework requires specifying levels across all five dimensions simultaneously.

5.1.4 Governance as Institutional Infrastructure

Governance is treated by the frameworks reviewed, when at all, as a set of policies. The dynamics of data use in schools suggest governance is better understood as infrastructure: embedded, participatory mechanisms that operate continuously rather than episodically. When governance is treated as an afterthought, decisions are made by default rather than design, with commercial platforms filling gaps that institutional capacity has left open. An integrated framework should position governance as the architecture within which all other competencies operate, including, at the organizational level, competencies in procurement and vendor assessment, data ownership, and the boundary between institutional and third-party platforms. A related concern is reconciling accountability and improvement (Earl and Katz 2006, Ehren and Swanborn 2012, Schildkamp et al. 2014); DBDM has shown that external evaluation can serve as input for internal data use cycles when embedded within a culture of trust rather than blame (Schildkamp 2019, Schildkamp et al. 2017), and an integrated framework should build in the competencies to navigate this tension and resist the narrowing effects high-stakes accountability has on the goals data use serves.

5.2 Alignment with the European Regulatory and Institutional Landscape

A contemporary EDL framework must internalize the regulatory landscape as a structural design parameter. The GDPR (European Parliament and Council 2016) and the AI Act (European Parliament and Council 2024) create a legal environment in which educational data use is subject to transparency, data minimization, purpose limitation, and human oversight. These are constitutive constraints, not external add-ons. DigComp 2.2 (Vuorikari et al. 2022) and DigCompEdu (Redecker 2017) provide European scaffolding but lack the granularity an integrated EDL framework requires; parallel work in higher education shows the field moving toward models with explicit progression levels (Echtenbruck et al. 2025).

The intersection of data and AI literacy is an emerging frontier. As AI-powered tools enter schools, educators must evaluate transparency and explainability, the representativeness of training data, and possible bias in AI-generated recommendations. These competencies sit at the intersection of the technical and ethical dimensions, and the AI Act makes their integration legally necessary.

The DIKW hierarchy (Ackoff 1989) offers an epistemic scaffold for integration. Applied to EDL, progression from raw data through contextualized interpretation to evidence-informed practice and, ultimately, principled data governance requires the simultaneous activation of all five dimensions at increasing levels of sophistication. The highest level, Wisdom, is the capacity to ask not only what data tools can do but whether and how they should be used: the Know Why that enables teachers and leaders to assess power asymmetries, protect learners' rights, and resist inappropriate data mandates. This is precisely where the European Common Data Space for Education and Skills, premised on educational data remaining within European governance frameworks, makes integration a practical horizon: an EDL framework that does not equip school leaders to assess whether a platform aligns with these requirements falls short of what the policy environment demands.

5.3 Implications for Practice

For teacher educators, the persistence of assessment-centred training suggests programme design should move beyond standalone data courses toward integration across the curriculum: fewer than four in ten educators surveyed across five European countries had received formal data literacy training (Kostas et al. 2025). For school leaders, the absence of a dedicated competency profile is both a gap and an opportunity, with the design requirements above pointing to competencies in governance, procurement, team facilitation, and the mediation of accountability. For policy-makers, anchoring professional standards to a single tradition risks neglecting the ethical and organizational dimensions GDPR and the AI Act now require; a policy framework for EDL should be multidimensional by design and accompanied by differentiated pathways for teachers and school leaders.

6. Conclusion

This paper has mapped and compared four traditions that define the current landscape of EDL. Each makes a distinctive contribution, but none provides the comprehensive, role-differentiated, contextually adaptable model European education systems now require. The contributions are threefold. First, the five-dimension analytical matrix offers a reusable lens for evaluating EDL traditions as new models emerge. Second, the comparative analysis specifies where coverage breaks down and why the organizational and ethical gaps carry particular practical weight as schools rely on commercial platforms. Third, the four design requirements

(multidimensional coverage with interdependencies, role differentiation within a unified architecture, contextual adaptability through proficiency levels, and governance as institutional infrastructure) translate analytical findings into actionable parameters, grounded in the DIKW hierarchy and aligned with European regulatory and data sovereignty requirements.

The analysis has limitations. The selection of four traditions may not capture every relevant contribution; models such as Data Wise (Boudett et al. 2013) or Earl and Katz (2006) may offer complementary perspectives. The dimensional matrix is an inductive construct, and the three-level coding reflects authorial interpretive judgment without inter-rater reliability checks. Conceptual architectures are evaluated rather than empirical outcomes. Future research should prioritize empirical validation through school-based interventions across diverse European contexts, assessment instruments measuring EDL across all five dimensions and proficiency levels, and investigation of how organizational conditions mediate the link between individual data literacy and institutional outcomes. Meta-analytic evidence shows that data use professional development positively affects student achievement (g = .41), with heterogeneity better explained by school-level implementation conditions than by intervention design (Spiele et al. 2025); understanding these conditions is the next frontier.

Acknowledgements

This work has been supported by the SCHOOLWISE project (Grant Agreement no. 101259617), funded by the European Union under the Erasmus+ Programme, call ERASMUS-EDU-2025-PI-FORWARD. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency.

Ethics Declaration

Ethical clearance was not required.

AI Declaration

Generative AI was used only for final language editing and formatting, while all conceptual, analytical, and interpretive work was carried out by the authors.

References

Ackoff, R.L. (1989) “From data to wisdom”, Journal of Applied Systems Analysis, Vol 16, No. 1, pp 3-9.

Beck, J.S. and Nunnaley, D. (2021) “A continuum of data literacy for teaching”, Studies in Educational Evaluation, Vol 69, Article 100871. https://doi.org/10.1016/j.stueduc.2020.100871

Boudett, K.P., City, E.A. and Murnane, R.J. (2013) Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning, rev. ed., Harvard Education Press, Cambridge.

Downes, S. (2023) “Three frameworks for data literacy”, in Sampson, D.G., Ifenthaler, D. and Isaías, P. (eds.) Proceedings of the 20th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2023), IADIS Press, pp 107-115.

Earl, L.M. and Katz, S. (2006) Leading Schools in a Data-Rich World: Harnessing Data for School Improvement, Corwin Press, Thousand Oaks.

Echtenbruck, M.M., Fühles-Ubach, S., Naujoks, B. and Kaliva, E. (2025) “A data literacy competence model for higher education and research”, arXiv preprint arXiv:2504.15690. https://doi.org/10.48550/arXiv.2504.15690

Ehren, M.C.M. and Swanborn, M.S.L. (2012) “Strategic data use of schools in accountability systems”, School Effectiveness and School Improvement, Vol 23, No. 3, pp 257-280. https://doi.org/10.1080/09243453.2011.652127

European Commission (2020) Digital Education Action Plan 2021-2027, COM(2020) 624 final, European Commission, Brussels.

European Parliament and Council (2016) Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data (GDPR), Official Journal of the European Union.

European Parliament and Council (2024) Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act), Official Journal of the European Union.

Jabareen, Y. (2009) “Building a conceptual framework: Philosophy, definitions, and procedure”, International Journal of Qualitative Methods, Vol 8, No. 4, pp 49-62. https://doi.org/10.1177/160940690900800406

Kippers, W.B., Poortman, C.L, Schildkamp, K. and Visscher, A.J. (2018) “Data literacy: What do educators learn and struggle with during a data use intervention?”, Studies in Educational Evaluation, Vol 56, pp 21-31. https://doi.org/10.1016/j.stueduc.2017.11.001

Kostas, A., Paraschou, V., Tzortzoglou, F., Kritikos, G., Pitsikalis, S. and Lasica, I.-E. (2025) Data Literacy Strategies in Compulsory Education, Deliverable D2.1, DATA-READY Project, European Commission. https://data-ready.eu/wp-content/uploads/2025/08/DATA-READY-D2.1-290625-FINAL.pdf

Lee, J., Alonzo, D., Beswick, K., Abril, J.M.V., Chew, A.W. and Oo, C.Z. (2024a) "Dimensions of teachers' data literacy: A systematic review of literature from 1990 to 2021", Educational Assessment, Evaluation and Accountability, Vol 36, No. 2, pp 145-200. https://doi.org/10.1007/s11092-024-09435-8

Lee, J., Alonzo, D., Beswick, K., Oo, C.Z., Anson, D.W.J. and Abril, J.M.V. (2024b) "Data literacy of principals in K-12 school contexts: A systematic review", Educational Research Review, Vol 45, Article 100649. https://doi.org/10.1016/j.edurev.2024.100649

Mandinach, E.B. and Gummer, E.S. (2016) "What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions", Teaching and Teacher Education, Vol 60, pp 366-376. https://doi.org/10.1016/j.tate.2016.07.011

Ndukwe, I.G. and Daniel, B.K. (2020) "Teaching analytics, value and tools for teacher data literacy: A systematic and tripartite approach", International Journal of Educational Technology in Higher Education, Vol 17, Article 22. https://doi.org/10.1186/s41239-020-00201-6

Papamitsiou, Z., Filippakis, M.E., Poulou, M., Sampson, D., Ifenthaler, D. and Giannakos, M. (2021) "Towards an educational data literacy framework: Enhancing the profiles of instructional designers and e-tutors of online and blended courses with new competences", Smart Learning Environments, Vol 8, Article 18. https://doi.org/10.1186/s40561-021-00163-w

Pastore, S. (2026) "Teacher data literacy: A scoping review (2020-2024)", Cogent Education, Vol 13, No. 1, Article 2592397. https://doi.org/10.1080/2331186X.2025.2592397

Raffaghelli, J.E. (2017) "Data literacy in the context of big and open data: An educational challenge", Formazione & Insegnamento, Vol 15, No. 3, pp 299-324. https://doi.org/10.7346/-fei-XV-03-17_21

Raffaghelli, J.E. (2019) "Developing a framework for educators' data literacy in the European context: Proposal, implications and debate", in EDULEARN19 Proceedings, IATED, pp 10520-10530. https://doi.org/10.21125/edulearn.2019.2655

Raffaghelli, J.E., Manca, S., Stewart, B., Prinsloo, P. and Sangrà, A. (2020) "Supporting the development of critical data literacies in higher education: Building blocks for fair data cultures", International Journal of Educational Technology in Higher Education, Vol 17, Article 58. https://doi.org/10.1186/s41239-020-00235-w

Raffaghelli, J.E. and Stewart, B. (2020) "Centering complexity in 'educators' data literacy' to support future practices in faculty development: A systematic review of the literature", Teaching in Higher Education, Vol 25, No. 4, pp 435-455. https://doi.org/10.1080/13562517.2019.1696301

Redecker, C. (2017) European Framework for the Digital Competence of Educators: DigCompEdu, Publications Office of the European Union, Luxembourg.

Sampson, D., Papamitsiou, Z., Ifenthaler, D., Giannakos, M., Mougiakou, S. and Vinatsella, D. (2022) Educational Data Literacy, Springer, Cham. https://doi.org/10.1007/978-3-031-11705-3

Schildkamp, K. (2019) "Data-based decision-making for school improvement: Research insights and gaps", Educational Research, Vol 61, No. 3, pp 257-273. https://doi.org/10.1080/00131881.2019.1625716

Schildkamp, K., Karbautzki, L. and Vanhoof, J. (2014) "Exploring data use practices around Europe: Identifying enablers and barriers", Studies in Educational Evaluation, Vol 42, pp 15-24. https://doi.org/10.1016/j.stueduc.2013.10.007

Schildkamp, K., Poortman, C., Luyten, H. and Ebbeler, J. (2017) "Factors promoting and hindering data-based decision making in schools", School Effectiveness and School Improvement, Vol 28, No. 2, pp 242-258. https://doi.org/10.1080/09243453.2016.1256901

Schildkamp, K., Fosnæs, A.R., Lindvig, Y. and Wærness, J.I. (2025) "Student participation in data-informed decision making: From passive data sources to active data users", Journal of Professional Capital and Community, Vol 10, No. 3, pp 337-351. https://doi.org/10.1108/JPCC-06-2024-0100

Schumacher, C. and Ifenthaler, D., eds. (2025) International Perspectives on Educational Data Literacy: Frameworks, Contexts, and Practices, Routledge, London

Spiele, S., Janssen, J.J.H.M., Schildkamp, K. and Poortman, C.L. (2025) "The effect of teacher data use professional development interventions on student achievement: A meta-analysis and qualitative comparative analysis", Studies in Educational Evaluation, Vol 87, Article 101528. https://doi.org/10.1016/j.stueduc.2025.101528

Vuorikari, R., Kluzer, S. and Punie, Y. (2022) DigComp 2.2: The Digital Competence Framework for Citizens - With New Examples of Knowledge, Skills and Attitudes, JRC Report No. JRC128415, Joint Research Centre, Seville.

Digital Learning Environments and Student Satisfaction: A Learning Analytics Perspective

Sarmīte Mickeviča and Signe Bāliņa
University of Latvia, Riga, Latvia

sarmite.mickevica@hotmail.com
signe.balina@lu.lv

Abstract: Higher education is increasingly characterised by heterogeneous student populations, digitally mediated learning environments, and the growing use of artificial intelligence (AI). This study examines how student complexity, digital readiness, learning-environment preferences, and perceived employability may inform the development of AI-enabled Student Relationship Management (SRM). The conceptual framework integrates relationship marketing, the Industrial Marketing and Purchasing (IMP) network perspective, the Nordic School's value co-creation logic, and learning analytics. Using EUROSTUDENT 8 data, the study combines a comparative descriptive analysis of six European higher education systems with exploratory correlation analysis across all countries for which comparable indicators are available. The findings reveal a pronounced "working student" profile among mature learners, characterised by substantial employment alongside continued academic commitments. Digital access and self-assessed professional digital skills are comparatively well-established across age groups, while high digital readiness shows no meaningful association with high perceived employability. Most importantly, greater online-learning intensity does not automatically produce stronger alignment with student preferences. Higher employment intensity is associated with stronger preferences for online learning, yet some mature student groups experience more online provision than they would ideally prefer. The findings therefore highlight a critical distinction between digitalisation and flexibility. AI-enabled SRM may contribute by supporting adaptive alignment between institutional delivery models, student circumstances, and differentiated support needs rather than by increasing digital intensity alone. The study contributes to educational technology and relationship-management literature by demonstrating that the strategic necessity of AI-enabled SRM arises primarily from growing student complexity and institutional adaptability challenges rather than from deficiencies in baseline digital readiness.

Keywords: Learning Analytics; Digital Learning Environments; Higher Education; Student Relationship Management; Digital Readiness; Student Heterogeneity

1. Introduction

Higher education is increasingly transforming from institution-centred systems into interconnected, data-driven digital ecosystems (Benavides et al., 2020). Within these ecosystems, AI applications are expanding across areas including prediction, assessment and evaluation, personalised support, intelligent tutoring, and student-learning management (Crompton & Burke, 2023; Bond et al., 2024). From a relationship-marketing and network perspective, universities no longer operate as isolated educational providers, but as actors embedded within complex value co-creation systems involving students, employers, digital platforms, and institutional stakeholders. Within a public service logic framework, higher education institutions function primarily as value facilitators, providing structural and digital resources that students, as active value creators, integrate into their everyday practices to generate value-in-use (Grönroos, 2019).

At the macro level, higher education ecosystems are becoming increasingly heterogeneous due to shifting student demographics, including the growing participation of non-traditional, working, and lifelong learners, as well as increasing international mobility (OECD, 2019; Hauschildt, Vögtle & Gwosć, 2018; Hauschildt, 2024). At the same time, digital transformation has expanded the use of online learning environments, digital administration, and learning analytics (Benavides et al., 2020; Darvishi et al., 2024; Gašević et al., 2022). Together, these developments challenge traditional "one-size-fits-all" approaches and increase the need for flexible, personalised, and data-driven student support and institutional management.

Recent systematic reviews show that AI research in higher education remains largely organised around specific applications, including learning, teaching, assessment, administration, profiling, prediction, adaptive systems and personalisation (Chiu et al., 2023; Crompton & Burke, 2023; Bond et al., 2024). Existing AI research frequently focuses on profiling, automation, and dropout prediction (Zawacki-Richter et al., 2019), while comparatively less attention has been devoted to how student heterogeneity, digital readiness, learning-environment preferences, and employability perceptions can inform the development of AI-enabled SRM from a networked ecosystem perspective. Chiu et al. (2023) specifically call for a more holistic perspective on the role of AI in education, while recent evidence also indicates comparatively limited attention to administrative and institutional applications of AI in higher education (Crompton & Burke, 2023; Bond et al., 2024). Taken together

with the SRM literature (Gholami et al., 2018; Rigo et al., 2016), this suggests an underdeveloped connection between AI-enabled capabilities and broader SRM perspectives.

This study aims to examine patterns of student heterogeneity, digital readiness, learning-environment preferences, and perceived employability across European higher education systems and to explore their implications for the development of adaptive AI-enabled SRM. Drawing on comparative EUROSTUDENT 8 data and integrating relationship-marketing theory, ecosystem perspectives, and learning analytics, the study contributes to the understanding the strategic role of AI-enabled SRM in digitally transformed higher education institutions.

2. Theoretical Foundation

2.1 Relationship Marketing and Value Co-creation

Relationship marketing theory provides a theoretical foundation for understanding how universities manage interactions with students and other stakeholders. The Nordic School and the Industrial Marketing and Purchasing (IMP) Group are particularly relevant for analysing relationship management within higher education ecosystems.

The Nordic School emphasises interaction and value co-creation as central to long-term relationships. Within public service logic, organisations act primarily as value facilitators, while beneficiaries create value-in-use by integrating available resources into their own contexts (Grönroos, 2019; Gummesson, 2008; Payne & Frow, 2017). When applied to higher education, this framework positions students as active value creators rather than passive recipients, integrating institutional resources into their learning and work-life circumstances. From this perspective, AI-enabled SRM may support value co-creation by helping institutions align communication, services, and support with heterogeneous student profiles and constraints.

2.2 Higher Education as a Networked Ecosystem

Complementing the Nordic School perspective, the IMP Group conceptualises organisations as actors embedded within complex relationship networks rather than isolated entities (Håkansson & Snehota, 2006). Value creation therefore occurs within interconnected systems involving multiple stakeholders.

This network perspective can be extended through the concept of ecosystems, where interactions are structured across interdependent actors, technologies, and institutional processes. Universities increasingly operate within digitally enabled ecosystems involving students, employers, government institutions, and technology providers, reflecting the broader digital transformation of higher education (Benavides et al., 2020).

Following Adner (2017), ecosystems consist of interrelated actors, activities, positions, and links that collectively create value. Within higher education, these include teaching, learning, recruitment, support services, and institutional coordination processes. In this structure, students occupy a dual role as both customers and co-creators of educational value and knowledge.

AI functions within this ecosystem as an integrating mechanism capable of strengthening coordination, interaction, and data exchange across institutional networks. Consequently, AI-enabled SRM can be understood not merely as a technological tool, but as a relationship-management mechanism operating across interconnected higher education ecosystems.

2.3 Learning Analytics and the Evolution from CRM to AI-Enabled SRM

Learning analytics concerns the measurement, collection, analysis, and reporting of data about learners and their contexts to understand and optimise learning and learning environments (Siemens, 2012). In higher education, learning analytics has developed from descriptive indicators and dashboards towards prediction, intervention, personalised feedback, and institutional decision support (Gašević et al., 2022). Importantly, learning analytics models and insights are highly context-sensitive, reinforcing the need to move beyond “one-size-fits-all” approaches (Gašević et al., 2016).

Learning analytics and AI are also sociotechnical rather than neutral systems. Their design and implementation may create risks related to surveillance, reductionism, power asymmetries, and limited transparency if appropriate safeguards are absent (Buckingham Shum & Luckin, 2019; Mouta et al., 2024; Selwyn, 2019).

Nevertheless, learning analytics can provide an analytical foundation for AI-enabled SRM by identifying patterns relevant to student engagement, support needs, and institutional decision-making.

The evolution from Customer Relationship Management (CRM) towards Student Relationship Management (SRM) reflects a transition from transactional administration to student-centred relationship management. While higher education CRM has often focused on administrative functions (Rigo et al., 2016), SRM emphasises engagement, support, and long-term relationships (Gholami et al., 2018). AI may extend these capabilities through prediction, personalised support, AI-assisted interaction, and adaptive systems. Recent reviews identify prediction, AI assistants, intelligent tutoring, assessment, student-learning management, and personalisation among major AI applications in higher education (Crompton & Burke, 2023; Bond et al., 2024).

Building on these theoretical perspectives, this study conceptualises AI-enabled SRM as an integrated ecosystem capability comprising three interconnected components: strategic, operational, and analytical.

- The **strategic component** concerns adaptive value co-creation and institutional responsiveness to heterogeneous student needs. Drawing on the Nordic School and IMP Group perspectives, it focuses on aligning institutional resources, study pathways, and support with differentiated student profiles and circumstances (Grönroos, 2019; Håkansson & Snehota, 2006).
- The **operational component** concerns digitally mediated interaction and service delivery. AI-enabled SRM may extend beyond administrative automation through AI-assisted communication, personalised interaction, and adaptive student-support services (Rigo et al., 2016; Gholami et al., 2018; Crompton & Burke, 2023).
- The **analytical component** concerns learning analytics, prediction, and proactive support. It enables the identification of patterns related to engagement, academic risk, and potential dropout, supporting earlier institutional intervention and data-informed decision-making (Siemens, 2012; Gašević et al., 2022; Bond et al., 2024).

Overall, AI-enabled SRM can be conceptualised as a strategic ecosystem capability integrating relationships, institutional processes, data, and technology to support more adaptive, discipline-sensitive, and student-centred higher education management (Gholami et al., 2020). Within this framework, digitalisation functions not merely as a technological instrument, but as a strategic capability enabling institutional flexibility, responsiveness, and competitiveness within increasingly globalised and heterogeneous higher education ecosystems.

3. Methodology

3.1 Research Design and Data Source

This study employs a non-experimental comparative quantitative design based on secondary EUROSTUDENT 8 data (2021–2024), combining descriptive comparison and exploratory correlation analysis. It does not evaluate the implementation or effectiveness of an active AI-enabled SRM system. Instead, it examines patterns of student heterogeneity, digital readiness, time constraints, learning-environment preferences, and perceived employability that may inform the development of adaptive student-support models. The analysis is informed by a learning analytics perspective and draws on the EUROSTUDENT modules “Digitalisation”, “Time Budget”, “Assessment of Studies and Performance”, and “Equipment at Home”.

The exploratory correlation analysis uses all EUROSTUDENT 8 countries for which comparable indicators are available, examining aggregated macro-level associations rather than individual-level relationships. This broader analytical scope captures substantial cross-national variation in student characteristics and study conditions. The approach is consistent with learning analytics research emphasising the importance of context in which learning and technology use occur (Gašević et al., 2016). For detailed descriptive comparison and visualisation, six higher education systems were selected as an illustrative subset: Estonia (EE), Finland (FI), Lithuania (LT), Latvia (LV), the Netherlands (NL), and Poland (PL). This subset is used for contextual interpretation rather than statistical generalisation.

3.2 Analytical Framework and Selected Variables

The selected indicators are interpreted in relation to the three AI-enabled SRM dimensions developed in Section 2: operational, analytical, and strategic. These are not treated as directly measured constructs. The operational dimension is reflected in digital-service provision and alignment between actual and preferred study modes; the analytical dimension in student heterogeneity, time allocation, digital readiness, and perceived employability;

and the strategic dimension in the interpretation of these patterns in relation to potential adaptive institutional responses.

The empirical analysis is organised around four domains.

- **Student Complexity**. Age groups, fields of study based on ISCED-F 2013 classifications (UNESCO Institute for Statistics, 2015), and time allocation between taught studies, personal study, and paid employment.
- **Digital Readiness**. Self-assessed professional digital skills and access to essential digital infrastructure, including a computer and reliable internet connection.
- **Learning Environment Alignment**. Comparison between actual and preferred online and in-person study modes and satisfaction with digitally available services.
- **Perceived Employability.** Self-assessed preparedness for national and international labour markets, interpreted as perceived labour-market readiness rather than an objective employment outcome.

3.3 Data Analysis Procedures

Data were analysed using descriptive statistics, comparative analysis, and exploratory correlations.

- **Descriptive and Visual Analysis.** Cross-country comparisons and visualisations focus on the six-country subset. Student time budgets across age groups were presented using stacked bar charts. Learning-environment alignment was assessed by comparing reported actual and preferred online and in-person study modes, with differences interpreted as indicators of alignment or mismatch. Scatter plots were used to visualise these relationships.
- **Correlation Analysis.** Pearson correlations were calculated across aggregated country-by-age-group observations for all countries with comparable data. The analysis yielded $N = 92$ observations for time-budget, study-preference, and digital-readiness analyses ($df = 90$), and $N = 88$ for digital-readiness and employability analyses ($df = 86$). Fisher's r-to-z transformation was used to calculate 95% confidence intervals (CI). Given the aggregated structure of the data, correlations are interpreted as exploratory macro-level associations and do not imply causal or individual-level relationships.

4. Results and Comparative Analysis

4.1 Student Complexity and the 'Working Student' Model

EUROSTUDENT 8 data reveal substantial heterogeneity in students' time allocation between taught studies, personal study, and paid employment. Across the six-country subset, employment intensity increases considerably with age. In LV, students aged under 22 spend approximately 14 hours per week in paid employment, compared with approximately 23 hours among older age groups. Mature students (30+) in LV and PL spend 31.9 and 35.9 hours per week in paid employment, respectively, while maintaining substantial study commitments. In contrast, FI shows lower employment intensity among mature students (21.4 hours per week).

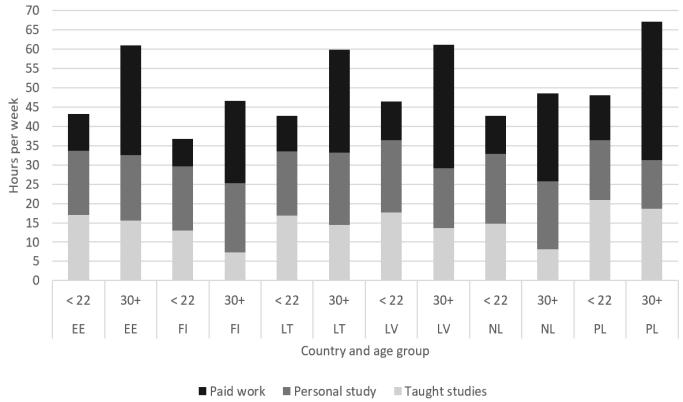


Figure 1: Weekly time-budget distribution of younger (<22) and mature (30+) students across the six-country comparative subset.

Figure 1 shows that mature students combine substantially higher employment intensity with continued academic commitments. In several systems, their combined weekly workload exceeds 57 hours. Differences by

field of study further contribute to student heterogeneity, with Medicine and Dental Studies showing particularly high study-time intensity and ICT students exhibiting comparatively high employment participation. These patterns indicate a pronounced "working student" profile among mature learners and highlight substantial variation in student time constraints.

4.2 Student Digital Readiness and Digital Capital

EUROSTUDENT 8 indicates a generally high baseline of digital capital across participating higher education systems. Cross-country mean values are high for self-assessed professional digital skills (4.3), access to computers (4.8), desks (4.6), sufficient internet connections (4.5), and quiet places to study (4.1) (Schirmer, 2024). Although country-level variation remains, these results indicate that most student populations have broadly favourable access to the skills and resources required for digitally enabled study.

Within the six-country subset, 59.7% of students aged under 22 (*SD* = 0.039) and 57.0% of students aged 30+ (*SD* = 0.027) assess their professional digital skills as "completely sufficient". In LV, the corresponding share is 58.6%, compared with 66.9% in EE and 64.7% in NL. These patterns are consistent with arguments that traditional assumptions concerning a generational digital divide and the technological superiority of so-called "digital natives" may be becoming less applicable in contemporary higher education contexts (Janschitz & Penker, 2022).

The exploratory correlation analysis across all available EUROSTUDENT 8 country-by-age-group observations indicates generally weak associations between student time allocation and self-assessed professional digital skills ($N = 92$, $df = 90$). The share of students reporting their digital skills as "completely sufficient" shows a weak negative association with time spent in taught studies ($r = -.142$, $p = .178$, 95% CI [-.337, .065]), a negligible positive association with personal study time ($r = .101$, $p = .337$, 95% CI [-.106, .300]), and a weak positive association with paid employment ($r = .250$, $p = .016$, 95% CI [.048, .433]). Similarly, the share reporting their digital skills as "not sufficient at all" is weakly associated with taught studies ($r = .048$, $p = .648$, 95% CI [-.158, .251]), personal study time ($r = -.224$, $p = .032$, 95% CI [-.410, -.020]), and paid employment ($r = -.035$, $p = .740$, 95% CI [-.238, .171]). By contrast, paid employment is moderately negatively associated with personal study time ($r = -.525$, $p < .001$, 95% CI [-.659, -.358]), indicating that increasing employment intensity is more clearly associated with time-allocation pressure than with lower perceived digital readiness. Overall, the results suggest that workload intensity and self-assessed digital readiness represent relatively distinct dimensions of student experience.

4.3 Learning Environment Alignment and Institutional Digital Provision

Actual and preferred online and in-person study modes vary substantially across countries and age groups. LV and FI are among the more online-oriented systems in the six-country subset, but greater online intensity does not necessarily correspond to stronger alignment with student preferences.

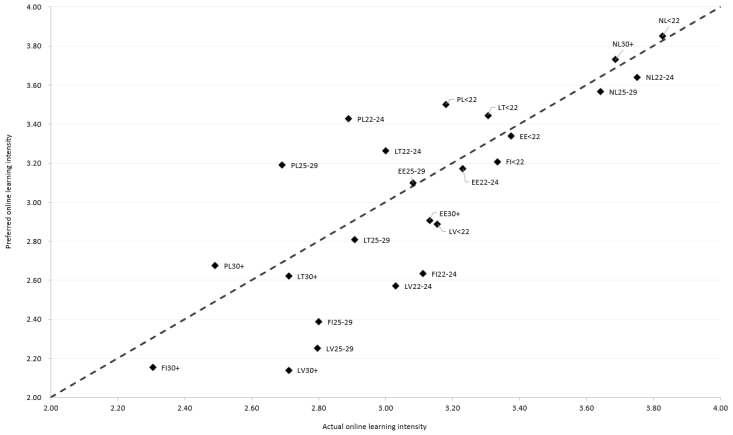


Figure 2: Alignment between actual and preferred online and in-person study modes across countries and age groups. The five-point scale ranges from predominantly online learning (1) to predominantly in-person learning (5). Points below the diagonal indicate more online learning than preferred; points above indicate a preference for more online learning than currently experienced.

Figure 2 shows that younger age groups generally lie closer to the alignment line, while several mature groups, particularly in LV and FI, display greater mismatches. Some mature groups therefore experience more online learning than they would ideally prefer.

Employment intensity is also associated with learning-environment preferences. Across country-by-age-group observations (N = 92), paid employment shows a strong positive association with a preference for completely online learning (r = .677, p < .001, 95% CI [.548, .774]) and a moderate negative association with a preference for predominantly in-person study (r = -.466, p < .001, 95% CI [-.612, -.289]). Thus, higher employment intensity is associated with stronger demand for online flexibility, while the alignment analysis shows that more online provision does not necessarily produce better alignment.

A second analytical component concerns students' satisfaction with the digital availability of institutional services. Descriptive results indicate substantial variation across service categories. Within the six-country comparative subset, satisfaction with the digital availability of online examinations is particularly high in LV (79%), with similarly high levels observed in EE and LT. Descriptively, however, satisfaction appears to vary considerably across different types of digitally available services, with counselling receiving comparatively lower positive evaluations than several transactional services.

4.4 Student Digital Readiness and Perceived Employability

Perceived labour-market preparedness varies across the six-country subset: NL shows comparatively high international preparedness, EE exhibits stronger national preparedness, while LV and LT report more moderate international preparedness. The exploratory correlation analysis (N = 88, df = 86) shows no meaningful association between "completely sufficient" self-assessed digital skills and high perceived employability. Associations are non-significant for both national (r = .034, p = .751, 95% CI [-.176, .242]) and international labour markets (r = -.080, p = .459, 95% CI [-.285, .132]). These findings indicate that high self-assessed digital readiness is not systematically associated with high perceived labour-market preparedness at the aggregated country-by-age-group level.

5. Discussion

The findings of this study indicate that the student experience in digitally enabled higher education cannot be interpreted independently of broader patterns of employment, time allocation, and study behaviour. The six-country comparison reveals a pronounced "working student" profile among mature learners, particularly in LV and PL, where substantial paid employment is combined with continued academic commitments. This pattern reinforces the limitations of uniform student-support models and is consistent with learning analytics research arguing that educational interventions should remain sensitive to contextual and learner heterogeneity rather than relying on "one-size-fits-all" approaches (Gašević et al., 2016).

Importantly, the findings suggest that flexibility and digitalisation should not be treated as synonymous concepts. Greater digital delivery does not automatically translate into stronger alignment with student preferences. Although working students demonstrate stronger preferences for online-oriented and flexible learning arrangements, mature student groups in several ecosystems – particularly LV and FI – also exhibit signs of potential digital overprovision, where actual online-learning intensity exceeds preferred levels. This indicates that effective digital transformation depends not on maximising online delivery itself, but on maintaining alignment between institutional delivery models and heterogeneous student preferences.

The findings further suggest that this alignment challenge cannot be explained simply by insufficient student digital readiness. Basic digital access and self-assessed professional digital skills are comparatively well-established across the analysed systems and age groups, consistent with the broader concept of digital capital as a combination of skills, access, and resources enabling participation in digitally mediated environments (Ragnedda et al., 2020). Moreover, high self-assessed digital readiness shows no meaningful association with high perceived labour-market preparedness. Digital competence may therefore be interpreted primarily as an enabling condition for participation in digitally transformed higher education rather than as a sufficient explanation of differences in student experience or perceived employability.

From an AI-enabled SRM perspective, these findings shift the strategic emphasis from increasing digital intensity towards adaptive alignment. The potential value of AI-enabled SRM lies not primarily in expanding the volume of digital interaction, but in helping institutions identify heterogeneous student profiles, recognise differing time constraints and learning preferences, and adjust communication and support accordingly. In this sense, AI-

enabled SRM may provide a mechanism for translating digitalisation into flexibility rather than treating the two as equivalent. Strategically, it may support alignment between institutional delivery models and differentiated student needs; operationally, AI-assisted communication and advising may extend the availability and responsiveness of support; analytically, learning analytics may help identify patterns requiring institutional attention. However, the present study does not evaluate the effectiveness of these AI-enabled mechanisms themselves. Rather, it identifies contextual conditions that may inform their future design and empirical validation.

At the same time, adaptive support should not be equated with automated decision-making. Learning analytics and AI systems are not neutral infrastructures; their design and implementation may embed institutional assumptions, reinforce existing inequalities, or reduce complex student experiences to simplified predictive categories (Buckingham Shum & Luckin, 2019; Mouta et al., 2024). Recent evidence also suggests that AI-supported performance does not necessarily translate into equivalent independent performance once assistance is removed, highlighting the importance of preserving learner agency and human judgement (Darvishi et al., 2024). AI-enabled SRM should therefore be implemented as a support mechanism rather than a deterministic management system, supported by clear governance concerning data privacy, transparency, accountability, equity, and ongoing evaluation.

Taken together, the findings suggest that the strategic challenge of digital transformation in higher education is shifting from basic technological access towards institutional adaptability. The key issue is not simply whether higher education becomes more digital, but whether digital provision remains aligned with heterogeneous student needs and circumstances. The distinction between digitalisation and flexibility is therefore central: greater digital intensity does not automatically produce better alignment with student preferences, whereas adaptive institutional capacity may be more important than technological intensity itself. Within this context, AI-enabled SRM represents a potential strategic mechanism for supporting more responsive and student-centred higher education while preserving meaningful student choice, agency, and human oversight.

6. Conclusions

This study empirically examined student and learning-environment conditions that may inform the future development of AI-enabled SRM by analysing student complexity, digital readiness, learning-environment preferences, and perceived employability. By integrating a learning analytics perspective with the IMP Group network perspective and the Nordic School's value co-creation logic, the study demonstrates the limitations of uniform "one-size-fits-all" approaches within increasingly heterogeneous higher education environments.

First, the findings reveal that student complexity is increasingly shaped by the emergence of the "working student" model. Mature students frequently combine substantial employment commitments with ongoing academic workloads, creating significant time-pressure conditions that increase the demand for flexible and online-oriented study arrangements. These patterns reinforce the need for student-support models that are responsive to heterogeneous life circumstances rather than relying on traditional full-time student assumptions.

Second, the analysis indicates that basic digital access and self-assessed professional digital skills are comparatively well-established across the analysed systems and age groups. High digital readiness is not systematically associated with high perceived labour-market preparedness, suggesting that digital competence functions primarily as an enabling condition for participation in digitally transformed higher education rather than as a sufficient explanation of differences in perceived employability.

Third, and most importantly, the findings demonstrate that digitalisation and flexibility should not be treated as synonymous concepts. Greater online-learning intensity does not automatically produce stronger alignment with student preferences. While higher employment intensity is associated with a stronger preference for online learning, some mature student groups already experience more online provision than they would ideally prefer. Effective flexibility therefore depends on the alignment between institutional delivery models and heterogeneous student needs rather than on maximising digital delivery itself.

From an AI-enabled SRM perspective, these findings suggest that the strategic value of AI may lie less in increasing digital intensity or administrative automation and more in supporting adaptive alignment. AI-enabled SRM may help institutions identify differentiated student profiles, recognise varying time constraints and learning preferences, and provide more responsive communication, advising, and support. However, the present study does not evaluate the effectiveness of such AI-enabled mechanisms and therefore treats them as potential institutional capabilities requiring further empirical validation.

Given that the study relies on aggregated cross-country and country-by-age-group indicators, it does not establish causal or individual-level relationships. Future research should therefore validate the proposed AI-enabled SRM framework using micro-level institutional data and student-level studies, with particular attention to adaptive support, student engagement, perceived employability, transparency, student agency, and the effectiveness of AI-enabled interventions.

Ethics Declaration

This study is based on secondary analysis of aggregated, anonymised, and publicly available data obtained from the EUROSTUDENT 8 database. The research did not involve direct interaction with human participants, primary data collection, or the processing of identifiable personal information. Consequently, formal approval from an institutional ethics committee was not required.

AI Declaration

During the preparation of this paper, AI tools (large language models) were used to support language refinement, academic proofreading, and technical clarification of selected data-processing procedures using Microsoft Excel. These tools were not used for autonomous data analysis, interpretation of findings, or the generation of scientific conclusions.

All conceptual development, empirical interpretation, analytical reasoning, and final conclusions represent the original work and responsibility of the authors.

References

Adner, R. (2017) 'Ecosystem as structure: An actionable construct for strategy', *Journal of Management*, 43(1), pp. 39–58. Available at: https://doi.org/10.1177/0149206316678451.

Benavides, L.M.C., Arias, J.A.T., Serna, M.D.A., Bedoya, J.W.B. and Burgos, D. (2020) 'Digital transformation in higher education institutions: A systematic literature review', *Sensors*, 20(11), p. 3291. Available at: https://doi.org/10.3390/s20113291.

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S.W. and Siemens, G. (2024) 'A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour', *International Journal of Educational Technology in Higher Education*, 21, p. 4. Available at: https://doi.org/10.1186/s41239-023-00436-z.

Buckingham Shum, S.J. and Luckin, R. (2019) 'Learning analytics and AI: Politics, pedagogy and practices', *British Journal of Educational Technology*, 50(6), pp. 2785–2793. Available at: https://doi.org/10.1111/bjet.12880.

Chiu, T.K.F., Xia, Q., Zhou, X., Chai, C.S. and Cheng, M. (2023) 'Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education', *Computers and Education: Artificial Intelligence*, 4, p. 100118. Available at: https://doi.org/10.1016/j.caeai.2022.100118.

Crompton, H. and Burke, D. (2023) 'Artificial intelligence in higher education: The state of the field', *International Journal of Educational Technology in Higher Education*, 20, p. 22. Available at: https://doi.org/10.1186/s41239-023-00392-8.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. and Siemens, G. (2024) 'Impact of AI assistance on student agency', *Computers & Education*, 210, p. 104967. Available at: https://doi.org/10.1016/j.compedu.2023.104967.

Gašević, D., Dawson, S., Rogers, T. and Gasevic, D. (2016) 'Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success', *The Internet and Higher Education*, 28, pp. 68–84. Available at: https://doi.org/10.1016/j.iheduc.2015.10.002.

Gašević, D., Tsai, Y.-S. and Drachsler, H. (2022) 'Learning analytics in higher education – Stakeholders, strategy and scale', *The Internet and Higher Education*, 52, p. 100833. Available at: https://doi.org/10.1016/j.iheduc.2021.100833.

Gašević, D., Siemens, G. and Sadiq, S. (2023) 'Empowering learners for the age of artificial intelligence', *Computers and Education: Artificial Intelligence*, 4, p. 100130. Available at: https://doi.org/10.1016/j.caeai.2023.100130.

Gholami, H., Saman, M.Z.M., Mardani, A., Streimikiene, D., Sharif, S. and Zakuan, N. (2018) 'Proposed analytic framework for student relationship management based on a systematic review of CRM systems literature', *Sustainability*, 10(4), p. 1237. Available at: https://doi.org/10.3390/su10041237.

Gholami, H., Zakuan, N., Saman, M.Z.M., Sharif, S. and Kohar, U.H.A. (2020) 'Conceptualizing and operationalizing the student relationship management strategy: Towards a more sustainable-based platform', *Journal of Cleaner Production*, 244, p. 118707. Available at: https://doi.org/10.1016/j.jclepro.2019.118707.

Grönroos, C. (2019) 'Reforming public services: Does service logic have anything to offer?', *Public Management Review*, 21(5), pp. 775–788. Available at: https://doi.org/10.1080/14719037.2018.1529879.

Gummesson, E. (2008) *Total relationship marketing*. 3rd edn. Oxford: Butterworth-Heinemann.

Hauschildt, K., Vögtle, E.M. and Gwosć, C. (2018) *EUROSTUDENT VI overview and selected findings: Social and economic conditions of student life in Europe*. Bielefeld: W. Bertelsmann Verlag. Available at: https://doi.org/10.3278/104-274w.

Hauschildt, K. (ed.) (2024) *Social and economic conditions of student life in Europe: EUROSTUDENT 8 Synopsis of indicators 2021–2024*. Bielefeld: wbv Publikation. Available at: https://doi.org/10.3278/6001920ew.

Håkansson, H. and Snehota, I. (2006) 'No business is an island: The network concept of business strategy', *Scandinavian Journal of Management*, 22(3), pp. 256–270. Available at: https://doi.org/10.1016/j.scaman.2006.10.005.

Janschitz, G. and Penker, M. (2022) 'How digital are "digital natives" actually? Developing an instrument to measure the degree of digitalisation of university students – the DDS-Index', *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 153(1), pp. 127–159. Available at: https://doi.org/10.1177/07591063211061760.

Mouta, A., Pinto-Llorente, A.M. and Torrecilla-Sánchez, E.M. (2024) 'Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education', *International Journal of Artificial Intelligence in Education*, 34, pp. 1166–1205. Available at: https://doi.org/10.1007/s40593-023-00384-9.

OECD (2019) *Benchmarking higher education system performance*. Paris: OECD Publishing. Available at: https://doi.org/10.1787/be5514d7-en.

Payne, A. and Frow, P. (2017) 'Relationship marketing: Looking backwards towards the future', *Journal of Services Marketing*, 31(1), pp. 11–15. Available at: https://doi.org/10.1108/JSM-11-2016-0380.

Ragnedda, M., Ruiu, M.L. and Addeo, F. (2020) 'Measuring digital capital: An empirical investigation', *New Media & Society*, 22(5), pp. 793–816. Available at: https://doi.org/10.1177/1461444819869604.

Rigo, G.-E., Pedron, C.D., Caldeira, M. and Araújo, C.C.S. (2016) 'CRM adoption in a higher education institution', *Journal of Information Systems and Technology Management*, 13(1), pp. 45–60. Available at: https://doi.org/10.4301/S1807-17752016000100003.

Schirmer, H. (2024) *Digitalisation of teaching, learning, and student life*. EUROSTUDENT 8 Topical Module Report. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).

Selwyn, N. (2019) 'What's the problem with learning analytics?', *Journal of Learning Analytics*, 6(3), pp. 11–19. Available at: https://doi.org/10.18608/jla.2019.63.3.

Siemens, G. (2012) 'Learning analytics: Envisioning a research discipline and a domain of practice', in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York: ACM, pp. 4–8. Available at: https://doi.org/10.1145/2330601.2330605.

UNESCO Institute for Statistics (2015) *International standard classification of education: Fields of education and training 2013 (ISCED-F 2013)*. Montreal: UNESCO Institute for Statistics.

Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F. (2019) 'Systematic review of research on artificial intelligence applications in higher education – where are the educators?', *International Journal of Educational Technology in Higher Education*, 16(1), p. 39. Available at: https://doi.org/10.1186/s41239-019-0171-0.

SMAP-ST: Smart MBTI-Adaptive Public Speaking Tutor

Leonardo Mogianesi[1,2], Stelios Kapetanakis[3] and Giacomo Nalli[4]
[1]University of Modena and Reggio Emilia, Modena, Italy
[2]University of Camerino, Camerino, Italy
[3]Distributed Labs, Distributed Analytics Solutions, 30 Churchill Pl, London, UK
[4]Computer Science, Science and Technology, Middlesex University London, UK

leonardo.mogianesi@unimore.it
stelios@distributedanalytics.co.uk
g.nalli@mdx.ac.uk

Abstract: Up to 80% of graduate students in the UK struggle to speak confidently in front of an audience, whether during class presentations, seminars, assessments or interviews. A major reason for this is the lack of a consistent support during their school education for developing strong public speaking skills whilst most teachers do not feel confident, with the tools they have, to support learners in developing these skills. Consequently, many primary and secondary school students do not have enough opportunities to develop their public speaking skills. Beyond institutional gaps, individual differences such as personality traits can significantly impact a student's confidence and ability to speak in public. For some students, public speaking can trigger overwhelming anxiety or cognitive overload. Students with more introverted or sensitive traits, for example, may feel intense discomfort when asked to perform in public. Personality patterns can create a range of barriers that traditional classroom environments rarely accommodate. These patterns affect more than just academic performance; they also limit opportunities after graduation. Without targeted support, many students may miss out on opportunities simply because there is a lack of effective support for their diverse communication needs. SMAP-ST (Smart MBTI-Adaptive Public Speaking Tutor) is an intelligent tutoring system powered by AI. It is designed to support oral presentation training in an online environment that combines text and voice-based interactions with a virtual tutor. Thanks to interaction with SMAP-ST, students can rehearse presentations on selected topics through guided interaction and receive immediate, personalized feedback on their communication skills. The system's distinctive feature is its MBTI-based personalisation, whereby the tutor adapts the tone and structure of the feedback to suit different personality types. This helps students to identify their own weaknesses during speech, boost their confidence and fluency, and be effective at the same time because it fits with their personality traits. The system is evaluated by analysing collected data during user interaction with the SMAP-ST to identify the effectiveness of the response generated for each personality trait.

Keywords: Intelligent Tutoring System, Personality Traits, Generative AI, Inclusive Education, Distance Education

1. Introduction and Background

Speaking in public has been reported as a significant challenge for many students at university level. Specifically in the UK, studies at two universities showed that up to 80% of students experience social anxiety during oral presentations and speaking to public(Russell et al., 2012). Indeed, many undergraduate and postgraduate learners often struggle to give a speech in front a large audience, whether it is a presentation, pitch or job interview and experience as reported "a sense of fear" (Goberman et al., 2011). A 2024-25 survey from the Department for Education (DfE) shows, indeed, that only 21% of instructors feel "very confident" about their ability and available tools to support their learners with spoken language and listening skills (Department for Education, 2025). This consequently impacts students negatively as they do not have consistent and effective support and opportunities to practice public speaking in front of a general audience. This fear of public speaking is also influenced by other factors such as fear of being judged, lack of confidence, stress and anxiety (Grieve et al., 2021). These factors also have further effects when combined with specific student personality traits. For example, introverts return a higher level of social anxiety than extroverts (Sigarlaki et al., 2023), extroverts generally experience less fear of negative evaluation than introverts (Duffy et al. 2018) while perfectionists could feel negative emotions (depression, stress) when their performance differs from their expectations (Simonet al., 2025). The literature demonstrates that real-time feedback has a positive impact on public speaking training by making the experience more immersive, improving participants' motivation and enhancing the final performance. As public speaking is affected by individual factors such as personality, training could also be improved by being tailored to participants' individual circumstances (Chollet et al., 2022). However, giving effective feedback is not an easy task for the teachers who often struggle to provide the right recommendations to students due to diverse curricula, intense pressure to delivery, high level of admin, etc. As a result learners receive not effective feedback which usually focuses more on the summative assessment rather than constructive recommendations for actual improvement(Bock et al., 2024). The paper aims to bridge this reported gap. We implement the "Smart MBTI-Adaptive Public Speaking Tutor (SMAP-ST)" an AI-powered intelligent

tutoring system that can provide to students real-time and specific feedback based on their speech strengths or challenges and tailored to their personality traits to maximise their learning experience. This paper is structured as follows, Section 2 resides the literature around our area, Section 3 presents our Methodology, whilst Section 4 present our results. Finally, Section 5 concludes by presenting the work against relevant benchmarks and follows with the future work.

2. Literature Review

Research on intelligent tutoring systems (ITS) has long highlighted the value of digital environments that support learning through guided interaction. Specific and corrective feedback has also been recognized as a central component of this process in traditional ITS (Yazdani, 1986). This perspective is much more relevant for oral presentation training, where improvement happens through iterative and long practice. (Heideman et al., 2022). An example of online tutoring systems provide students with recommendations to improve the quality of the java code, giving static feedback based on the missing elements in the code (Nalli et al. 2023). Over the years, the dialogic ITS has become central in the literature thanks to a conversational interaction treated as a key mechanism through which learning can be supported (Graesser et al., 2001). In the dialogic ITS, the exchange between learner and system is not marginal, because questions, responses, clarification, and progressive refinement shape the educational process (Heideman et al., 2022) . However, traditional and Dialogic Intelligent Tutoring systems work well in some specific structured topics, but less well in the so called ill-defined domain (Fournier-Viger et al., 2010) Recent work has extended this line of research toward spoken language improvement through conversational systems. The use of conversational intelligent tutoring systems for oral skill development has shown that spoken production, transcription, and AI-mediated feedback can be integrated into a single workflow (Pérez-Ortiz et al., 2024). A similar direction also emerges in generative AI systems designed for English-speaking training, where the model supports the learner by processing spoken or transcribed output and returning feedback in a form that is immediate and educationally meaningful (Lu et al., 2025). The growing spread of AI models has further strengthened this area of research. The potential of AI in education has been widely recognized, especially in relation to tutoring, feedback generation, and learner support (Huber et al., 2024) (Xing et al., 2025). Nevertheless, the AI models implemented for the speaking training are focused more on pronunciation or phonetic rather than the whole complete performance (Chen et al., 2016). The literature also notes that the use of LLMs in education requires critical revision of the generated output rather than simple reliance on it (Huber et al, 2024). Another relevant aspect concerns adaptation to learner characteristics such as the students' personalities. The relationship between MBTI profiles and student behaviour in educational systems has been explored as a possible factor influencing engagement with digital learning environments (Alseitova et al., 2024). Although personality-based adaptation should not be treated as deterministic, this line of work offers a rationale for examining whether differences in feedback style may shape how learners perceive and use tutor responses. However, there is no LLM public speaking training tool which provides tailored feedback based on MBTI profiles. Overall, SMAP-ST lies at the intersection of intelligent tutoring systems, conversational tutoring, large language models in education, and AI-supported oral skills training. Its main contribution consists in bringing these directions together within a single chat-based environment that also explores MBTI-informed feedback adaptation as a design choice.

3. Methodology

The SMAP-ST intelligent tutoring system consists of a chat-based interaction tool that acts as a "virtual" tutor. It provides formative feedback on students' speech, considering their personality traits in the response to maximise learning support. Specifically, SMAP-ST runs on the shell and requires as input the student speech in MP3 format to initialise the process. The file needs to be uploaded into a dedicated folder so that the system can recognise and transcribe it using a speech-to-text API. Once the transcript is obtained, SMAP-ST starts the processing of the data as shown in Figure 1 which illustrates the system architecture.

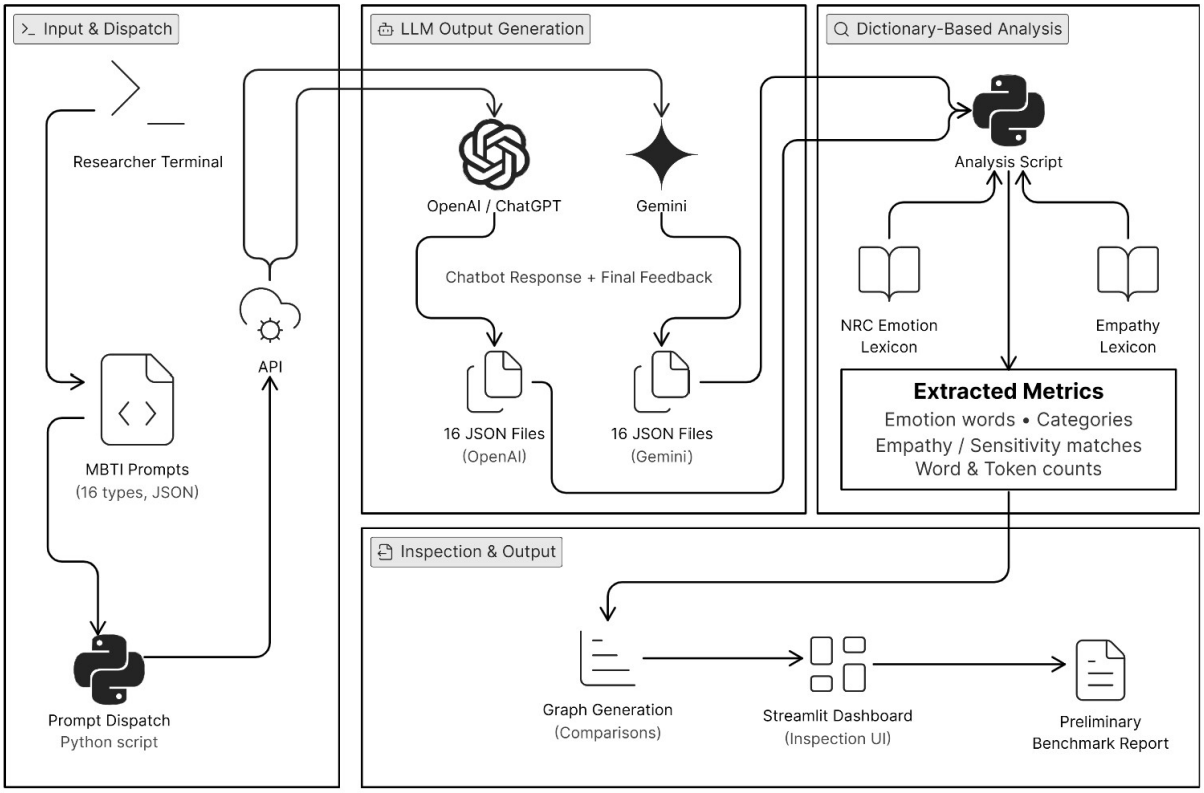


Figure 1: SMAP-ST system architecture and feedback-generation workflow.

The experimental pipeline was built around a Python-based AI agent developed specifically for this study. The agent was designed to interact with the APIs of both OpenAI/ChatGPT and Google/Gemini and to initialize each chat session through a system prompt. The system prompt consisted of sixteen different variants, each corresponding to one MBTI personality type. The prompt used by the system changed according to the MBTI profile selected by the user when running the program. The MBTI personality can be identified by filling a specific test which return four characters that identify one of the 16 personalities(Bharadwaj et al., 2018)(Nalli et al, 2025). For the purposes of this experimental phase, the selection of the MBTI personality was defined through terminal execution. This selection allowed the AI agent of SMAP-ST to generate responses tailored to students' personality by modifying only the specific MBTI system prompt without affecting the rest of the workflow. The first step of the software implementation consisted of creating a JSON collection containing sixteen system prompts, each corresponding to one MBTI type. These prompts were developed using insights from "Differentiation through Personality Types", a book which collects the most effective strategies for instructors to differentiate the delivering of the content to the students based on their personalities dimensions, with the aim to improve their own motivation and learning experience (Kise, 2007) . The AI agent used personality classifications to shape its responses, adjusting the answers according to personality-oriented profiles. In this study, MBTI was used only as a design-oriented personalization framework rather than as a diagnostic psychological instrument. Each MBTI-specific prompt is processed by two large language model (LLMS) providers: OpenAI/ChatGPT 5.4 and Google/Gemini 2.5 Pro. In order to ensure a consistent comparison, the same prompt-set was used for both providers. The same transcript of the speech (input) was also used across all experimental runs, together with the previously mentioned personality-specific system prompt. The transcript simulates a short public speaking task in which the speaker had to present themselves in a professional context. Specifically, the text addressed three guiding questions: the speaker's identity, the role the speaker was aiming for, and the reasons why an employer should hire them. Using the same base input across all MBTI profiles and both LLMs allowed the comparison on the effects of the personality-oriented prompts and the model provider, this design ensured that any observed differences in the generated outputs were attributable to the MBTI-oriented prompts and to the LLM provider, rather than to variations in the input speech content. For each personality type, two textual outputs were generated by each model: a chatbot response and a final feedback message. The first output represented the conversational response of the ITS. This interaction could continue over multiple iterations if additional questions or follow-up requests were submitted to the system by the student throughout the chat. However, for the purposes of this preliminary benchmark, the session was limited to a single question–

answer exchange, followed by the final feedback message mentioned above. All outputs generated by both providers were stored into JSON files. In total, thirty-two JSON files were produced: sixteen files for OpenAI/ChatGPT and sixteen files for Google/Gemini, corresponding to the sixteen MBTI personality types for each LLM. Each JSON file contained the generated chatbot response and final feedback associated with the specific personalityoriented prompt. Once the iteration between the student and ITS was over, a Python analysis script was used to compare the generated JSON files with two lexical resources: the NRC Emotion Lexicon (Mohammad et al., 2013) and the empathy lexicon (Sedoc et al., 2020), both taken from dedicated repositories. The first was used to identify emotion-related vocabulary in the generated feedback, while the second was used to identify empathy-related terms. The script parsed each JSON file, extracted the textual content, and counted the number of matched words in each dictionary. This analysis produced the following outputs: matched emotion words, emotion categories, empathy-related matches, sensitivity-related matches, and token counts. These indicators were calculated for each personality type and for each provider. The extracted match counts, emotion-category frequencies, empathy and sensitivity indicators, and token counts were then used to generate matching comparisons between OpenAI/ChatGPT and Google/Gemini. The resulting graphs reported empathy and sensitivity match percentages, emotion-related counts, and average response length indicators. The interpretation of differences between the two LLM was supported by these visualizations, which allowed the generated feedback to be compared across the sixteen MBTI profiles. Finally, a Streamlit (Streamlit, 2026) inspection interface was used to review the generated JSON files, the matched words, the emotion categories, and the graphical outputs. Ultimately, this methodological process led to the development of a preliminary benchmark report comparing the lexical and emotional features of feedback generated by OpenAI/ChatGPT and Google/Gemini.

4. Results and Discussion

A benchmark was conducted to compare the text generated by OpenAI/ChatGPT and Google/Gemini by matching the AI-generated responses (chatbot response and final feedback) provided by SMAPST, with the selected dictionaries. The analysis focused on empathy, sensitivity, output length, and emotional tone. Empathy and sensitivity were measured though a match-making process between the words included in the responses produced by LLMs and the words included in the related dictionaries (Mohammad et al., 2013). Emotional tone was analysed through a separate emotion dictionary (Sedoc et al., 2020). The benchmark does not aim to prove that one model is pedagogically superior to the other, but to identify which provider produces more effective feedback based on the selected lexical categories and for each MBTI personality. Figure 2 reports the percentage of words matched, obtained by comparing the AI-generated feedback with the empathy and sensitivity dictionaries across the sixteen MBTI personality traits.

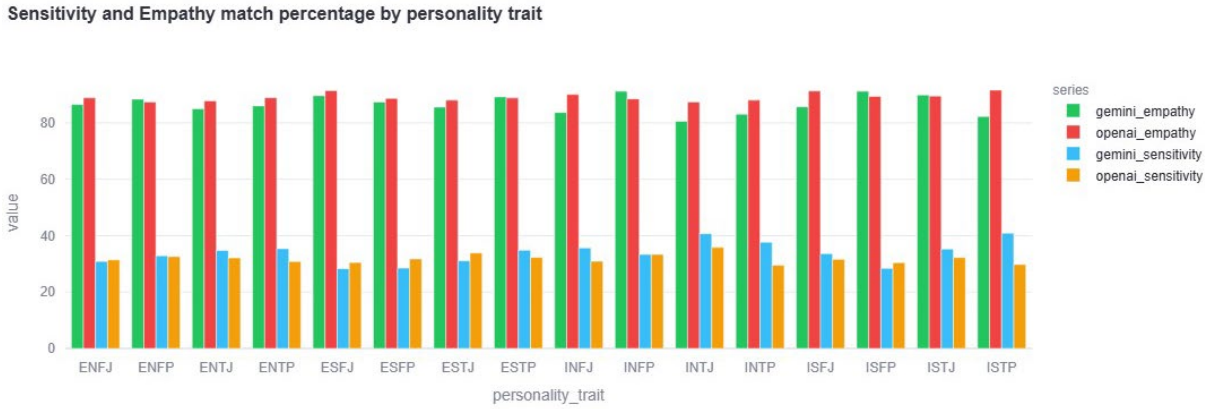


Figure 2: Sensitivity and empathy match percentage by personality.

The percentage-based comparison is important because it reduces the effect of response length and allows a more balanced comparison between providers. Both models show high matching with empathy-related terms across personality traits, while sensitivity-related matches are consistently lower, with OpenAI/ChatGPT reaching an average empathy score of 89.12% and Google/Gemini reaching 86.58%. Sensitivity-related matches were consistently lower for both providers, with OpenAI/ChatGPT obtaining 31.80% and Google/Gemini 33.87% on average. Therefore, OpenAI/ChatGPT showed a slightly higher empathy-related matching rate, while Google/Gemini showed a slightly higher sensitivity-related matching rate. However, the differences between the two providers remained relatively small, corresponding to 2.53% for empathy and 2.07% for sensitivity. This data

indicates that both models generate feedback contained vocabulary associated with supportive communication, but OpenAI/ChatGPT appears more closely aligned with the empathy-related dictionary used in the assessment framework than Google/Gemini.

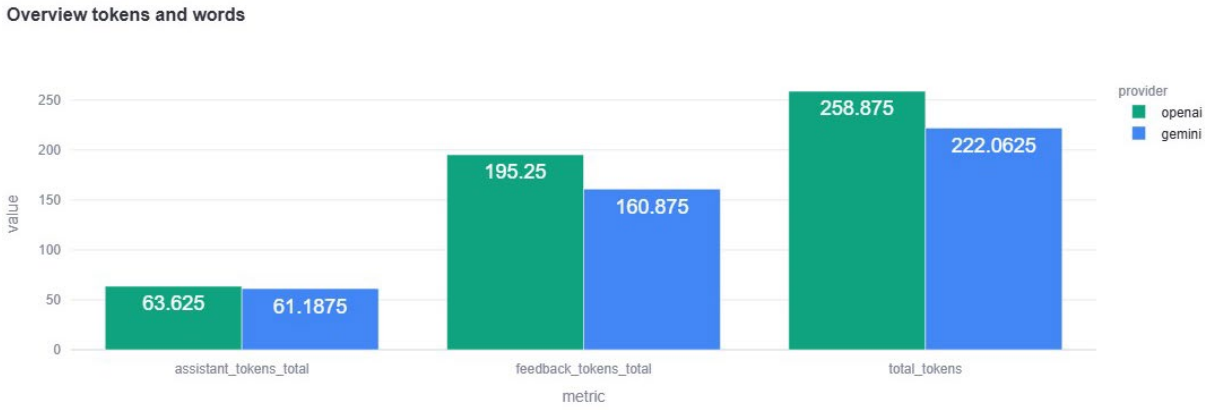


Figure 3: Comparison between OpenAI/ChatGPT and Google/Gemini on average token volumes: assistant tokens, feedback tokens, and total tokens.

Figure 3 provides an overview of the average number of assistant tokens, feedback tokens, and total words produced by each LLM. OpenAI/ChatGPT generated longer feedback responses on average, reaching approximately 194 words and 1,233 characters, while Google/Gemini generated approximately 159 words and 1,033 characters. OpenAI/ChatGPT outputs were longer by about 35 words and 200 characters on average. This difference should be considered when interpreting dictionary-based matches, since longer responses may naturally contain a higher number of terms associated with empathy or sensitivity dictionaries. For this reason, Figure 2 reports percentage-based values rather than only unique word matches, since percentages make the comparison between providers less dependent on the different output lengths.

Table 1: Average emotion counts by LLM provider.

Provider	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
ChatGPT	1.25	8.0	0.75	1.5	6.88	1.81	27.0	1.06	2.25	12.81
Gemini	2.25	8.69	1.38	2.94	6.0	3.44	27.12	1.12	2.69	11.25

Figure 4 shows the average distribution of emotion-related matches obtained by comparing the feedback generated by the two providers with the NRC Emotion Lexicon dictionary. Both models produced feedback mainly matching positive, trust-oriented, and anticipatory terms, with positive terms accounting for approximately 42% for OpenAI/ChatGPT and 41% for Google/Gemini, trust-oriented terms for approximately 20% for OpenAI/ChatGPT and 17% for Google/Gemini, and anticipatory terms for approximately 12% for OpenAI/ChatGPT and 13% for Google/Gemini. Negative emotion categories appear less frequently, with negative-related terms accounting for approximately 3% for OpenAI/ChatGPT and 5% for Google/Gemini, and fear, anger, sadness, and disgust remaining below approximately 5% for both providers. This is relevant for SMAP-ST because public speaking feedback should not only identify weaknesses, but also maintain a supportive and encouraging tone. The similarity between OpenAI/ChatGPT and Google/Gemini in this analysis suggests that both providers generate feedback with a comparable emotional profile, although OpenAI/ChatGPT shows a slightly higher number of trust-related matches. Overall, the benchmark indicates that OpenAI/ChatGPT produced feedback with stronger lexical matching with the predefined empathy dictionary, reaching approximately 89% compared with 87% for Google/Gemini, while Google/Gemini showed a slightly higher sensitivity-related matching rate, with approximately 34% compared with 32% for OpenAI/ChatGPT . Google/Gemini also produced shorter responses on average, with approximately 159 words compared with 194 words for OpenAI/ChatGPT. Based on these results, OpenAI/ChatGPT was selected for the main feedback generation analysis. However, these findings should be interpreted as evidence of lexical and emotion correspondence between AI-generated responses and the dictionaries, rather than direct evidence of pedagogical effectiveness. Dictionary-based matching captures the presence of relevant terms, but it does not fully measure the quality, accuracy, or usefulness of the feedback from the learner's perspective. For this reason,

the benchmark should be considered a preliminary model-selection step, to be complemented by user-based evaluation of perceived usefulness, clarity, confidence improvement, and usability.

5. Conclusion

SMAP-ST is an AI-based tutoring system for public speaking practice. The system personalizes the feedback based on MBTI personality prompts. The software is implemented in Python and it was tested with two LLMs providers: OpenAI/ChatGPT and Google/Gemini. The same input text was used across all MBTI profiles to compare which LLM align better, doing a pattern matching based on two dictionaries that are NRC Emotion Lexicon and the Empathy Lexicon. OpenAI/ChatGPT showed a slightly stronger lexical alignment with the empathy category, reaching approximately 89% compared to 87% for Google/Gemini, while Google/Gemini obtained a slightly higher sensitivity score, with approximately 34% compared to 32% for OpenAI/ChatGPT. Google/Gemini also produced shorter responses, with approximately 159 words on average compared with 194 words for OpenAI/ChatGPT, while maintaining a similar score in terms of emotional responses. The results suggest that MBTI-oriented prompting can influence LLM feedback. For example, in the empathy category, OpenAI/ChatGPT reached high values for ESFJ and ISTP, with approximately 91% and 92%, respectively, while Google/Gemini reached its highest values for INFP and ISFP, both around 91%. In the sensitivity category, Google/Gemini obtained particularly high scores for INTJ and ISTP, both around 41%, whereas OpenAI/ChatGPT showed lower sensitivity scores for the same profiles, with approximately 36% for INTJ and 30% for ISTP. SMAP-ST can be a step towards AI tutoring system that helps people become more confident with public speaking. Although the findings are preliminary due to the lack of pedagogical analysis through student evaluation, the obtained results demonstrate high potential and a promising future outlook. Future work will involve conducting user testing, with the aim to evaluate the clarity and usability of the software, as well as the confidence of the students using the SMAP-ST intelligent tutoring system.

Ethics Declaration

This is not applicable as the study only consists of a stand-alone software implementation, and doesn't involve human participants, personal data or interaction with users.

AI Declaration

The AI hasn't been adopted for the content development. A translator tool was used to refine some sentences in the paper. The OpenAI/ChatGPT and Google/Gemini API the has been linked to the software developed in the paper.

References

Alseitova, A., Oliveira, W., Li, Z., Shi, L., and Hamari, J. (2024). "The Relationship Between Students' Myers-Briggs Type Indicator and Their Behavior within Educational Systems", 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 3-7

Bharadwaj, S., Sridhar, S., Choudhary, R., and Srinath, R. (2018). "Persona traits identification based on Myers-Briggs Type Indicator (MBTI)-a text classification approach", 2018 international conference on advances in computing, communications and informatics (ICACCI), pp. 1076-1082.

Bock, T., Thomm, E., Bauer, J., and Gold, B. (2024). "Fostering student teachers' research-based knowledge of effective feedback", European Journal of Teacher Education, Vol 47, No. 2, pp. 389-407.

Chen, N. F., and Li, H. (2016). "Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning", 2016 Asia-Pacific Signal and Information Processing Association annual summit and conference, pp. 1–7.

Chollet, M., Marsella, S., and Scherer, S. (2022). "Training public speaking with virtual social interactions: Effectiveness of real-time feedback and delayed feedback", Journal on Multimodal User Interfaces, Vol. 16, No. 1, pp. 17-29.

Department for Education (2025). School and College Voice: December 2024. Research and analysis. Available at: https://www.gov.uk/government/ publications/school-and-college-voice-omnibus-surveys-for-2024-to-2025/school-andcollege-voice-december-2024.

Duffy, K. A., Helzer, E. G., Hoyle, R. H., Fukukura Helzer, J., and Chartrand, T. L. (2018). "Pessimistic expectations and poorer experiences: The role of (low) extraversion in anticipated and experienced enjoyment of social interaction", PloS one, Vol. 13, no.7.

Fournier-Viger, P., Nkambou, R., and Nguifo, E. M. (2010). "Building intelligent tutoring systems for ill-defined domains", Advances in intelligent tutoring systems, pp. 81-101.

Goberman, A. M., Hughes, S., and Haydock, T. (2011). “Acoustic characteristics of public speaking: Anxiety and practice effects”, Speech communication, Vol. 53, No.6, pp. 867-876.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., and Harter, D. (2001). “Intelligent tutoring systems with conversational dialogue”, AI magazine, Vol. 22, no.4.

Grieve, R., Woodley, J., Hunt, S. E., and McKay, A. (2021). “Student fears of oral presentations and public speaking in higher education: a qualitative survey”, Journal of Further and Higher Education, Vol.45, no.9, pp. 1281-1293.

Heideman, P. D., and Laury, J. E. (2022). “Ultra-short presentations with immediate in-class public feedback to enhance skill development with low class time and instructor time”, College Teaching, Vol. 70, no. 2, pp. 196-205.

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., and Ninaus, M. (2024). “Leveraging the potential of large language models in education through playful and game-based learning”, Educational Psychology Review, Vol. 36, no.1, p. 25.

Kise, J. A. (2007). “Differentiation through personality types: A framework for instruction, assessment, and classroom management”, Corwin Press.

Lu, C.-T., Chen, Y.-J., Wu, T.-Y., and Lu, Y.-Y. (2026). “An Intelligent English-Speaking Training System Using Generative AI and Speech Recognition, Applied Sciences, Vol. 16, no. 1, p. 189.

Mohammad, Saif M. and Peter D. Turney (2013). “*NRC Emotion Lexicon*. English. National Research Council Canada”, available at : https://nrc-publications.canada.ca/eng/view/ object/?id=0b6a5b58-a656-49d3-ab3e-252050a7a88c.

Nalli, G., Culmone, R., Perali, A., and Amendola, D. (2023). “Online tutoring system for programming courses to improve exam pass rate”, Journal of e-Learning and Knowledge Society, Vol. 19, no. 1, pp. 27-35.

Nalli, G., Kapetanakis, S., and Nguyen, K. A. (2025). “AI Intelligent Tutoring System Tailored to the Students' Personality and Neurodiversity”, European Conference on e-Learning, pp. 297-305.

Pérez-Ortiz, J. A., Esplà-Gomis, M., Sánchez-Cartagena, V. M., Sánchez-Martínez, F., Chernysh, R., Mora-Rodríguez, G., and Berezhnoy, L. (2024). “A conversational intelligent tutoring system for improving english proficiency of non-native speakers via debriefing of online meeting transcriptions”, In Proceedings of the 13th workshop on natural language processing for computer assisted language learning, pp. 187-198.

Russell, G., and Topham, P. (2012). “The impact of social anxiety on student learning and well-being in higher education”, Journal of mental health, Vol. 21, no.4, pp. 375-385.

Sedoc, J., Buechel, S., Nachmany, Y., Buffone, A., and Ungar, L. (2020). “Learning word ratings for empathy and distress from document-level user responses”, In Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 1664-1673.

Sigarlaki, M. A. S., and Setiawati, L. S. (2023). “Social Anxiety Disparity Between Extroverted and Introverted Psychology Students During COVID-19 at University" X" ”, Journal of Health and Behavioral Science, Vol.5, no.3, pp. 313-328.

Simon, P. D., Salanga, M. G. C., and Aruta, J. J. B. R. (2025). “The distinct link of perfectionism with positive and negative mental health outcomes”, Frontiers in psychiatry, Vol.16.

Xing, W., Nixon, N., Crossley, S., Denny, P., Lan, A., Stamper, J., and Yu, Z. (2025). “The use of large language models in education”, International Journal of Artificial Intelligence in Education, Vol.35, no.2, pp. 439-443.

Yazdani, M. (1986). Intelligent tutoring systems: an overview. Expert systems, 3(3), 154-163.

From "AI is Supi" to "Democracy is in Danger": Users' Perceptions About AI

Dagmar Monett
Computer Science Dept., Berlin School of Economics and Law, Germany

dagmar.monett-diaz@hwr-berlin.de

Abstract: Educational institutions worldwide have introduced courses, training, guidelines, and similar documentation which range from allowing unlimited use of artificial intelligence (AI)-based products to their complete ban in specific settings. Tangentially, and with the spread of chatbots and similar software programs for fast producing textual and other types of information, there are also increasing concerns regarding how and what students actually learn or should, as well as whether outsourcing reading comprehension, academic writing, critical thinking, and other cognitive skills, once paramount in educational contexts, are enhanced or atrophied with the help of AI, and how academic integrity might be devalued as a result of the latter. Tangentially, because the harms are often acknowledged but the use of AI is simultaneously promoted, which is known as responsible realism and is a form of ethics washing. This paper presents part of an empirical, exploratory study whose goals include the analysis of users' perceptions about the use of AI and their awareness of possible harms. To gather opinions on these and related topics, a survey was conducted to which a total of 535 university members participated. All responses were scrutinized using qualitative content analysis. This allowed for the extraction of categories that depict the whole spectrum of their stances regarding AI. The results contribute to understanding users' awareness, opinions, concerns, and needs regarding the use of this technology. They also help in uncovering common tropes that might be preventing university members from assessing generative AI harms critically.

Keywords: AI, AI Harms, Education, Critical AI, Responsible Realism

1. Introduction

Four months after ChatGPT's release on November 30, 2022, an internal webinar on AI-supported writing was organised by the university's IT department to which about 40 people, mainly lecturers, participated. The first presenter opened the webinar with the question *"Will you use ChatGPT in your teaching in Summer 2023?"* The answers from the participants were: *Yes*, 32%; *No*, 25%; and *Not sure*, 42%. The contents then included a brief description about ChatGPT (and generative AI), what could be done with it (at the time), the advantages of using it ("for saving time" being one of them), the legal issues if used for grading students' works, as well as the supposed benefits of "the AI" everybody was talking about. It followed a discussion with a Q&A session. I remember my talking about the AI hype, plagiarism, and issues the technology already had, and how even publishing houses and universities were sharing editorials on either banning its use or clarifying that the only authors should be humans and not software programs. I shared several publications on the chat, like e.g. (Brainard, 2023; COPE, 2023; Flanagin et al., 2023; Nature, 2023; Thorp, 2023). The same question was asked at the end. The answers this time, first to my surprise and then to my disappointment, were: *Yes*, 65%; *No*, 20%; and *Not sure*, 15%, an evident shift that could suggest a lack of awareness of the harms and issues of the technology. How about other university members?

This paper presents part of an exploratory empirical study which goals include, among others, the analysis of university members' awareness of the harms of and their perceptions about AI. It follows the methodology for qualitative content analysis (QCA) suggested in (Mayring, 2014; 2025), which first step consists in *defining concrete research questions* (Mayring, 2014, p. 10). The research questions guiding this work are:

> RQ1: To what extent are university members aware of selected issues and current or potential harms associated with generative AI?
>
> RQ2: What positive and negative perceptions of generative AI are expressed by university members, and to what extent do these perceptions reflect recurring tropes identified in the literature?

The second step of QCA consists in *linking the research question(s) to theory* (Mayring, 2014, p. 11). Much has been written regarding users' perceptions about the use of AI. For instance, Miranda (2025) surveyed university members from an Italian university and found that although positive towards AI, participants were sceptical of its utility for younger learners. Unterweger et al. (2026) document students' concerns about data privacy and academic integrity, among others. In Colbran et al.'s (2026) study about chatbots use, students manifested concerns e.g. about academic integrity and the accuracy of results. Erkılıç and Cifci (2025), in contrast, queried ChatGPT about similar concerns and other academic violations, and suggest that knowing about ChatGPT's issues helps to address problems of unethical behaviour.

Little has changed in some contexts despite abundant research exposing the dangers of the uncritical adoption of generative AI, especially in education (Atelier Ecopol, 2025; Birhane, 2021; Crawford, 2021; Guest et al., 2026; Markelius et al., 2024; Monett & Paquet, 2025). Some users find chatbots' writing "abilities" a perfect substitute for doing the tasks themselves. Others, surrounded by the current wave of AI-based applications, might actually conform to what goes with the AI hype, see technical solutionism as unavoidable, neglect critical harms, and take cognitive growth for granted (Monett & Paquet, 2025).

The next sections introduce the methodological approach underneath QCA and how the data was gathered, analysed, presented and evaluated. I.e. the steps three to seven suggested by Mayring: *definition of the research design*, *definition of the sample or material and the sampling strategy*, *methods of data collection and analysis*, *processing and presentation of results*, and *discussion* (Mayring, 2014, p. 11ff).

2. Methods

The methodological approach used in this research is QCA, "a mixed methods approach: assignment of categories to text as qualitative step, working through many text passages and analysis of frequencies of categories as quantitative step" (Mayring, 2014, p. 10). The text material consists of open-ended answers provided by respondents to survey questions. The research design is *explorative*, where the content analysis is performed following an *inductive* category formation procedure (Mayring, 2014, p. 79ff). The selection criterion includes, but is not limited to, the reasons for or against the use of generative AI in the context of education and research. *Systematically*, categories and sub-categories are built depending on those reasons, i.e. by analysing their *semantic connotation*. General comments, suggestions, concerns, and common tropes (or clichés) are grouped according to their *type*. Their occurrences are quantitatively analysed, *exemplarily*.

To gather university members' opinions, a survey was conducted in July 2025. This paper focuses *only* on the results concerning both participants' perceptions and knowledge about the role that generative AI-based programs are playing in 14 concrete forms of harm as documented in the literature. Reality shows how negative aspects and AI harms are underestimated or ignored not only at the individual but also at the institutional level and beyond. Is educating university members about such harms necessary? How can this knowledge influence institutional policies, teaching, and learning?

2.1 Participants

17855 people were invited to participate in the survey: 13952 students (undergraduate and postgraduate, the latter being e.g. Master's students), 3520 adjunct lecturers (those with teaching contracts and external to the university), 288 professors (those in university positions), and 95 researchers (those being e.g. doctoral students and/or in research positions). The invitation was sent via email to four corresponding university mailing lists which were made available by the IT department of the university. The survey was active two weeks, from July 9 to July 24, 2025, and was provided in two languages, German and English.

A total of 535 people (3%) completed the survey. This should be more than enough for an exploratory study in qualitative analysis (Braun & Clarke, 2025). One respondent belongs to a group not explicitly targeted with the survey and is thus discarded. All other responses (i.e. 534) are analysed, i.e. the sample for QCA is not such but the whole set of responses: 325 from undergraduate (60.9%) and 88 postgraduate (16.5%) students, 104 from adjunct lecturers or professors (19,5%), and 17 from researchers including PhD students (3.2%).

2.2 Data Collection

The survey was created using Google Forms. Options to collect personal data there were deactivated. Apart from questions related to the use of generative AI in teaching and learning, two general questions are key to the analysis presented in this paper: a multiple-choice question comprising 14 current AI harms and one open-ended question to collect any other opinion about generative AI (see Table 1).

2.3 Data Analysis

No generative AI-based tools were used in this research project nor in the production of this paper, the main being a *conscientious objection* (Atelier Ecopol, 2025) to generative AI because of the many proven harms of this technology, especially in, but not limited to, education and academia (see e.g. Guest et al., 2026). Furthermore, Mayring (2025) found that summaries generated by ChatGPT are not suited for QCA and are error-prone,

superficial, and do not recognize hidden text content. Other authors do not recommend the use of generative AI for qualitative analysis either (Jowsey et al., 2025; Nguyen & Welch, 2025; Radu & Rocher, 2026).

Table 1: Questions regarding participants' awareness of AI harms and added opinions on generative AI-related topics (In the German version of the questionnaire, those were questions Q16 and Q17, respectively)

[Q34] Are you aware of the possible role the usage of generative AI plays in... *Check all that apply.* ☐ ...increasing CO2 emissions? ☐...increasing water consumption? ☐...increasing energy consumption? ☐...becoming dependent on AI? ☐...replacing jobs? ☐...extending bias and discrimination? ☐...the loss of knowledge quality? ☐...expanding security breaches? ☐...cognitive offloading? ☐...the loss of human agency? ☐...degrading learning abilities? ☐...violating intellectual property rights? ☐...spreading fake news, misinformation and disinformation? ☐...allowing data workers exploitation? ☐ I am not aware of any of the issues mentioned above.
[Q35] Are there any other comments you would like to add regarding the use of generative AI? Please, add them below.

The results of the open-ended question, Q35 from Table 1 and its corresponding one in German, were analysed manually by following the methodological approach introduced above. Tools for interactive qualitative content analysis (like QCAmap (Mayring, 2025)) were dismissed since no consent from respondents was collected in this regard. In addition, tropes regarding generative AI were analysed by drawing on the notion of *responsible realism*, an instance of ethics washing introduced in (Guersenzvaig & Monett, 2026).

2.4 Quality Criteria

Although carried out carefully and based on the long experience (also in writing) of the author in the topics mentioned above, QCA was done by only one person, i.e., no intra/-inter-coder agreement can be calculated. This is not an issue in qualitative analysis, though (Braun & Clarke, 2025). The results presented here should be considered *exploratory*. To partially counter this limitation, and as suggested by (Mayring, 2014, p. 114), part of the textual material and its categorisation was sent to a colleague, who verified the semantic validity and analysis regarding tropes about generative AI.

3. Results and Discussion

All participants (n=534) answered question Q34 (see Table 1): 479 (89.7%) answered the German and 55 (10.3%) the English version of the questionnaire. Figure 1 shows the total number of occurrences for each harm option, and Table 2 shows their distribution depending on the participant group.

Only 6 participants (1,1% from the total) indicated not being aware of any of the harms. 12 lecturers (11,5% of that group) and 1 researcher (5,9% of that group) do not use generative AI; all of them are aware of at least 2, and 7 of them of at least 11 from 14 harms where generative AI might play a role in. One lecturer indicated not having used generative AI, but in a different question indicated using it for generating ideas and suggestions, which is a contradiction and thus is not counted as not using AI. Several harms rank low for all three groups. As the data shows, there are types of harms where even students know much better than lecturers and researchers. Lecturers, as a group and on average, are actually the least aware of the harms, at least as indicated by the answers they provided. Also possible: some participants might be ignoring the harms or do not see them as

critical enough to stop using AI. This might be a reason for the uncritical adoption of generative AI in their teaching, a serious problem according to (Guest et al., 2026).

96 respondents (18%) answered the open-ended question Q35, 65 of them being students and the rest, lecturers and researchers. 80 comments were collected from the version in German of the questionnaire, and 16 from the one in English. The length of the comments varies from 1 to 327 words.

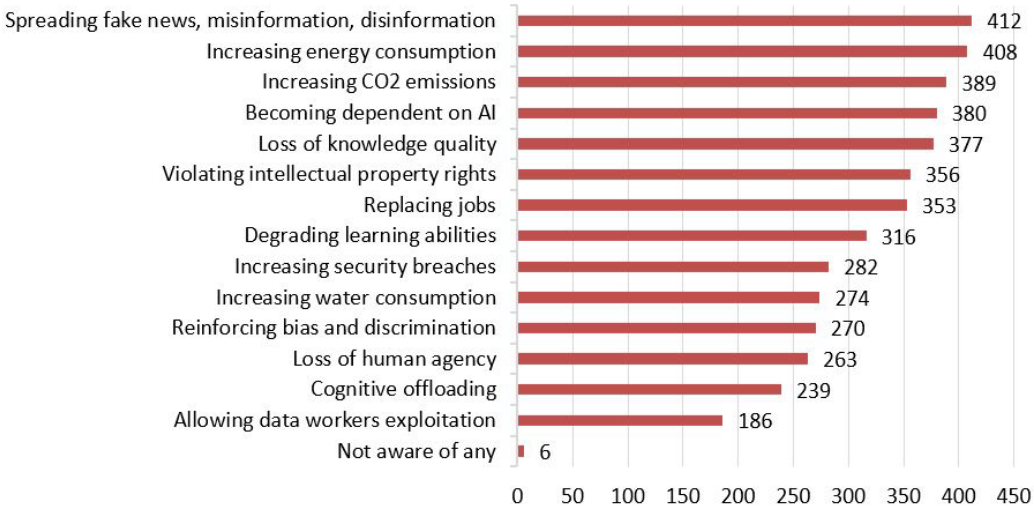


Figure 1: Participants' acknowledged awareness: Possible role the usage of generative AI plays in different harms

Table 2: Participants' acknowledged awareness: Distribution and percentage relative to the participant group (emphasis: top 3 highest and top 3 lowest scores)

Awareness of…	Students	Lecturers	Researchers
Spreading fake news, misinformation, disinformation	325 (**78,7%**)	73 (70,2%)	14 (**82,4%**)
Increasing energy consumption	312 (**75,5%**)	82 (**78,8%**)	14 (**82,4%**)
Increasing CO2 emissions	294 (71,2%)	81 (**77,9%**)	14 (**82,4%**)
Becoming dependent on AI	306 (**74,1%**)	61 (58,7%)	13 (76,5%)
Loss of knowledge quality	295 (71,4%)	72 (69,2%)	10 (58,8%)
Violating intellectual property rights	266 (64,4%)	78 (**75,0%**)	12 (70,6%)
Replacing jobs	285 (69,0%)	58 (55,8%)	10 (58,8%)
Degrading learning abilities	238 (57,6%)	68 (65,4%)	10 (58,8%)
Increasing security breaches	216 (52,3%)	55 (52,9%)	11 (64,7%)
Increasing water consumption	217 (52,5%)	44 (**42,3%**)	13 (76,5%)
Reinforcing bias and discrimination	188 (**45,5%**)	69 (66,3%)	13 (76,5%)
Loss of human agency	219 (53,0%)	36 (**34,6%**)	8 (**47,1%**)
Cognitive offloading	187 (**45,3%**)	45 (43,3%)	7 (**41,2%**)
Allowing data workers exploitation	143 (**34,6%**)	35 (**33,7%**)	8 (**47,1%**)
Not aware of any	4 (1,0%)(*)	1 (1,0%)(**)	1 (5,9%)

*: 3 other students also indicated being aware of almost all the harms, which is a contradiction.

**: 1 other researcher also indicated being aware of almost all the harms, which is a contradiction.

One by one, each comment was analysed, thoroughly, and its semantic connotation categorised. The following broad sets were constructed along this process, inductively (concrete examples are showed in Tables 3 to 5):

- General comments (positive or negative) regarding generative AI (e.g. related to the respondent's experience)

- Concerns regarding the use of generative AI (i.e. concrete concerns the respondent has)
- Needs or suggestions mentioned by the respondent (some respondents mention them explicitly)

The categories and sub-categories, created also inductively from the topics addressed in the comments, were:

- Education (assessment, grading, curriculum, teaching, training, learning, learning methods, cognitive abilities, cognitive skills, human abilities)
- Ethics (awareness, ethical aspects, good practice, requirement)
- AI use (AI in teaching, AI in learning, AI in research, software coding)
- Issues (AI limitations, challenges, harms, risks, laziness, lock-in, stolen data, plagiarism, replacement, jobs)
- Governance (guideline, regulation, banning)
- Pushback (refusal, opposition, denial)
- Tropes (AI hype, fear of missing out (FOMO), ignorance about functioning, inevitability, category error, false generalization, uncritical adoption, responsible realism, industry push, efficiency, time)
- Survey (content-related)

These categories did not exist *a priori* but were built *systematically* after analysing each of the 96 comments that were provided. From those categories, the most frequently mentioned topics were *AI use* (e.g. whether AI should be used or not, as a support tool or not, good in research or not, good for learning or not, etc.) with more than 60 mentions in total; *harms* (e.g. to individuals, study programme, institution, society, etc.) with more than 50 mentions, and *tropes* (see below) with more than 45 mentions.

Table 3: Examples of general aspects regarding generative AI, grouped by semantic connotation

Positive Aspects:	Negative or Problematic Aspects:
Increasing importance of AI-related skills	AI is promoted everywhere, especially at the university
Use of AI is important in education	AI is in all products without being asked; it interrupts work
AI can be good in research	There is an overuse of AI
AI is a very important tool for learning	AI is not only ChatGPT
AI is helpful for coding	Regulation does not prevent use
AI should be freely available for use in writing	There is no motivation not to use AI
AI for spell and grammar check is Ok	AI as a tool is good for those who know how to check it
AI is useful	Not all students use AI but there is pressure from the university
AI is supi	Lecturers using AI is not well seen
	Lesser suitable assessment was needed because of AI
	AI is pushed by industry
Both positive and negative aspects in the same comment:	Other general aspects:
Generative AI has both pros and cons	Developing own's cognitive abilities and experiences is important for learning and in education
Use in research can be positive but also problematic	Own's achievements and learning shouldn't be replaced (by tools' output)
AI makes the work easier but not always better	Students who use AI have advantage but learn less
	There are law suits against AI

Table 4: Examples of participants' concerns regarding the use of generative AI, grouped by semantic connotation

Related to Teaching and Learning:	Involving Duty, Responsibility, or Ethical Behaviour:
There is no possibility to know about AI usage	There is no space for opposition (to the use of AI)
The learning effect when using AI is smaller or poorer	There is disregard for harms
The use of AI foments laziness	Licenses are purchased despite the harms or without consulting users
Students replace thinking completely with AI	

Related to Teaching and Learning:	Involving Duty, Responsibility, or Ethical Behaviour:
Some students use it to produce whole writing works	Adoption without consulting; technology lock-in
Group work is affected (when some use AI and others not)	Students are not aware of or clarified about their responsibility for the results
Students' use of AI should be disclosed but who will check anyways?	There is little awareness about responsibility structures
Lecturers often fail to explain content well; thus, AI should be used	There is no discussion about ethical considerations and negative impact
There is a possible skill degradation	Banning means disadvantage
There might be unfairness when grading	There is AI hype
Use of AI should not be supported at the university	Some of the effects are controversial or have not been clearly proven
Grade should not be affected if students use AI because lecturers cannot explain content well	Used wrong, one could learn little, become dependent, and be replaceable
Plagiarism is an issue	Replacing own's actions completely is problematic
Accusing someone of plagiarism based on the use of AI is unjustified	
Excessive and especially undeclared use of AI lowers the academic standards of the program	
Related to the Technology:	**Other General Concerns:**
The output of generative AI is often false or not existent	Not using AI is seen as a “blocker for progress”
Use should be careful because of errors or non-existent sources	Not using AI is a career disadvantage
Output should not be trusted; it should be verified first	Tasks that once made fun might disappear
Output does not meet scientific standards	AI exacerbates harms
That using it saves time is questionable	There are many harms and risks
Many people have no knowledge about how AI functions	Interests of a few over interests of humanity
AI has limitations like dependency on training (stolen) data	There is a potential for threats to life
Online content is flooded with AI-generated content	

Table 5: Needs or suggestions, grouped by semantic connotation

Related to Teaching and Learning:	Involving Duty, Responsibility, or Ethical Behaviour:
Session on how to use AI efficiently in research and teaching	Use of AI must be disclosed
Faculty must be trained to deal with ChatGPT	Use of AI should be allowed
How to let students use AI	Use of AI should be regulated and controlled by experts
How to test/verify the use of AI when grading	AI should not be banned, but regulated
How to use AI to personalize asynchronous learning	AI output must be assessed critically
New assessment forms are needed	Responsibility should not be outsourced to faculty
A basic understanding of software coding is needed	AI tools should be used with moderation
Teacher training on the good and also on the risks of AI	Pros and cons of the use of AI should be evaluated
More information about pros and cons for learning	Discussion on ethical implications is needed
Learning activities must be adapted, cannot be the same	Faculty should be provided with clear guidelines
AI should be integrated into the learning process	Guidelines not only for students, but also for faculty
AI should be freely available for use in writing	
Related to the Technology:	**Other General Suggestions:**
More training of AI on human research methods	Need to determine which human tasks make sense or not anymore
Prompt engineering as lecture	Differentiating between the human and the artificial will be essential
AI compliance with data protection regulations when doing research is needed	

The following examples of comments contain instances of the category *AI use*:

> *"AI as support is good as long as it is not used to write complete works."* (Lecturer; translated from German)

> *"I hate AI, but sometimes by deadlines and academic stress means that I have no choice but to save time in some instances by finding literature and clarifying concepts. However, when I have done this in the past ChatGTP had produced information that is not factually correct, so I don't actually know if it saves me time."* (Postgraduate student)

The following examples of comments contain instances of the category *harms*:

> *"There is a huge disregard for the harms of generative AI."* (Lecturer)

> *"The welfare of humanity must take precedence over the interests of a few. Unfortunately, that is not the case at present. AI is exacerbating the problems caused by digitalization.* ***Democracy is in danger."*** (Undergraduate student; translated from German)

Reflecting on the users' perceptions about AI, and as manifested in the comments to the open-ended question Q35, it is evident that there is much to be learnt and done regarding generative AI. The whole spectrum of opinions—from positive to negative, for or against AI, from demonstrably true to demonstrably false, from ethical to unethical—is present across all participant groups. Some perspectives, however, ignore the limits of AI (Guest et al., 2026; Selwyn, 2024).

Drawing on the concept of *responsible realism*, a form of ethics washing and "stance in which Higher Education institutions [or individuals] acknowledge the harms and ethical problems of commercial AI, yet still promote its use, shifting the burden of mitigation onto individual end users by calling to 'responsible use'" (Guersenzvaig & Monett, 2026, p. 1), each comment was analysed to detect if it contained traces of this kind of narrative or similar ones. As expected, responsible realism was found in the respondents' comments, at least in 13 of them (14,8%). Several examples are presented in Table 6, the last column being their type.

Table 6: Examples of common tropes (fallacies, clichés) about generative AI, together with their type

Trope, Fallacy, Cliché	Type
Assuming that chatting with a software program is an important skill	Overvalue of skill
No AI use is a blocker for progress	Techno-solutionism
AI cannot be stopped	Inevitability
AI is useful and not problematic when it functions right and someone is responsible	Ethical washing
The tool is good if the output is analysed critically	Critical washing
AI will dramatically transform higher education and research	AI hype, inevitability
"Prompt instructor" will be a job opportunity	AI hype
Banning generative AI makes no sense	Devaluing of regulation
Education should be transformed to the "new reality" of AI	Determinism
AI is increasingly relevant for companies; thus, it should be taught at the university	Inevitability, category error
Those using AI will replace those not using it	FOMO
Potential or competitiveness not possible if students do not use AI or learn how to use it	FOMO
AI is just a tool	Disregard for biases
AI is a partner	Anthropomorphism
Everybody uses AI; those who say they don't are lying	False generalisation
Teach how to use it properly and issues will be mitigated	Wishful thinking, critical washing
Delegating cognitive tasks to AI is good: more time available and fun when learning	Ignorance about how learning occurs
Harms depend on how AI is used	Ignorance about how the technology works, ethical washing

There was no category formation in this part of the analysis. Those are well-known tropes, easy to detect based on the author's experience and publications, and as defined in the research literature (see e.g. Guersenzvaig & Monett, 2026; Guest et al., 2026; Morozov, 2013; Postman, 1992).

Based on determinism and inevitability, and recreating cases of techno-solutionism, anthropomorphism, AI hype, and even ignorance about both how generative AI actually works and the socio-technical conditions that determine it, responsible realism provides a cover for unethical practices, unfortunately. Those users' perspectives might risk undermining the goals of education (Postman, 1995). As Selwyn (2024) reminds us:

> "The future of AI and education is *not* a foregone conclusion that we simply need to adapt to. Instead, the incursion of AI into education is definitely something that can be resisted and reimagined."

4. Conclusion

The findings presented above provide a varied perspective regarding the research questions. Yet, they should be taken as a casual snapshot of much more complex reasons, dynamics, and conditions regarding the use of generative AI. People, tools, and users' perceptions are constantly changing. Refusal, rejection, and organisation are evolving too. The findings were considered when designing two micro-credentials courses on Critical AI that will be offered to university members starting 2027. Interested readers may reach out to request more details about them.

Other university members were excluded from participating, but staff and other administrative members are also generative AI users and might influence its adoption. Another limitation is the possible bias some respondents might have had when answering the open-ended question: at least seven of them were visibly upset with the question about their awareness of the harms. It is known, for example, that the rejection of these technologies and criticism of the *status quo* is often accompanied by erasure or even retaliation (Täuber, 2026). To partially prevent this, anonymous participation was considered. But if there is no open culture for discussion about the most acute AI issues affecting education and research, we cannot expect sincerity and a civilised discussion, needless to mention principled action.

Acknowledgements

Especial thanks go to Ariel Guersenzvaig for the fruitful discussions on responsible realism, to Maja Günther and Titus Reiche for their support in preparing the questionnaire, and to the anonymous participants to the survey.

Ethics Declaration

No personal data was collected. The questionnaire was reviewed by the central division "Research Support Services" of the university and officially approved on July 7th, 2025.

AI Declaration

No generative AI was used. Translations were done by the author.

References

Atelier Ecopol (2025, December 4) "Faced with generative AI, conscientious objection" (Original title in French: Face à l'IA générative, l'objection de conscience), *ATelier d'ÉCOlogie POLitique*, https://informatique.up8.edu/actu/2025-2026.html#2025-12-04_17-30

Birhane, A. (2021) "The Impossibility of Automating Ambiguity", *Artificial Life*, 27(1):44–61.

Brainard, J. (2023) "As scientists explore AI-written text, journals hammer out policies", *Science*, *379*(6634):740–741.

Braun, V., and Clarke, V. (2025) "Reporting guidelines for qualitative research: a values-based approach", *Qualitative Research in Psychology*, *22*(2):399–438.

Colbran, S., Jha, M., and Schiavone, C. (2026) "Understanding student perspectives on generative AI chatbots: a human-centred mixed-methods study in higher education", *International Journal of Educational Technology in Higher Education*, *23*, 28.

COPE (2023) "Authorship and AI tools", *COPE Council*, Committee on Publication Ethics, https://doi.org/10.24318/cCVRZBms

Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press.

Erkılıç, E., and Cifci, I. (2025) “Ethical Issues and Violations in Using Chatbots in Academic Writing and Publishing: The Answers from ChatGPT”, *Journal of Multidisciplinary Academic Tourism*, *10*(1):111–120.

Flanagin, A., Bibbins-Domingo, K., Berkwits, M. and Christiansen, S.L. (2023) “Nonhuman ‘Authors’ and Implications for the Integrity of Scientific Publication and Medical Knowledge”, *Journal of the American Medical Association*, *329*(8):637–639.

Guest, O., Suárez, M., Müller, B.C.N. and van Meerkerk, E. et al. (2026) “Against the Uncritical Adoption of 'AI' Technologies in Academia”, *Digital Culture & Education*, *16*(2):85–118.

Jowsey, T., Braun, V., Clarke, V., Lupton, D. and Fine, M. (2025) “We Reject the Use of Generative Artificial Intelligence for Reflexive Qualitative Research”, *Qualitative Inquiry*, https://doi.org/10.1177/10778004251401851

Markelius, A., Wright, C., Kuiper, J., Delille, N. and Kuo, Y.-T. (2024) “The mechanisms of AI hype and its planetary and social costs”, *AI Ethics*, *4*:727–742.

Mayring, P. (2014) *Qualitative Content Analysis: Theoretical foundation, basic procedures and software solution*, Social Science Open Access Repository, https://www.ssoar.info/ssoar/handle/document/39517

Mayring, P. (2025) “Qualitative Content Analysis With ChatGPT: Pitfalls, Rough Approximations and Gross Errors. A Field Report”, *Forum Qualitative Social Research*, *26*(1).

Miranda, S. (2025) “Artificial Intelligence in Education: An Exploratory Survey to Gather the Perceptions of Teachers, Students, and Educators Around the University of Salerno”, *Education Sciences*, *15*(8):975.

Monett, D. and Paquet, G. (2025) “Against the commodification of education — if harms then not AI”, *Journal of Open, Distance, and Digital Education*, *2*(1):1–24.

Morozov, E. (2013) *To save everything, click here: The folly of technological solutionism*, PublicAffairs.

Nature (2023) “Tools such as ChatGPT threaten transparent science; here are our ground rules for their use”, *Nature*, *613*:612.

Nguyen, D.C. and Welch, C. (2025) “Generative Artificial Intelligence in Qualitative Data Analysis: Analyzing—Or Just Chatting?” *Organizational Research Methods*, *29*(1):3–39.

Postman, N. (1992) *Technopoly: The surrender of culture to technology*, Vintage Books.

Postman, N. (1995) *The end of education: Redefining the new value of school*, Vintage Books.

Radu, R. and Rocher, L. (2026) “Attributing and situating knowledge cannot be left to language models”, *Nature Machine Intelligence*, *8*:295.

Selwyn, N. (2024) “On the Limits of Artificial Intelligence (AI) in Education”, *Nordisk tidsskrift for pedagogikk og kritikk*, *10*(1):3–14.

Täuber, S. (2026) “Erasing critical scholarship: How attacks on academic freedom undermine democracy”, *Organization*, 1–19.

Thorp, H.H. (2023) “ChatGPT is fun, but not an author”, *Science*, *379*(6630):313.

Unterweger, M., Hörmann, C., Kuka, L., Sabitzer, B. (2026) “The AI generation at school: A large-scale survey on student practices, perceptions, and worries”, *Social Sciences & Humanities Open*, *13*, 102821.

AI and Academic Integrity: Student Perceptions of Plagiarism and Ethical Use at the University of Fort Hare

Bongo Mqukuse and Nathunathi Mvunge
University of Fort Hare, East London, South Africa

bmqukuse@ufh.ac.za
nmvunge@ufh.ac.za

Abstract: The rapid adoption of Artificial Intelligence (AI) in higher education is transforming how students engage with learning, research and academic writing, while raising concerns about plagiarism, authorship, originality and academic integrity. This study explored undergraduate students' perceptions of plagiarism and the ethical use of generative AI at the University of Fort Hare. A mixed-methods, interpretive design was used, drawing on pre-workshop responses from 347 students and post-workshop responses from 127 students. Quantitative items were analysed descriptively, while open-ended responses were analysed thematically using Braun and Clarke's approach. The findings show that students entered the workshop with substantial prior exposure to AI but uneven understanding of ethical use. Among post-workshop respondents, 96.9% rated their understanding of responsible AI use as good or excellent, while 86.6% agreed or strongly agreed that they understood when AI use could become plagiarism or academic misconduct. The study highlights the need for continuous AI literacy development, clear institutional guidance and assessment practices that promote responsible AI use while preserving student authorship and independent thinking.

Keywords: Artificial Intelligence; Academic Integrity; Plagiarism; Ethical Use; Student Perceptions; Higher Education; Mixed-Methods Research; Thematic Analysis.

1. Introduction and Background

Generative Artificial Intelligence (AI) is reshaping higher education by influencing teaching, learning, assessment and academic writing. Tools such as ChatGPT, Gemini and Copilot support idea generation, summarisation, feedback and writing, but also raise concerns about authorship, originality and academic integrity (Kasneci et al., 2023; Tlili et al., 2023). Generative AI complicates conventional understandings of plagiarism because students can produce fluent text without directly copying an identifiable source. This creates uncertainty about acceptable assistance, disclosure and authorship (Cotton, Cotton and Shipway, 2024; Chan, 2025). These issues are particularly relevant to the University of Fort Hare (UFH), where students increasingly encounter generative AI in academic work. Context-specific evidence is needed to understand how students distinguish legitimate learning support from inappropriate assistance and what guidance they require. Against this background, this study explores undergraduate students' perceptions of plagiarism and the ethical use of generative AI at UFH. It assesses students' use and understanding of AI before and after an AI literacy workshop and examines their perceptions of responsible use, plagiarism, authorship and institutional support. The study contributes evidence from a historically disadvantaged South African university and offers recommendations for AI literacy, institutional guidance and assessment practices that support responsible AI use while protecting academic integrity and independent thinking.

2. Literature Review

2.1 Artificial Intelligence and Generative AI in Higher Education

Higher education's teaching and learning methods are being drastically changed by artificial intelligence (AI), especially generative AI tools like ChatGPT. According to Kasneci et al. (2023), these technologies allow students to produce ideas, summarize knowledge, obtain personalized explanations, and assist with academic writing. Additionally, personalized learning, greater accessibility, and more effective academic support are all made possible by generative AI (Tlili et al., 2023). Concerns about student reliance, authorship, academic integrity, and the accuracy of AI-generated data have been highlighted by its increasing use. Cotton, Cotton, and Shipway (2024) draw attention to the difficulties generative AI poses for traditional methods of academic misconduct evaluation. Universities must therefore strike a balance between the educational opportunities provided by AI and responsible integration into teaching, learning, and evaluation as well as explicit ethical guidelines (Perkins et al., 2024).

2.2 AI as a Learning and Academic Writing Support Tool

More university students are using generative AI as a tool to help with learning and academic writing. Tools like ChatGPT can help students brainstorm, organise arguments, summarise research, explain complex ideas, and make their writing clearer (Kasneci *et al*., 2023). These AI tools can also give quick, personalised feedback, helping students work more independently (Tlili et al., 2023). Still, relying too much on AI-generated content can make it harder for students to think critically, write on their own, and build subject knowledge. Using AI in academic writing also brings up questions about who is the real author, how original the work is, and how much AI help should be mentioned (Cotton, Cotton and Shipway, 2024). Due to this, students need clearer guidance and AI literacy skills so they can use generative AI to support their learning without risking academic integrity.

2.3 Academic Integrity, Plagiarism, Originality and Authorship

Generative AI has complicated traditional understandings of academic integrity, plagiarism, originality and authorship in higher education. While AI can support learning, students may struggle to determine when AI assistance becomes academic misconduct or compromises the originality of their work (Cotton, Cotton and Shipway, 2024; Gruenhagen *et al*., 2024). Chan (2025) demonstrates that students' understandings of "AI-giarism" vary, highlighting uncertainty around acceptable AI use. These concerns extend to authorship and disclosure when AI contributes substantially to academic work (Lund *et al*., 2025). Consequently, clearer institutional expectations regarding transparency, acknowledgement and ethical AI use are increasingly important (Perkins, 2023; Perkins *et al*., 2024).

2.3.1 Student Perceptions of AI-Assisted Academic Work

Emerging South African research suggests that university students generally recognise generative AI as a valuable tool for learning, academic support and improving efficiency, while remaining concerned about its ethical implications. Megbowon (2025), studying students at a South African university, found that perceptions of generative AI encompassed its effects on learning as well as concerns about ethical use. Similarly, research involving engineering students at Stellenbosch University highlights the growing integration of generative AI alongside concerns about ethics and academic integrity (Jooste, Wolff and Joubert, 2025). More recently, Pramjeeth (2026) found mixed levels of confidence and ethical engagement among 1,866 students across five South African private higher education institutions, including uncertainty about integrity and disclosure of AI use. These findings demonstrate the need for clearer institutional guidance, ethical AI literacy and further context-specific research within South African universities.

2.4 Ethical Use, Disclosure and Responsible AI

Ethical use of generative AI in higher education requires transparency, accountability and responsible engagement with AI-generated content. Students should critically evaluate AI outputs and disclose AI assistance where required, particularly when it contributes to assessed academic work. Research shows that uncertainty remains around disclosure practices and acceptable levels of AI use, including in the South African higher education context (Pramjeeth, 2026). Responsible AI use therefore extends beyond avoiding plagiarism to include verifying AI-generated information, maintaining student authorship and acknowledging AI assistance appropriately (Bittle and El-Gayar, 2025). Clear institutional guidance and AI literacy are consequently important for helping students distinguish between legitimate academic support and unethical AI use.

2.5 Institutional Policies, AI Literacy and Assessment

The rapid adoption of generative AI requires universities to develop clear institutional policies that distinguish acceptable AI-assisted learning from academic misconduct. Research shows that unclear or inconsistent policies can create uncertainty for both students and lecturers regarding permissible AI use in assessment (Gruenhagen *et al*., 2024; Moorhouse *et al*., 2025). AI literacy is therefore essential to enable students to critically evaluate AI-generated information and use these technologies responsibly. In the South African context, institutional approaches remain uneven, with differences in how universities regulate and communicate acceptable AI use (Jacobs-Basadien and Pather, 2026). Assessment practices should consequently provide explicit guidance on permitted levels of AI involvement and align AI use with intended learning outcomes (Perkins *et al*., 2024).

2.6 South African Higher Education Context and Research Gap

Generative AI is increasingly influencing teaching, learning and assessment across South African higher education, while institutional responses remain uneven. Tarisayi (2025) found gaps in South African university plagiarism policies, particularly in addressing AI-generated content and emerging forms of authorship. More recently, Jacobs-Basadien and Pather (2026) found that only 46% of South African universities examined had institution-wide generative AI guidelines, with considerable variation in how ethical principles were addressed. Student evidence also indicates uncertainty about ethical AI use and disclosure practices in South African higher education (Pramjeeth, 2026). Despite this emerging scholarship, limited research has examined how students at historically disadvantaged public universities understand plagiarism, authorship and ethical AI use. This study addresses this gap by investigating undergraduate students' perceptions at the University of Fort Hare, thereby providing context-specific evidence that may inform institutional policy, AI literacy and academic integrity practices.

3. Normative Theory of Academic Integrity and its Application

The Normative Theory of Academic Integrity provides an ethical foundation for understanding how students make decisions about acceptable and unacceptable academic practices. Academic integrity is conceptualised through six fundamental values: honesty, trust, fairness, respect, responsibility and courage (International Centre for Academic Integrity [ICAI], 2021). These principles are particularly relevant in the context of generative Artificial Intelligence (AI), where students must make decisions about plagiarism, authorship, originality, acknowledgement and disclosure of AI-generated content. The six values can be applied to this study as follows:

- Honesty refers to students being truthful and transparent about their use of generative AI in academic work (ICAI, 2021).
- Trust concerns maintaining confidence in the authenticity of students' academic work and critically evaluating information produced by AI.
- Fairness relates to ensuring that AI use does not provide an inappropriate advantage or undermine the requirements of an assessment.
- Respect involves recognising intellectual contributions, acknowledging sources and adhering to institutional expectations regarding AI use.
- Responsibility requires students to remain accountable for submitted work, including verifying, evaluating and appropriately using AI-generated information.
- Courage involves making ethical academic decisions even where inappropriate AI use may provide an easier means of completing academic tasks (ICAI, 2021).

The framework is applicable to this study because it provides a theoretical lens through which University of Fort Hare students' perceptions of plagiarism, originality, authorship, disclosure and ethical AI use can be interpreted. The framework will assist in examining whether students perceive particular forms of AI assistance as consistent with or contrary to established principles of academic integrity. However, these six values will guide the interpretation of the findings rather than serve as predetermined themes, as themes will be generated from participants' responses through thematic analysis.

The Artificial Intelligence Assessment Scale (AIAS) developed by Perkins *et al*. (2024) can further complement the Normative Theory of Academic Integrity by providing a practical framework for understanding different levels of acceptable AI involvement in assessment. The AIAS was developed to provide greater clarity and transparency concerning the appropriate integration of generative AI into educational assessment (Perkins *et al*., 2024). Together, the two frameworks allow the study to examine both the ethical principles underlying students' perceptions and the practical boundaries between acceptable and inappropriate AI-assisted academic work.

4. Methodology

4.1 Research Design

This study adopted a mixed-methods research design within an interpretive orientation, guided by the Normative Theory of Academic Integrity and its values of honesty, trust, fairness, respect, responsibility and courage (ICAI, 2021). Data were collected using structured pre- and post-workshop questionnaires administered to undergraduate students who participated in an AI literacy workshop at the University of Fort Hare. The

instruments combined closed-ended questions examining AI use, academic integrity and confidence with open-ended questions on students' experiences and perceptions of plagiarism and ethical AI-assisted work. Quantitative data were analysed descriptively, while qualitative responses were analysed thematically using Braun and Clarke (2006; 2021). The framework was subsequently used as an interpretive lens to relate emerging themes to responsible AI use.

4.2 Sampling

The study employed a purposive sampling strategy that included undergraduate students who participated in the AI literacy workshop and were therefore able to reflect on AI-assisted academic work. The pre-workshop questionnaire generated 347 responses and the post-workshop questionnaire generated 127 responses. The two datasets were analysed at group level because the questionnaires did not contain identifiers that would permit individual pre- and post-workshop responses to be matched. Purposive sampling was appropriate because participants had direct experience relevant to the phenomenon under investigation (Campbell et al., 2020).

4.3 Data Collection

Data were collected through structured pre- and post-workshop questionnaires containing closed-ended and open-ended questions. The pre-workshop instrument examined prior AI use, frequency and purposes of use, and students' initial understanding of academic integrity. The post-workshop instrument assessed perceived understanding of responsible AI use, recognition of plagiarism and academic misconduct, confidence in ethical AI use, and recommendations for institutional support. Open-ended questions enabled students to explain what they had learned and how they intended to use AI in future academic work. Participation was voluntary and responses were collected without direct personal identifiers.

4.4 Data Analysis

A mixed-methods approach to data analysis was employed, integrating descriptive statistics with thematic analysis. Quantitative responses from the closed-ended questionnaire items were analysed using frequencies and percentages to identify patterns in students' AI use, understanding of academic integrity, perceptions of plagiarism and confidence in using generative AI responsibly. The descriptive findings provided an overview of students' perceptions before and after the AI literacy workshop.

Qualitative data from the open-ended responses were analysed using Braun and Clarke's (2006; 2021) thematic analysis. Students' narratives were examined for recurring phrases, meaning units, sentences and keywords, which were coded and grouped into broader categories and themes. The analysis followed six phases: familiarisation with the data, generating initial codes, developing potential themes, reviewing themes, defining and naming themes, and producing the final report (Braun and Clarke, 2006).

The thematic analysis was primarily inductive, allowing themes to emerge from the students' responses rather than imposing predetermined categories. This process generated themes relating to responsible and ethical AI use, plagiarism and academic misconduct, AI as a learning support tool, independent thinking, verification of AI-generated information, and the need for institutional guidance and continued AI literacy support.

Following the development of the themes, the Normative Theory of Academic Integrity was used as an interpretive lens to understand how students' perceptions and experiences related to the principles of honesty, trust, fairness, respect, responsibility and courage (ICAI, 2021). The framework therefore did not determine the initial coding or themes but provided an ethical lens through which the qualitative findings were interpreted within the context of AI-assisted academic work.

5. Results

In line with the Normative Theory of Academic Integrity, descriptive statistics and thematic coding were used to evaluate students' responses before and after the AI literacy workshop. The analysis drew on pre-workshop responses from 347 students and post-workshop responses from 127 students. The framework provided an interpretive lens for understanding students' perceptions and experiences of generative AI in relation to the fundamental values of honesty, trust, fairness, respect, responsibility and courage (ICAI, 2021). The themes emerging from the data demonstrate how students understood AI as an academic support tool, negotiated the

boundaries between acceptable AI use and academic misconduct, and reflected on their responsibility for maintaining academic integrity.

5.1 AI as an Academic Learning Support Tool

The pre-workshop responses established students' existing experience with generative AI and provided a basis for understanding their perceptions of AI-assisted academic work. Of the 347 students who completed the pre-workshop questionnaire, 300 (86.5%) indicated that they had previously used AI tools such as ChatGPT, Gemini or Copilot for academic purposes. The most frequently reported use was understanding academic concepts (256 responses), followed by summarising information (142), brainstorming ideas (123), editing grammar and language (74) and writing assignments (36).

Qualitative responses similarly revealed that students viewed AI as a mechanism for obtaining explanations, simplifying difficult concepts, generating ideas and supporting academic tasks. These findings indicate that students entered the workshop with substantial practical exposure to generative AI. However, their prior use did not necessarily demonstrate an equivalent understanding of the ethical boundaries associated with AI-assisted academic work. The findings therefore established the importance of examining not simply whether students use AI, but how they understand their responsibility when incorporating AI into academic learning.

5.2 Responsible and Ethical AI Use

The post-workshop findings indicated high levels of perceived understanding of responsible and ethical AI use. Of the 127 post-workshop respondents, 61.4% rated their understanding as excellent and 35.4% as good. Combined, 96.9% therefore rated their understanding as either good or excellent

Qualitative feedback reinforced these findings, with recurring phrases referring to using AI "responsibly", "ethically", "correctly" and as an "assistant". One student stated, *"AI is not for doing my whole assignment, but it assist me to do it,"* while another indicated, *"We should use AI more responsible."* These narratives reflect an emerging understanding of **honesty and responsibility**, two principles of the Normative Theory of Academic Integrity (ICAI, 2021). Students appeared to recognise that they remain responsible for their academic work even when generative AI is used to support the learning process.

5.3 Plagiarism, Academic Misconduct and Authorship

Students' understanding of plagiarism and academic misconduct represented another prominent theme. Post-workshop findings showed that 52.0% agreed and 34.6% strongly agreed that they understood when the use of AI could become plagiarism or academic misconduct. Combined, 86.6% of respondents therefore expressed a positive understanding of the ethical boundaries surrounding AI-assisted academic work.

Students' narratives further demonstrated this awareness. One participant stated that they had learnt *"To not copy and paste from AI,"* while another indicated that *"Using AI irresponsible lead to plagiarism."* These statements demonstrate an emerging distinction between using AI as academic assistance and presenting AI-generated material as one's own work. Within the Normative Theory of Academic Integrity, these responses particularly reflect **honesty, respect and responsibility**, as students recognised the importance of maintaining their own intellectual contribution and taking responsibility for submitted academic work (ICAI, 2021).

5.4 Independent Thinking and Responsible Use

The students' narratives also demonstrated an understanding that generative AI should support rather than replace independent thinking. Students repeatedly referred to AI as a "guide", "assistant", "support" or form of "help". One student stated, *"You can use AI to guide you but not to think for you,"* while another explained, *"I should use Ai as an assistant tool not to relia on it."*

These responses suggest that students recognised the importance of maintaining intellectual ownership when using generative AI. This finding particularly reflects the academic integrity principle of **responsibility**, since students remain accountable for demonstrating their own knowledge, understanding and academic skills. It also reflects **honesty**, as responsible AI use requires students to ensure that submitted work represents their own intellectual contribution. The findings therefore indicate that students were beginning to conceptualise AI as a supplementary learning resource rather than a replacement for their own academic thinking.

5.5 Trust and Critical Evaluation of AI-Generated Information

The post-workshop responses also revealed an increased awareness of the limitations and reliability of generative AI. Students recognised that AI-generated information may be inaccurate and therefore requires verification before being incorporated into academic work. One student explained, *"I learnt that AI can bring up wrong things… and it can bring non existing information."* Another participant stated, *"I learned that AI has hallucinations hence it produces inaccurate references."*

These narratives demonstrate the importance of **trust and responsibility** within AI-assisted academic work. Trust does not imply accepting AI-generated information without question; rather, students demonstrated an understanding that they have a responsibility to critically evaluate and verify information. This was further illustrated by one student who emphasised *"double-checking AI outputs for accuracy and learning how to prompt it correctly for my coursework."* The findings therefore suggest that responsible AI use extends beyond avoiding plagiarism to include critical evaluation, verification of sources and accountability for the accuracy of academic work.

5.6 Confidence in Responsible AI Use

Students' confidence in applying their understanding of responsible AI use was also examined. Following the workshop, 41.7% of respondents indicated that they were moderately confident in using AI responsibly while maintaining academic integrity, while 34.6% reported being very confident. A further 19.7% were slightly confident, while only 3.9% indicated that they were not confident. Combined, 76.3% of respondents therefore described themselves as moderately or very confident.

These findings suggest that most post-workshop respondents perceived themselves as having the capacity to apply responsible AI practices to their academic work. However, the proportion of students who remained only slightly confident indicates that the development of responsible AI practices requires continued support. Within the Normative Theory of Academic Integrity, this finding is particularly relevant to **responsibility and courage**, as ethical academic practice requires students not only to understand appropriate behaviour but also to apply these principles when completing academic tasks independently.

5.7 Institutional Guidance and Continued AI Literacy Support

The need for continued institutional guidance emerged strongly from students' recommendations. Students frequently requested additional workshops, tutorials, practical demonstrations and clearer guidance on responsible AI use. One participant recommended *"More workshops to teach about AI,"* while another requested *"Clear university guidelines on acceptable AI usage in assignments."* Another suggested that the University should *"Post short videos on Blackboard to help alleviate the unethical use of AI by students."*

These findings demonstrate that students regarded responsible AI use as both an individual and institutional responsibility. Their requests for consistent guidelines are particularly relevant to **fairness,** since students require clear and equitable expectations concerning acceptable and unacceptable AI use in academic assessment. Students also requested support from lecturers, tutors, the Teaching and Learning Centre and the library, as well as guidance on plagiarism detection, credible academic sources and verification of AI-generated information. The findings therefore indicate that responsible AI integration requires an institution-wide approach combining academic integrity principles, AI literacy, clear assessment expectations and ongoing student support.

Overall, the findings demonstrate that students entered the workshop with substantial experience of generative AI but varying levels of understanding of its ethical implications. The post-workshop responses reflected stronger awareness of responsible AI use, plagiarism, independent thinking and the need to verify AI-generated information. When interpreted through the Normative Theory of Academic Integrity, the findings particularly demonstrate the relevance of **honesty, trust, fairness, respect and responsibility** to students' engagement with generative AI. Although **courage** forms part of the framework, it was less explicitly evident in students' narratives and was therefore not treated as a dominant empirical finding. This ensures that the theoretical interpretation remains grounded in the themes that emerged from the data rather than imposing all six values on participants' responses.

6. Discussion

The findings indicate that the AI literacy workshop provided a context for students to reflect on responsible generative AI use through the Normative Theory of Academic Integrity (ICAI, 2021). Although students had substantial prior AI experience, extensive use did not necessarily correspond with clear understanding of ethical boundaries. Their emphasis on avoiding direct copying, verifying outputs and using AI as an assistant reflects honesty, trust and responsibility. This supports research showing that students value generative AI while remaining uncertain about plagiarism, authorship and acceptable assistance (Cotton, Cotton and Shipway, 2024; Gruenhagen et al., 2024; Chan, 2025).

Students' calls for clearer guidelines and further training also reflect fairness and respect and show that responsible AI use cannot depend on individual judgement alone. Universities should provide transparent expectations, continuous AI literacy and assessment guidance (Kasneci et al., 2023; Perkins et al., 2024). This is particularly relevant in South Africa, where institutional responses remain uneven (Tarisayi, 2025). Overall, the findings show that ethical AI engagement requires student accountability, critical evaluation and institutional support.

7. Conclusion

The study concludes that generative AI is an important academic support tool, but extensive use does not necessarily ensure understanding of ethical boundaries. Post-workshop respondents demonstrated high awareness of responsible AI use, plagiarism, verification and independent thinking. Interpreted through the Normative Theory of Academic Integrity, the findings particularly emphasise honesty, trust, fairness, respect and responsibility. The study contributes context-specific evidence from a historically disadvantaged South African university and highlights the importance of combining AI literacy with clear institutional guidance.

8. Recommendations

The University of Fort Hare should provide continuous AI literacy and academic integrity training, supported by clear guidelines on acceptable AI use, disclosure and assessment. Lecturers should communicate AI expectations for individual tasks and encourage students to verify AI-generated information and retain intellectual ownership. Practical resources should be available through Blackboard and other institutional platforms, while policies and assessment practices should be reviewed as generative AI evolves.

Ethics Declaration

The author confirms that this study was conducted in accordance with relevant ethical standards and institutional guidelines for research involving human participants. Ethical clearance was obtained from the Inter-Faculty Human Research Ethics Committee: IFHREC-26/05-0439. All participants were fully informed about the purpose and procedures of the study, and their voluntary participation was secured through written informed consent. The privacy, confidentiality, and anonymity of participants have been strictly maintained, and all data were handled in compliance with applicable data protection regulations

AI Declaration

Artificial intelligence tools, specifically ChatGPT (OpenAI), were used during the preparation of this paper to support language refinement, clarity of expression and the organisation of ideas. The authors critically reviewed and verified all AI-assisted content and retained full responsibility for the analysis, interpretation, academic arguments and final manuscript. The use of AI was undertaken in accordance with applicable ethical and institutional guidelines.

References

Bittle, K. and El-Gayar, O. (2025) 'Generative AI and academic integrity in higher education: A systematic review and research agenda', Information, 16(4), 296. doi:10.3390/info16040296.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', Qualitative Research in Psychology, 3(2), pp. 77-101. doi:10.1191/1478088706qp063oa.

Braun, V. and Clarke, V. (2021) Thematic Analysis: A Practical Guide. London: SAGE.

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D. and Walker, K. (2020) 'Purposive sampling: Complex or simple? Research case examples', Journal of Research in Nursing, 25(8), pp. 652-661. doi:10.1177/1744987120927206.

Chan, C.K.Y. (2025) 'Students' perceptions of "AI-giarism": Investigating changes in understandings of academic misconduct', Education and Information Technologies, 30, pp. 8087-8112. doi:10.1007/s10639-024-13151-7.

Cotton, D.R.E., Cotton, P.A. and Shipway, J.R. (2024) 'Chatting and cheating: Ensuring academic integrity in the era of ChatGPT', Innovations in Education and Teaching International, 61(2), pp. 228-239. doi:10.1080/14703297.2023.2190148.

Gruenhagen, J.H., Sinclair, P.M., Carroll, J.A., Baker, P.R.A., Wilson, A. and Demant, D. (2024) 'The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments', Computers and Education: Artificial Intelligence, 7, 100273. doi:10.1016/j.caeai.2024.100273.

International Center for Academic Integrity (ICAI) (2021) The Fundamental Values of Academic Integrity. 3rd edn. Clemson, SC: International Center for Academic Integrity.

Jacobs-Basadien, M. and Pather, S. (2026) 'Assessing AI policy and guideline progress within the South African public higher education system: Insights from documented policies', Universal Access in the Information Society, 25, Article 71. doi:10.1007/s10209-026-01340-9.

Jooste, J.L., Wolff, K.E. and Joubert, J.G. (2025) 'Engineering students' perceptions and use of generative artificial intelligence', Computer Applications in Engineering Education, 33, e70064. doi:10.1002/cae.70064.

Kasneci, E. et al. (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', Learning and Individual Differences, 103, 102274. doi:10.1016/j.lindif.2023.102274.

Lund, B., Mannuru, N.R., Teel, Z.A., Lee, T.H., Ortega, N.J., Simmons, S. and Ward, E. (2025) 'Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in higher education', AI in Education, 1(1), 2. doi:10.3390/aieduc1010002.

Megbowon, F.K. (2025) 'Generative AI in higher education: Perception and attitude of students from a university in South Africa', Studies in Learning and Teaching, 6(2), pp. 542-555. doi:10.46627/silet.v6i2.641.

Moorhouse, B.L. et al. (2025) 'Generative AI in higher education: A global perspective of institutional adoption policies and guidelines', Computers and Education: Artificial Intelligence, 8, 100348. doi:10.1016/j.caeai.2024.100348.

Perkins, M. (2023) 'Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond', Journal of University Teaching & Learning Practice, 20(2). doi:10.53761/1.20.02.07.

Perkins, M., Furze, L., Roe, J. and MacVaugh, J. (2024) 'The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment', Journal of University Teaching & Learning Practice, 21(6). doi:10.53761/q3azde36.

Pramjeeth, S. (2026) 'Ethics in the age of generative AI: Student perceptions, use and transparency in South African private higher education', Journal of Education, 103, pp. 42-61. doi:10.17159/2520-9868/i103a03.

Tarisayi, K.S. (2025) 'Lustre and shadows: Unveiling the gaps in South African University plagiarism policies amidst the emergence of AI-generated content', AI and Ethics, 5, pp. 245-251. doi:10.1007/s43681-023-00333-1.

Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R. and Agyemang, B. (2023) 'What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education', Smart Learning Environments, 10, 15. doi:10.1186/s40561-023-00237-x.

From Cognitive Overload to Individualized Learning: AI Tutoring in Electrical Engineering

Jens Müller, Annalena Meier, Matthias Heinz, Kilian Göller and Alexander Alfred Zyla
TUD Dresden University of Technology, Dresden, Germany

jens.mueller1@tu-dresden.de

Abstract: First-year students in electrical engineering often encounter significant challenges in their undergraduate programs, particularly due to heterogeneous prior knowledge in mathematics and physics, as well as difficulties in managing their own learning processes. These issues frequently result in cognitive overload, poor exam performance, and high dropout rates. Traditional teaching methods, especially in large cohorts, struggle to address these challenges effectively, as excessive tutorial time is spent on basic calculation steps rather than fostering deeper, subject-specific understanding. To address this, a domain-specific AI tutor was piloted during the winter semester of 2025/2026 in the first-semester course "Fundamentals of Electrical Engineering". The AI, based on the Google Gemini 2.5 Pro language model, was enhanced with Retrieval-Augmented Generation (RAG) to integrate tailored teaching materials, such as annotated lecture transcripts and detailed solution manuals. Configured as a Socratic tutor, the AI provided individualized, asynchronous support by guiding students through tutorial preparation, verifying their calculation steps, and assisting with step-by-step troubleshooting. This tool was tested with a seminar group of 18 students who got access to the AI tutor via the platform you.com. The pilot project included continuous data collection throughout the semester. Weekly online reflection forms captured students' usage patterns, perceived task difficulty, and feedback on the AI's usefulness or limitations. These data were complemented by two comprehensive surveys conducted at midterm and end of semester. The study presents the technological configuration and didactic applications of the AI tutor, analyzing its impact across different student performance levels. The results indicate high acceptance of the AI tutor, particularly for verifying calculations and resolving short-term learning obstacles. However, the analysis also revealed limitations, especially in handling subject-specific visual and topological contexts. The findings underscore the importance of fostering critical AI literacy among students and highlight the need for future multimodal AI models to better support engineering education. In addition, the use of AI in undergraduate studies serves to build competencies within the context of a potential core curriculum.

Keywords: AI Tutor, Blended Learning, Fundamentals of Electrical Engineering, Artificial Intelligence, Learning Environment, E-Learning, Higher Education

1. Introduction

Reducing student dropout rates in engineering is a key challenge for higher education institutions. In Europe and North America, approximately 40 % to 50 % of students drop out of their engineering courses during the first year, highlighting that most problems arise in the critical initial phase of university education (Sultana, Khan and Abbas, 2017). University dropout is the result of multiple complex causes, including personal, family, sociocultural, and economic factors, which may interact with one another. The transition from secondary school to university studies is a sensitive phase in the academic career, characterized by cognitive, emotional, and socio-economic stress factors (Casanova *et al.*, 2021).

Several key factors have been identified as contributing to this development, including highly heterogeneous prior knowledge of mathematics and physics among incoming students, and general difficulties with independent learning organization and time management. However, the primary cause of early academic failure is being overwhelmed by complex subject-specific content, arising from a substantial discrepancy between the university's curricular requirements and high school graduates' actual initial prerequisites (Tayebi, Gómez and Delgado, 2021). This continuous academic overload leads to poor exam results at the end of the first semester, evokes frustration, and frequently ends in the final abandonment of the degree program. This situation is further intensified in modern education by concurrent financial cuts in the higher education sector. These budgetary restrictions create a significantly more challenging institutional starting position, as the available funds for sufficient teaching staff, student tutors, and modern technical equipment for students are increasingly reduced (Hartmann, 2025; John, 2026).

To effectively counter this structural deficit, novel technological approaches for supporting students during this critical phase are gaining importance (Fierro Saltos *et al.*, 2025). This paper presents and empirically evaluates a course-specific AI tutor designed to provide asynchronous, personalized learning support. Within the framework of this study, concrete research questions are addressed: The study investigates the specific ways in which students utilize the provided AI tutor throughout their first academic semester. Building upon this, the analysis focuses on the detailed description of specific usage scenarios over the entire semester and differentiating interaction behaviors within distinct student performance groups. Finally, the paper addresses the question of

what technical limitations and systemic boundaries occur during the practical application of the AI tutor in the students' daily learning routines.

2. Related Work

The theoretical foundation of this work is based on the Cognitive Load Theory (CLT), which states that the human working memory has a strictly limited cognitive capacity (Sweller, 1988). In the field of electrical engineering, the subject's inherent complexity results in a very high intrinsic cognitive load. The primary educational objective is therefore to reduce extraneous cognitive load while maximizing germane cognitive load. However, particularly during the initial phases of university studies, students are confronted with high amounts of both intrinsic loads (inherent difficulty of the task or material itself) and extraneous loads (e.g. organization of study process) simultaneously. This continuous state of cognitive overload is one of the main reasons of university dropout (Naylor, Baik and Arkoudis, 2018; Tight, 2020).

A significant challenge for learners is the transition from procedural knowledge – involving the algorithmic execution of mathematical rules – to conceptual knowledge, which requires recognizing and applying abstract mathematical relationships. This transition becomes especially difficult when complex content is covered too quickly or is significantly beyond the students' prior knowledge level (Koudsia and Kirchner, 2024). Deficits in secondary school education and a lack of routine in performing basic mathematical calculations often mean that exercise sessions and seminars remain on a purely procedural level. Elementary calculation steps, such as solving equations, computing derivatives, or integrating functions, must be carried out in detail. As a compensation strategy, students often try to learn standard problems through pattern matching and overlearning, without truly comprehending the underlying theoretical principles. This ultimately leads to a severe lack of deep conceptual understanding (Kim and Pak, 2002). Furthermore, traditional teacher-centered instruction methods are not scalable due to the high heterogeneity of the student body. In such learning environments, low-performing students often fail at basic algebraic tasks, while high-performing students remain chronically unchallenged.

To address these challenges, the use of generative AI has been introduced in higher education. The overall goal is to promote sustainable and equitable education, as AI enables highly personalized learning environments and facilitates a shift from teacher-centered to student-centered instruction. Current AI systems in education generally fall into three main categories: intelligent tutoring systems, adaptive learning platforms, and AI-driven dialogue systems or chatbots (Asgar, 2025). Their implementation has shown potential for performance improvements, particularly within STEM disciplines (Kestin *et al.*, 2025) and they are most effective when used to supplement human teachers rather than replace them outright (Batsaikhan and Correia, 2024).

AI-driven dialogue systems based on large language models (LLMs) have been used to provide flexible, dialogue-based learning support. (Lin, Huang and Lu, 2023). This is achieved by combining standard language models, such as GPT, Claude, or Gemini, with a domain-specific knowledge base. Specifically, retrieval-augmented generation (RAG) is employed in educational contexts to prevent hallucinations and to seamlessly integrate subject-specific teaching materials (Lewis *et al.*, 2021; Reicher *et al.*, 2025). To prevent AI from merely providing unreflective explanations, it can be configured as a Socratic tutor. This didactic approach emphasizes helping students to help themselves rather than simply transmitting factual knowledge (Santos-Guevara *et al.*, 2026). This approach enables self-paced learning in a psychologically safe environment where students can ask questions without fearing embarrassment due to a lack of knowledge.

Although it has been shown that AI-based learning tools can improve the knowledge, motivation, and self-efficacy of higher education students, their practical impact on higher-order cognitive processes and professional skill acquisition varies considerably and inconsistently across different disciplines (Luo *et al.*, 2025). Furthermore, they often perform poorly as competent pedagogical tutors due to a lack of standardized, learning-science-based evaluation frameworks and specific mistake remediation skills (Maurya *et al.*, 2025). A significant risk associated with this technology is "cognitive offloading", where a high dependence on AI leads to a decline in critical thinking and a reduced ability to solve problems independently (Qadir, 2023; Farrokhnia *et al.*, 2024; Krupp *et al.*, 2024).

Remaining technological limitations are particularly evident in the field of multimodality, where the integrated processing of voice, video, images and text is required. Moreover, AI systems still struggle with visual spatial "understanding" and the systematic analysis of complex scenarios. In the field of electrical engineering in particular, interpreting and understanding complex circuits remains an unsolved challenge (Skelic *et al.*, 2025).

3. Materials and Methods

The research design and methodology are based on a study conducted at the TUD, Dresden University of Technology, during the winter semester of 2025/26. During this period, an “AI tutor” was provided to 18 first-semester students enrolled in the course “Fundamentals of Electrical Engineering”. The didactic structure of the lecture follows an inverted classroom format: Students independently acquire the course content through self-study using interactive instructional videos, literature and digital tests. This is followed by an in-person session in the auditorium, to further develop conceptual understanding and discuss example problems. Each teaching unit concludes with the completion of four or five exercises, followed by a discussion of the solution approaches in a small-group seminar of approximately 20 students.

The course curriculum is structured to build foundational knowledge of electrical engineering progressively. Starting with basic electrical concepts and quantities (charge, current, voltage, power and energy), it then moves on to Kirchhoff's circuit laws and the characterization of resistive two-poles. Based on these principles, the syllabus then moves on to introduce elementary network analysis, covering topics such as voltage and current dividers, the superposition theorem and two-pole network theory. Subsequently, the course explores power dissipation and maximum power transfer. In the final phase, the curriculum advances to more complex circuit elements and systematic solving techniques, addressing dependent sources alongside nodal voltage analysis and mesh current analysis.

Google Gemini 2.5 Pro (Google LLC, 2026) was utilized and implemented as a RAG system via the you.com platform (SuSea Inc, 2026) for the technical realization of the interactive chat platform. A key feature of the system is the integration of subject-specific documents that are not regularly accessible to students. These include complete transcripts of the instructional videos as well as comprehensive model solutions for the exercises. The AI tutor was methodologically configured as a Socratic tutor, focusing on comprehension-enhancing questions rather than simply providing the complete solution. However, students could obtain the complete solution to a task by explicitly asking for it. To deepen the interaction, it was technically possible for students to upload copies of their own handwritten calculations in the form of photos or screenshots. Students were encouraged to use this feature to have their individual problem-solving approaches verified by the AI.

In line with the United Nations Sustainable Development Goals (SDGs) (United Nations, 2026) and the strategic objectives of both the TUD and the Faculty of Electrical Engineering and Information Technology, the specific foundational goals have been established for an AI tutor in electrical engineering. A key requirement is the utilization of a secure network host for the deployed AI models. Consequently, the approach incorporates the you.com pilot project at TUD, which ensures the privacy-compliant use of AI and LLM systems as a secure alternative to freely accessible models from commercial providers. Furthermore, the framework requires a secure, inclusive, and accessible design, coupled with the continuous evaluation and testing of the implemented AI systems. To guarantee reliable support, designated contact people possessing the necessary AI and technical expertise are continuously available to all users.

The data collection was designed as a longitudinal analysis, using weekly online reflection questionnaires to continuously monitor usage patterns and perceived task difficulty. These weekly instruments served for self-assessment and contained specific quantitative and qualitative items. Subjective difficulty was evaluated using the question “How easy was it for you to complete the tasks without the support of an AI tutor?” measured on a five-point scale ranging from “very easy” (1) to “very difficult” (5). Concrete usage behavior was documented via the prompt “Please select how you utilized the AI tutor for the task.” Multiple-choice options included: non-usage, asking for an initial approach, asking for hints on the next steps, verifying one's own calculation, checking the final result, searching specifically for errors in one's own calculation, and requesting an explanation of the entire calculation path. This was supplemented by a free-text field in which students were asked to briefly describe what they found helpful or unhelpful during their usage.

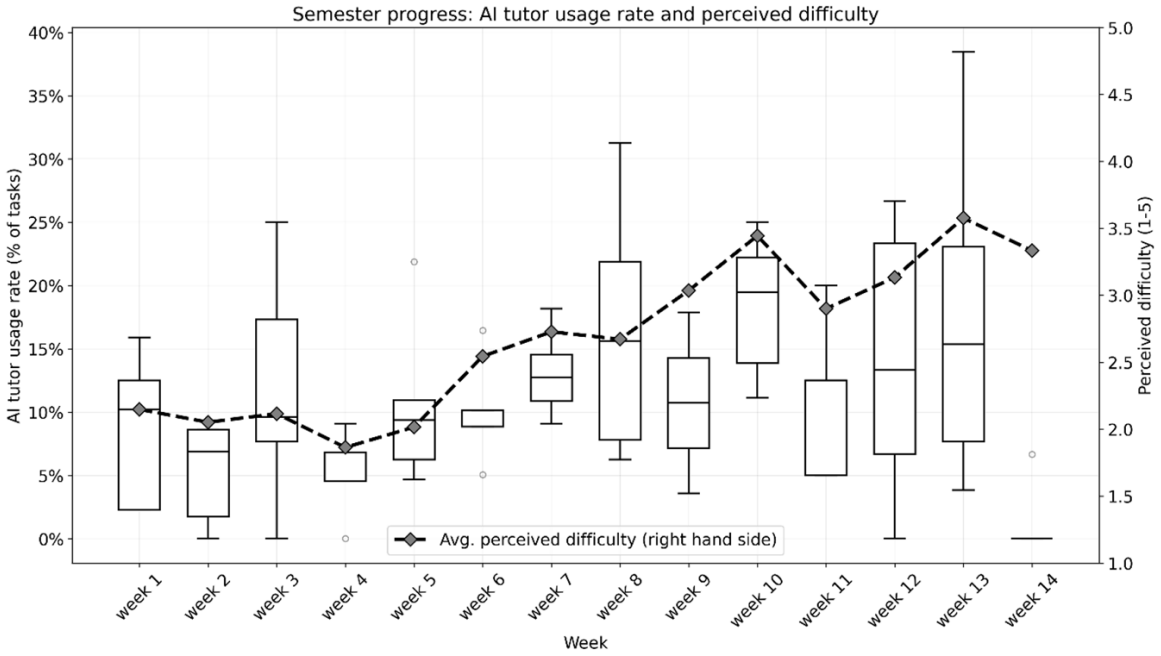


Figure 1: The usage rate of the AI tutor (the percentage of exercise tasks for which the AI tutor has been requested by all students) and the perceived average task difficulty over the 14-week semester.

In addition to the weekly data collection, the evaluation design included two cross-sectional surveys. A mid-term evaluation focusing on the general usefulness of the AI tutor and identifying immediate technical limitations was conducted in the middle of the semester. The study concluded with a final survey at the end of the semester, before the examinations. This served to evaluate the tool's effectiveness and prompted critical self-reflection from the students regarding their current and future use of AI in the context of their university studies.

4. Results and Discussion

Data on the students' interactions with the AI tutor has been collected over the 14-week winter semester. Initially, participation in the weekly feedback forms was high, with all 18 students responding in the first week. However, this number gradually decreased, with only five students responding by week 14. During the semester, the average perceived difficulty of the exercises increased steadily, rising from 2.15 in week 1 to 3.33 in week 14, on a 5-point scale (see Figure 1). Concurrently, the median usage rate of the AI tutor increased slightly, and the variance in AI tutor usage among the students also widened as the semester progressed, indicating diverging interaction strategies over time.

An analysis of the types of AI support requested and categorized by perceived task difficulty reveals distinct usage patterns (see **Error! Reference source not found.**). For tasks perceived as very easy (difficulty level 1), the AI tutor was largely ignored, with students opting not to use the system in 97 % of these cases. Across all difficulty levels, students rarely requested only the final result of an exercise. As the perceived difficulty increased to moderate and high levels, however, students predominantly used the AI tutor to verify their own calculations, search for errors or request hints and approaches for the next steps. Requests for a complete explanation of the calculation path were exclusively reserved for the most difficult tasks, specifically those rated as difficulty levels 4 or 5.

At the end of the course, the students' final exam scores were compared against their average perceived difficulty ratings for the exercises completed throughout the semester. As shown in Figure 3, the data indicates a weak negative correlation between these two. Students who rated the weekly tasks as more difficult on average tended to achieve lower scores in the final examination.

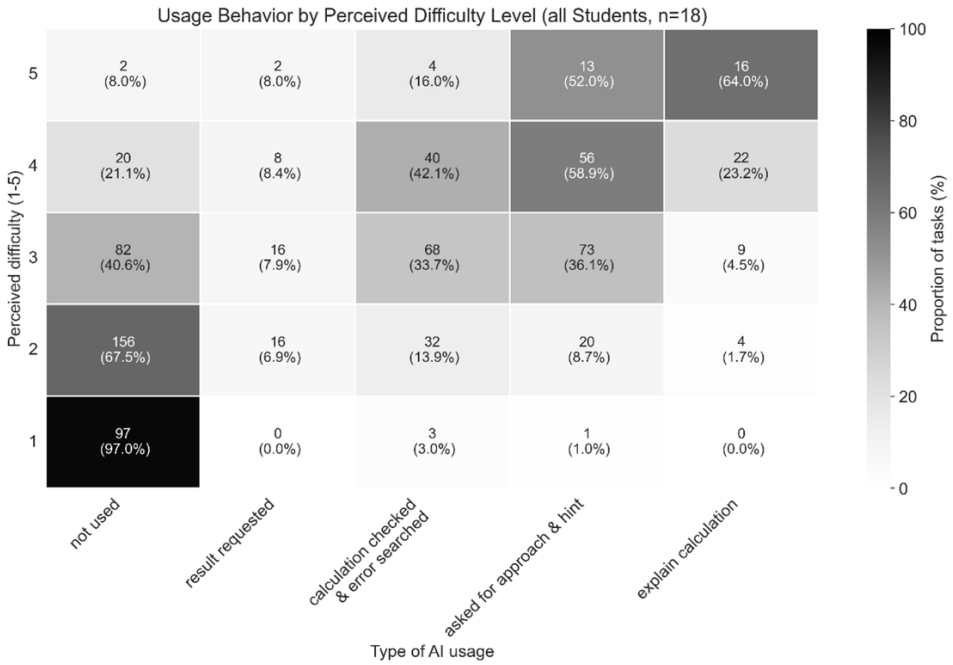


Figure 2: Usage of AI tutor by perceived difficulty of exercise tasks. Each entry gives the absolute number of responses and the percentage of tasks for which the respective type of AI usage was used. Because multiple choices were allowed, the total percentage in a row may exceed 100 %.

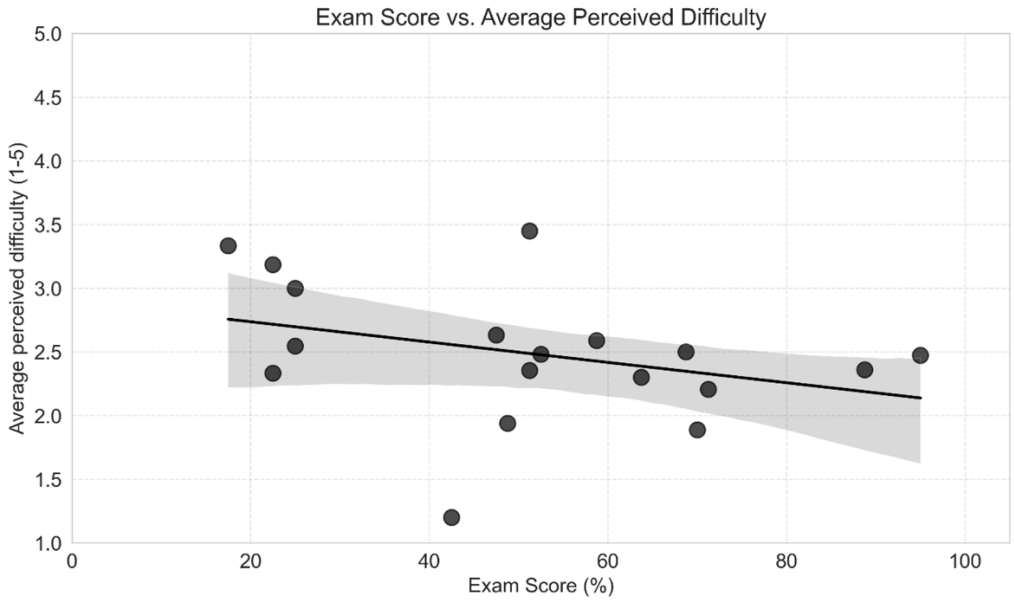


Figure 3: Average perceived difficulty of exercise tasks versus the score in the final exam, where each dot represents one student. The black line shows a linear regression, and the gray area indicates the 95 % confidence interval.

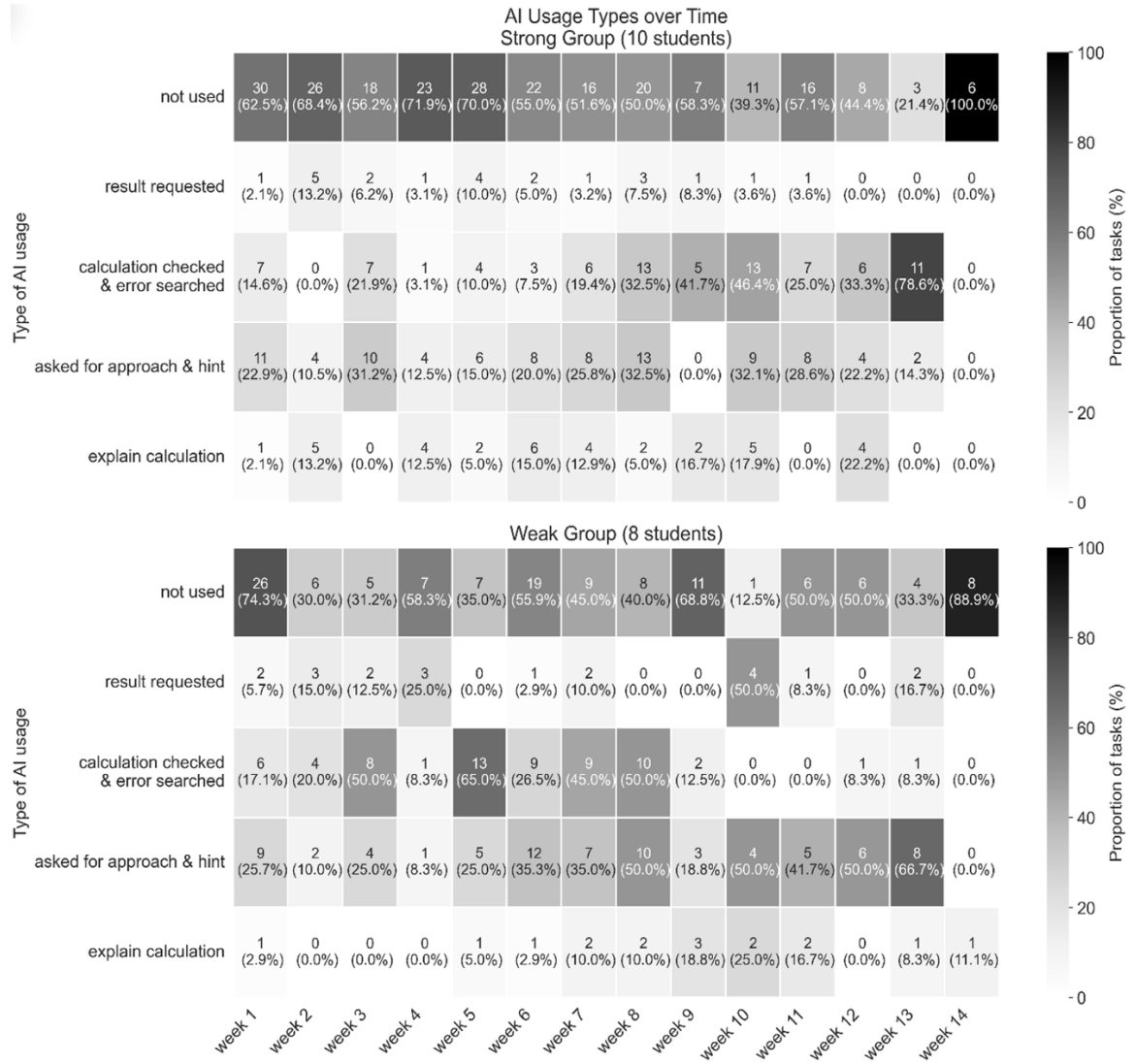


Figure 4: Usage of AI tutor over the semester for the cohort of high-performing students (top) and low-performing students (bottom). Each entry gives the absolute number of responses for all tasks of the respective week (up to 5 responses for each student possible) and the percentage of tasks for which the respective type of AI usage was used. Because multiple choices are allowed, the total percentage in a column may exceed 100 %. Example: In the lower diagram (weak group) at week 5 over all exercise tasks a total of 13 responses to the categories "own calculation checked" and "error in own calculation searched" were registered and the students of this cohort used the AI in 65 % of all tasks to "check own calculation" or to "find errors in own calculations".

To gain a better understanding of usage behaviour, the cohort was divided retrospectively into two groups based on final exam performance: higher-performing students (10 students), who scored at least 50 %, and lower-performing students (8 students), who scored below 50 %. Figure 4 shows the longitudinal data highlighting differences in how these two groups interacted with the AI tutor in terms of intensity and regularity. The stronger group demonstrated a more consistent usage pattern throughout the semester, frequently using the AI to check their calculation steps, particularly as the semester progressed. In contrast, the weaker group exhibited slightly different behavior: Their engagement in checking their own calculations dropped significantly in the later weeks. Towards the end of the semester, the proportion of tasks for which the AI was not used at all was often noticeably higher for the lower-performing group than for their higher-performing peers.

The didactic focus of the AI interactions also varied considerably between the two performance cohorts. The higher-performing students primarily utilized the AI tutor as a validation tool. Their requests to check calculations and search for errors increased substantially towards the end of the semester. For this group, the AI functioned as an efficiency tool to secure their own academic performance. In contrast, the lower-performing students utilized AI more frequently to overcome cognitive blockages. They regularly asked for general

approaches and hints to find an entry point into the tasks. Furthermore, during the first half of the semester (weeks 2 to 4), the weak group tended slightly more to request the final result outright (up to 25 % of tasks), whereas the strong group rarely exhibited this behavior. This suggests a stronger tendency among the weaker cohort to consume ready-made solutions.

As the inherent complexity and the perceived difficulty of the course material increased throughout the semester, the two groups reacted differently. The strong group responded by using the AI tutor more frequently to verify complex calculation steps. In contrast, the weaker group showed a growing tendency to fall behind. By week 14, 88.9 % of the weak group had not used the AI tutor. This high rate of non-usage, combined with the overall decline in questionnaire response rates, suggests potential resignation or abandonment of the exercise tasks altogether.

Qualitative data gathered from the open-text fields of the weekly questionnaires provides deeper insights into these usage patterns. Students appreciated the Socratic guidance, with one stating that "the approach helped me solve the task without the AI giving the answer directly, allowing me to complete it independently". The AI also served as a safety net, providing a "sense of security" by enabling students to "have the entire solution path explained and ask specific follow-up questions". One student specifically highlighted its value in error detection: "I make many careless mistakes, and the AI sometimes helps me with them."

In contrast, critical feedback highlighted significant technical and pedagogical limitations. Some users reported that "the AI makes gross errors" and noted that certain tasks were "too complex for the AI in its current state". The lack of multimodal capabilities was particularly criticized, with one student remarking that "the AI tutor is somehow not suitable for more graphical explanations" as certain concepts can only be understood "if someone explains them using the circuit diagram". Finally, some students found the AI's responses "too general," which "did not help move the calculation forward". These reported limitations can be attributed to two main factors. Firstly, the errors may originate directly from the system's inherent technical constraints. Alternatively, the issues may stem from how the AI model is operated by the users. A lack of AI literacy among students can result in inputs that are inadequate or poorly structured. Consequently, suboptimal prompting can exacerbate the model's tendency to hallucinate, leading directly to an increase in erroneous or overly general outputs.

The end-of-semester evaluation revealed that the participating students generally had a positive perception of the AI tutor. As shown in Figure 5Figure 5, the students particularly valued the system's ability to verify their individual calculation methods. The tool was recognized for its effectiveness in helping students to overcome short-term learning blockades, thereby successfully promoting independent and asynchronous learning outside of the traditional classroom environment and schedule.

The evaluation of the AI tutor revealed a clear correlation between the increasing perceived task difficulty and course content progression. Students perceived topics introduced later in the semester, such as network analysis, dependent sources and nodal voltage analysis, as significantly more complex than basic calculations involving elementary electrical quantities. Despite this increasing complexity, students reported very few technical issues when interacting with the AI tutor (see Figure 5). The final evaluation indicates that students generally found the AI's answers comprehensible. In some cases, however, they felt the tutor gave away the solution paths unprompted. Overall, the AI was perceived as a highly beneficial tool for finding an entry point into complex tasks. Students emphasized that the system enabled them to start assignments that they would otherwise not have been able to work on and prevented them from getting stuck without making progress. Consequently, using the AI tutor can significantly reduce cognitive overload by supporting the problem-solving process. Furthermore, the system proved highly effective in helping students identify computational errors that they would otherwise have overlooked.

The AI's ability to identify basic calculation errors could reduce the need for in-person teaching. By delegating routine error correction to the AI, valuable time during exercise sessions can be freed up for deeper, more conceptual discussions. However, it is important to emphasize that the AI tutor is not a standalone replacement for human instructors. Students strongly disagreed with the notion that a technically mature AI tutor eliminates the need for a weekly exercise session with a human teacher. Additionally, the results critically show that the AI tutor only marginally increased the students' intrinsic motivation to solve the tasks.

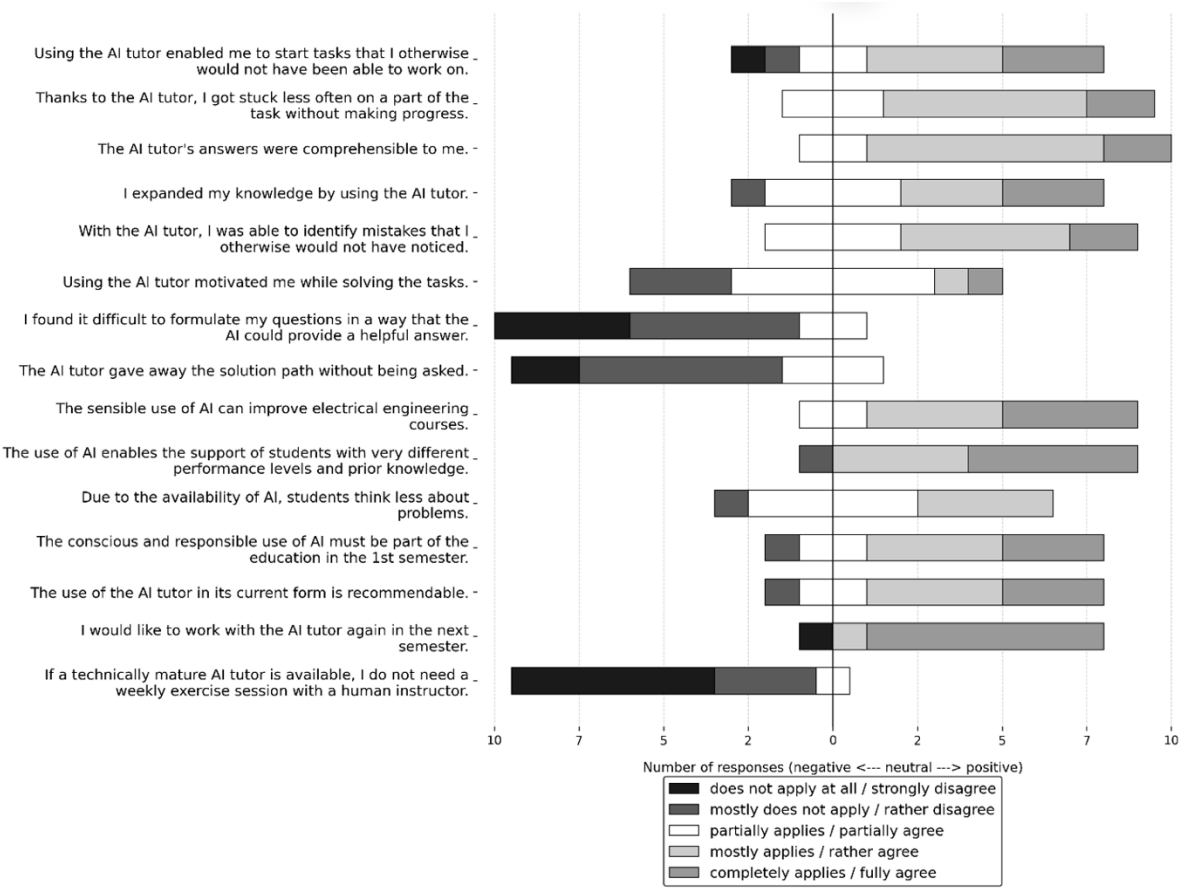


Figure 5: Final evaluation of the AI tutor. (excerpt; questions with redundant content have been omitted)

A highly critical finding of the final evaluation is the students' self-reflection regarding their own cognitive engagement: a large majority (9 out of 10 students) partially or rather agreed that the availability of AI leads to students thinking less about the problems (see Figure 5). This alarming trend highlights the urgent need to integrate “AI literacy” into the university curriculum. Students strongly agreed that learning the conscious and responsible use of AI must be a mandatory part of first-semester education. The ability to critically question and evaluate AI-generated results is an essential skill that must be actively taught to prevent over-reliance and superficial learning. Despite these concerns, there remains a strong consensus among the cohort that the sensible use of AI can generally improve electrical engineering courses.

Finally, the methodological limitations of the current study must be reflected critically: The overall validity of the findings is constrained by the small group size and a noticeably declining response rate in the feedback questionnaires over the semester. This drop in participation threatens the representativeness of the late-semester data. However, this is also consistent with the frequently observed decline in commitment to class attendance and the preparation of homework assignments over the semester. Furthermore, the technical implementation focused entirely on one specific language model. To meet the specific and complex demands of engineering education in the future, true multimodality will be required. Future systems must be capable of processing, interpreting, and *generating* circuit diagrams, technical charts, and visual schematics to provide adequate and comprehensive support.

5. Outlook and Conclusion

From the authors' perspective, future developments should focus on an integrated, AI-driven learning environment capable of continuously assessing individual learning progress to provide targeted and personalized recommendations on established and purpose-built teaching methods. To complement this student-centered approach, it is highly desirable to establish a data pipeline that provides instructors with aggregated, anonymized insights into exercise performance, usage patterns, student–AI communication, and emerging hallucinations. This would enable lecturers to adapt content dynamically, implement demand-oriented interventions and review or restructure complex topics based on empirical learning analytics.

In the current semester, the pilot project has been scaled up to include a larger cohort of around 200 second-semester students. This includes a comparative evaluation of different LLMs regarding their practical usability. However, expanding the use of these tools highlights the importance of fostering AI literacy among students and educators to ensure accurate and productive interactions. Although enabling users to customize and optimize AI tutors according to their specific needs offers significant potential for sustainable personalization, this fundamentally requires highly responsible and ethical engagement with generative AI. Ultimately, the core findings confirm that, while the AI tutor is an effective instrument for personalizing undergraduate education, it remains supplementary to, rather than a replacement for, human mentorship and supervision.

Ethics Declaration

Participation in this study was strictly voluntary. All collected data was analyzed anonymously, and prior informed consent was obtained from all participating students.

AI Declaration

During the preparation of this work, the authors used Google Gemini 3.1 Pro to generate Python program code for the creation of diagrams in Figures 1-5. Specifically, the generated code was used to process the database containing the feedback forms collected during the semester and the results from the mid-term and final evaluations, and to visualize the respective data series. Additionally, AI assistance was employed to refine the orthography, grammar, and language throughout the manuscript.

Sustainable Development Goals

Below are all the key United Nations Sustainable Development Goals (United Nations, 2026) that are relevant to this paper: Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities; Goal 5: Achieve gender equality and empower all women and girls; Goal 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all; Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation; Goal 10: Reduce inequality within and among countries; Goal 16: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.

References

Asgar, D. (2025) “Analysing the Effectiveness of Different AI-Based Tutoring Systems and their Impact on Education across Global Contexts: A Literature Review,” *The National High School Journal of Science* [Preprint]. Available at: https://nhsjs.com/2025/analysing-the-effectiveness-of-different-ai-based-tutoring-systems-and-their-impact-on-education-across-global-contexts-a-literature-review/.

Batsaikhan, B. (Zack) and Correia, A.-P. (2024) “The effects of Generative Artificial Intelligence on Intelligent Tutoring Systems in higher education: A systematic review,” *Studies in Technology Enhanced Learning*, 4(1). Available at: https://doi.org/10.21428/8c225f6e.33570bb1.

Casanova, J. *et al.* (2021) “University Dropout in Engineering: Motives and Student Trajectories,” *Psicothema*, 4(33), pp. 595–601. Available at: https://doi.org/10.7334/psicothema2020.363.

Farrokhnia, M. *et al.* (2024) “A SWOT analysis of ChatGPT: Implications for educational practice and research,” *Innovations in Education and Teaching International*, 61(3), pp. 460–474. Available at: https://doi.org/10.1080/14703297.2023.2195846.

Fierro Saltos, W.R. *et al.* (2025) “Leveraging Artificial Intelligence for Sustainable Tutoring and Dropout Prevention in Higher Education: A Scoping Review on Digital Transformation,” *Information*, 16(9), p. 819. Available at: https://doi.org/10.3390/info16090819.

Google LLC (2026) *Gemini 2.5 Pro Documentation*, *online*. Available at: https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/gemini/2-5-pro?hl=de (Accessed: May 21, 2026).

Hartmann, R. (2025) “Finanzielle Sorgen an Sachsens Hochschulen – Kürzungen treffen auch Gleichstellung und Lehre.” Available at: https://www.mdr.de/nachrichten/sachsen/politik/uni-hochschulen-finanzierung-sparen-kuerzungen-100.html.

John, C. (2026) “Another Wave of Faculty and Program Cuts Is Hitting American Universities. Here’s Who’s Cutting What — and Why It Won’t Stop.” Available at: https://www.universityherald.com/articles/80226/20260316/another-wave-faculty-program-cuts-hitting-american-universities-heres-whos-cutting-what.htm.

Kestin, G. *et al.* (2025) “AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting,” *Scientific Reports*, 15(1), p. 17458. Available at: https://doi.org/10.1038/s41598-025-97652-6.

Kim, E. and Pak, S.-J. (2002) “Students do not overcome conceptual difficulties after solving 1000 traditional problems,” *American Journal of Physics*, 70(7), pp. 759–765. Available at: https://doi.org/10.1119/1.1484151.
Koudsia, S. and Kirchner, M. (2024) “Reducing cognitive overload for students in higher education: A course design case study,” *Journal of Higher Education Theory and Practice*, 24(10), pp. 97–124.
Krupp, L. *et al.* (2024) “Unreflected Acceptance – Investigating the Negative Consequences of ChatGPT-Assisted Problem Solving in Physics Education.” Available at: https://doi.org/10.3233/FAIA240195.
Lewis, P. *et al.* (2021) “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” Available at: http://arxiv.org/abs/2005.11401.
Lin, C.-C., Huang, A.Y.Q. and Lu, O.H.T. (2023) “Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review,” *Smart learning environments*, 10(1), p. 41.
Luo, J. *et al.* (2025) “Design and assessment of AI-based learning tools in higher education: a systematic review,” *International Journal of Educational Technology in Higher Education*, 22(1), p. 42. Available at: https://doi.org/10.1186/s41239-025-00540-2.
Maurya, K.K. *et al.* (2025) “Unifying AI Tutor Evaluation: An Evaluation Taxonomy for Pedagogical Ability Assessment of LLM-Powered AI Tutors,” *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 1234–1251. Available at: https://doi.org/10.18653/v1/2025.naacl-long.57.
Naylor, R., Baik, C. and Arkoudis, S. (2018) “Identifying attrition risk based on the first year experience,” *Higher Education Research & Development*, 37(2), pp. 328–342. Available at: https://doi.org/10.1080/07294360.2017.1370438.
Qadir, J. (2023) “Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education,” *2023 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, pp. 1–9. Available at: https://doi.org/10.1109/EDUCON54358.2023.10125121.
Reicher, H. *et al.* (2025) “A Generative AI-Empowered Digital Tutor for Higher Education Courses,” *Information*, 16(4), p. 264. Available at: https://doi.org/10.3390/info16040264.
Santos-Guevara, A. *et al.* (2026) “Fostering Student Engagement and Learning Perception Through Socratic Dialogue with ChatGPT: A Case Study in Physics Education,” *Education Sciences*, 16(2), p. 184. Available at: https://doi.org/10.3390/educsci16020184.
Skelic, L. *et al.* (2025) “CIRCUIT: A Benchmark for Circuit Interpretation and Reasoning Capabilities of LLMs.” Available at: http://arxiv.org/abs/2502.07980.
Sultana, S., Khan, S. and Abbas, M.A. (2017) “Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts,” *International Journal of Electrical Engineering Education*, 54(2), pp. 105–118.
SuSea Inc (2026) *you.com chat*, *online*. Available at: https://you.com/chat (Accessed: May 21, 2026).
Sweller, J. (1988) “Cognitive load during problem solving: Effects on learning,” *Cognitive science*, 12(2), pp. 257–285.
Tayebi, A., Gómez, J. and Delgado, C. (2021) “Analysis on the Lack of Motivation and Dropout in Engineering Students in Spain,” *IEEE Access*, 9, pp. 66253–66265. Available at: https://doi.org/10.1109/ACCESS.2021.3076751.
Tight, M. (2020) “Student retention and engagement in higher education,” *Journal of Further and Higher Education*, 44(5), pp. 689–704. Available at: https://doi.org/10.1080/0309877X.2019.1576860.
United Nations (2026) *THE 17 GOALS*. Available at: https://sdgs.un.org/goals (Accessed: May 21, 2026).

Artificial Intelligence in Teaching and Learning: Opportunities, Tensions, and Transformation in Higher Education

Nathunathi Mvunge and Bongo Mqukuse
Teaching and Learning Centre, University of Fort Hare, South Africa

nmvunge@ufh.ac.za
bmqukuse@ufh.ac.za

Abstract: The growth of Artificial Intelligence (AI) is reshaping higher education globally, with increasing relevance for institutions such as the University of Fort Hare (UFH), which are navigating digital transformation within resource constrained contexts. At UFH, the emergence of AI tools presents both exciting opportunities to enhance teaching and learning, and complex challenges related to ethics, academic integrity, and digital readiness. While AI has the potential to support more flexible and personalized learning, there remains uncertainty about how it is being used and experienced by both students and lecturers. This study explores the use and implications of AI in teaching and learning at UFH, with a specific focus on students. It aims to examine the extent of AI adoption among students, how they engage with these tools in their everyday academic practices, and the implications of this engagement for the future of teaching, learning, and assessment within the institution. A mixed methods case study design was employed. Quantitative data was collected through survey questionnaires administered to students to identify patterns of AI use, perceived benefits, and levels of digital competence. This was complemented by qualitative data from semi-structured interviews and focus group discussions, which provided deeper insights into participants' experiences, concerns, and expectations. The study is informed by the Community of Inquiry (CoI) framework. The findings reveal that while students are increasingly turning to AI tools for academic support such as idea generation, summarization, and content clarification. Participants expressed optimism about AI's potential to enhance learning and concern about issues such as over reliance, plagiarism, and the erosion of critical thinking. The study also highlights gaps in institutional policy and development related to AI. The study concludes that AI can play a meaningful role in advancing learning at UFH, but its successful integration depends on fostering a shared `z understanding of ethical use, strengthening digital literacy, and providing ongoing support for both staff and students.

Keywords: Artificial Intelligence, Higher Education, Digital Literacy, Teaching and Learning, Academic Integrity

1. Introduction and Background

Artificial Intelligence (AI) has become an increasingly important part of teaching and learning in higher education. The growing availability of generative AI tools has changed the ways in which students access information, engage with learning content, develop ideas, and complete academic tasks. Students can now use AI tools to clarify difficult concepts, generate ideas, summarise information and obtain additional academic support outside the classroom. Valtonen et al (2021) argue that these developments present opportunities for universities to enhance learning and provide students with more flexible forms of academic support. At the same time, the growing use of AI raises important questions about how students use these technologies and what that use means for teaching, learning, and assessment. According to Yu (2024), the increasing accessibility of AI is particularly significant because students can independently incorporate these tools into their everyday academic practices. Unlike institutional learning technologies, which are generally introduced and supported through formal university structures, many AI tools are accessed directly by students through personal devices. This makes it important for higher education institutions to understand students' experiences with AI rather than to make assumptions about how these technologies are used. Kim et al. (2022) state that the focus should not only be on whether students use AI, but also on how they use it, what they use it for, and whether such use contributes to meaningful learning. These developments need to be considered within the broader South African higher education context, where students have different levels of digital access, digital literacy, and technological experience. Previous research at the University of Fort Hare (UFH) has highlighted the challenges that students experience when engaging with digital learning environments, including differences in digital literacy, access to reliable connectivity, and technological proficiency. These challenges remain relevant in the context of AI. Access to an AI tool does not necessarily mean that students have the skills to evaluate the information it produces, identify inaccuracies, or use the technology appropriately within academic contexts (Huang, 2023).

At UFH, AI therefore presents both an opportunity and a challenge for teaching and learning. According to Mohebbi (2025), AI tools have the potential to provide students with additional academic support and promote independent learning. However, concerns relating to over-reliance, plagiarism, academic integrity and the possible weakening of critical thinking need to be considered. The important issue is to understand how AI can

be positioned as a learning support rather than as a replacement for students' engagement with academic content and the development of their own knowledge and skills. Against this background, this study explores the use and implications of AI in teaching and learning at UFH, with a particular focus on students. It examines the extent of AI adoption among students, how they use AI tools in their everyday academic practices, the benefits and challenges they experience, and the implications of AI use for teaching, learning, and assessment. The study seeks to provide practical insights that can support the development of responsible AI practices, strengthen digital literacy, and inform academic development and institutional approaches to AI integration in higher education.

2. Literature Review

2.1 Artificial Intelligence and the Transformation of Pedagogical Practice

The emergence of AI has prompted renewed debate about the nature and purpose of teaching and learning in higher education. Bearman et al. (2023) argue that AI should not be viewed merely as another educational technology but rather as a development with significant implications for pedagogical practice and for the relationship between students, knowledge, and technology. This perspective is particularly relevant in the context of AI, which can produce explanations, arguments, and other forms of academic content. Jose et al. (2025) similarly highlight the capacity of AI to influence how students access, engage with, and construct knowledge. However, while AI's capabilities may expand students' access to academic support and information, they also challenge conventional understandings of learning, particularly when students rely on AI to generate content rather than engaging directly in inquiry and knowledge construction. The debate, therefore, is not simply about whether AI should be incorporated into higher education, but about how its integration can support meaningful learning without weakening the intellectual processes that underpin higher education. In this regard, the increasing use of AI requires lecturers to reconsider pedagogical approaches that emphasize inquiry, interpretation, critical reflection, and the application of knowledge, rather than the mere production of academic outputs.

This debate is further complicated by questions concerning human agency in the learning process. Cukurova (2025) argues that the interaction between students and AI systems requires greater attention to the distribution of agency between humans and technology. From this perspective, AI does not simply provide students with an additional source of information; it can increasingly participate in activities traditionally associated with the student, such as generating ideas, explaining concepts, and producing academic responses. This creates a tension between the potential of AI to enhance learning and the possibility that excessive reliance on AI may reduce students' engagement with independent thinking and academic judgment. While Bearman et al. (2023) emphasize the broader pedagogical transformation associated with AI, and Jose et al. (2025) draw attention to its capacity to reshape knowledge access and construction, Cukurova (2025) brings the debate towards the question of who retains responsibility for learning and knowledge construction when AI becomes an active participant in the learning process. Taken together, these perspectives suggest that effective AI integration cannot be based solely on providing students with access to AI tools. Instead, pedagogical designs should deliberately position students as active and critical participants who interrogate, evaluate, verify, and extend AI-generated outputs. AI should consequently function as a resource that supports human learning and judgment rather than a substitute for them. The literature therefore suggests a shift towards pedagogies that balance the affordances of AI with the preservation of learner agency, critical thinking, and meaningful knowledge construction.

2.2 AI-Mediated Knowledge Construction

The discussion of student agency and AI-supported pedagogy provides a foundation for understanding AI-mediated knowledge construction. Boretti (2026) identifies the reliability and nature of AI-generated knowledge as important concerns, noting that AI can produce responses that appear authoritative while containing factual inaccuracies, incomplete explanations or contextual limitations. This raises questions about how students evaluate and validate knowledge obtained through AI. However, rather than viewing these limitations solely as risks, Rana et al. (2025) argue that AI can create opportunities for deeper learning when students are required to verify, critique, and improve AI-generated outputs. Their position suggests that the challenge of unreliable AI-generated information can itself become a pedagogical opportunity for developing students' critical thinking and information-evaluation skills.

These perspectives therefore suggest that AI-mediated knowledge construction should involve active engagement rather than passive acceptance of AI-generated information. While Boretti (2026) emphasises the potential weaknesses of AI-generated knowledge, Rana et al. (2025) demonstrate how these weaknesses can encourage students to compare AI outputs with disciplinary knowledge, question assumptions and develop their own interpretations. This also reinforces the importance of human agency discussed in the previous section, as students remain responsible for evaluating information and making academic judgments. Consequently, the lecturer's role increasingly involves designing learning experiences that encourage students to interrogate, verify and contextualise AI-generated content. The educational value of AI, therefore, depends not simply on its ability to generate information but on the quality of the critical intellectual engagement surrounding its use.

2.3 Theoretical Framework

The Community of Inquiry (CoI) framework developed by Garrison, Anderson and Archer (2000) provides a useful theoretical perspective for examining learning in technology-mediated environments. The framework conceptualizes meaningful learning through the interaction of cognitive presence, social presence, and teaching presence. Its value for the study lies in its ability to shift attention from the technology itself to the quality of the learning experience created around it.

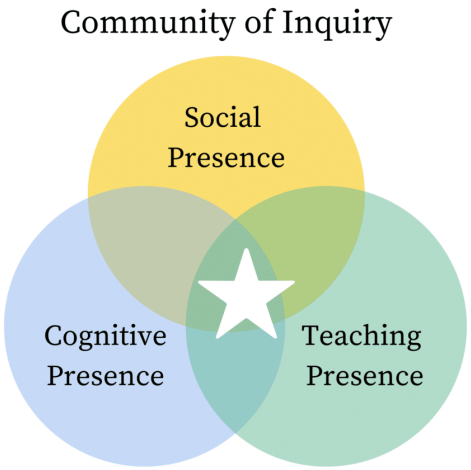


Figure 1: Community of Inquiry Framework

Cognitive presence is concerned with students' engagement in processes of inquiry, reflection and the construction of meaning. This dimension is particularly relevant to AI-supported learning because students may use AI as part of processes through which they explore and develop ideas. Social presence draws attention to communication and participation within a learning community, reminding us that technology-mediated learning remains relational. Teaching presence focuses on the design, facilitation and direction of learning and is therefore important for understanding how lectures establish purposeful boundaries and learning activities around AI use. The CoI framework consequently provides an appropriate lens for examining AI not as an isolated technological intervention but as part of a broader educational environment. It enables attention to be given to the interaction between students, lectures, knowledge and technology.

3. Methodology

3.1 Research Design

A parallel mixed-methods design was employed in the study to examine undergraduate students' experiences, perceptions, and challenges related to AI use. In this approach, quantitative data was collected through structured surveys, enabling statistical analysis of students' patterns of AI use, perceived usefulness, engagement, and challenges encountered when using AI tools for learning. At the same time, qualitative data were collected through open-ended survey questions, providing deeper insight into students' experiences, motivations, concerns, and contextual factors influencing their use of AI. The quantitative and qualitative data were collected concurrently but analysed separately before being integrated during interpretation to provide a more comprehensive understanding of students' engagement with AI-mediated learning. This concurrent design

ensured that both measurable trends and students' individual perspectives were considered in understanding the educational implications of generative AI. The combination of closed- and open-ended survey items further strengthened the study by providing both breadth and depth of evidence, allowing the findings to capture not only how students use AI but also why they engage with these technologies and under what circumstances. According to Clark (2019), such a mixed-methods approach is particularly appropriate for educational research involving emerging technologies, as it enables researchers to integrate numerical evidence with contextual accounts of students' experiences and perceptions.

3.2 Sampling

In this study, an academic bootcamp was conducted for first-year students from the Faculties of Science and Agriculture, and of Law, at the University of Fort Hare. The bootcamp was designed to support students' transition into university learning by addressing key academic and digital competencies, including a dedicated session on the effective and responsible use of AI in learning. Attendance at the bootcamp was mandatory for first-year students from the participating faculties. Following the bootcamp, a questionnaire was distributed to students through Google Forms to collect their perceptions, experiences, and engagement with AI in their academic learning. Of the approximately 600 students who attended the bootcamp, 347 completed and submitted the questionnaire, resulting in a response rate of approximately 57.8%. The responses provided the empirical data for examining students' perceptions and experiences of AI and its role in supporting teaching and learning in higher education.

3.3 Data Collection

In this study, data was collected using both quantitative and qualitative approaches to obtain a comprehensive understanding of students' perceptions and experiences of AI. Quantitative data were gathered through closed-ended questions and rating scales in the questionnaire administered via Google Forms. These responses generated numerical data that enabled the identification of patterns and trends relating to students' use of AI, perceived usefulness, engagement, and challenges associated with its use for academic purposes. In addition, qualitative data was collected through open-ended questions that provided students with opportunities to describe their experiences, views, and reflections on the use of AI in learning. These narrative responses provided contextual insights that extended beyond the numerical findings and helped capture students' motivations, concerns, benefits, and challenges associated with AI use. The integration of quantitative and qualitative responses, therefore, provided both breadth and depth in understanding how students engage with generative AI within the university learning environment.

3.4 Data Analysis

The data were analysed using both quantitative and qualitative approaches and interpreted through the Community of Inquiry (CoI) framework. Quantitative responses were analysed using descriptive statistics to identify patterns in students' use of AI, including the purposes for which AI tools were used and students' perceptions of their academic value. Qualitative responses were analysed thematically to identify recurring experiences, concerns and perspectives relating to the use of AI in learning. The interpretation of the findings was organised around the three CoI presences: cognitive, social and teaching presence. Cognitive presence was used to examine how students used AI to support understanding, knowledge construction, and critical engagement with academic content. Social presence is considered the extent to which AI influences students' interaction and engagement within the learning environment, while teaching presence focuses on the role of lecturer guidance and institutional support in shaping responsible and meaningful AI use. Using these three dimensions enabled the study to identify both the educational value and emerging challenges of AI use and to consider areas where further student support, academic development and institutional guidance may be required.

4. Results and Discussion

The findings were interpreted through the Community of Inquiry (CoI) framework, which conceptualises meaningful learning through the interaction of cognitive presence, teaching presence and social presence. In the context of this study, the framework was applied to understand how first-year students engage intellectually with AI, the extent to which they require structured academic guidance in using AI, and how AI may support interaction and participation within the broader learning environment.

4.1 Students' Exposure to and Use of Generative AI

The findings demonstrate that AI was already well established within students' academic practices prior to the AI presentation. Of the 347 respondents, 300 (86.5%) reported having used AI tools such as ChatGPT, Gemini, and Copilot for their academic studies, while only 47 (13.5%) reported not having done so. This suggests that first-year students are entering higher education with considerable prior exposure to AI. However, exposure to AI should not necessarily be taken as evidence that students possess the academic and digital literacy required to use these technologies appropriately.

Students also differed in the frequency with which they used AI. "Sometimes" was the most frequently selected response, appearing in 61.1% of responses, followed by "occasionally" (21.0%), "often" (12.7%) and "always" (5.5%). "Never" was selected by only 2.9%. Because a small number of students selected more than one frequency category, these percentages should be treated as multiple-response frequencies rather than mutually exclusive categories. Nevertheless, the overall pattern reinforces the finding that AI use was already part of the academic practices of most participating students.

4.2 Cognitive Presence: AI as a Tool for Knowledge Construction

The strongest evidence in the dataset relates to cognitive presence, which concerns students' construction and confirmation of meaning through reflection, exploration, and application of knowledge. Students predominantly reported using AI to support learning activities that involve understanding and processing academic information.

Table 1:

Purpose for Using AI	Number Selecting	Percentage of Respondents
Understanding concepts	256	73.8%
Summarising information	142	40.9%
Brainstorming ideas	123	35.4%
Editing grammar and language	74	21.3%
Writing assignments	36	10.4%

The finding that 73.8% of respondents use AI to understand concepts is particularly significant from a CoI perspective. It suggests that students primarily perceive AI as a cognitive support mechanism rather than simply a tool for producing assignments. Similarly, the use of AI for summarising information (40.9%) and brainstorming (35.4%) indicates that students are incorporating AI into processes of exploring, organising and making sense of academic knowledge.

These findings support the current study's emphasis on AI-mediated knowledge construction. AI appears to provide students with an additional space in which they can interrogate concepts, generate ideas and simplify complex information. However, cognitive presence requires more than obtaining information from AI. Meaningful learning depends on students being able to critically evaluate, verify, and integrate AI-generated information into their existing knowledge. The relatively smaller proportion using AI for writing assignments (10.4%) is also noteworthy, as it challenges an assumption that students primarily engage with AI to produce assessed work. Instead, the data suggest that its dominant function among these students is related to understanding and supporting learning.

4.3 Teaching Presence: The Need for Guidance and Responsible AI Use

The findings also reveal a strong relationship between students' experiences and teaching presence. Within the CoI framework, teaching presence involves the design, facilitation and direction of learning experiences. In an AI-mediated learning environment, this extends to lecturers and institutions providing students with guidance on how, when and why AI should be used. Students demonstrated varying levels of prior understanding of academic integrity and responsible AI use. "Fair" was selected in 45.2% of responses and "good" in 38.9%, while "excellent" was selected in only 13.0%. Poor and very poor understanding were selected in 6.6% and 1.4% respectively. Again, because some students selected multiple categories, these figures represent multiple-response frequencies.

More importantly, students' open-ended responses repeatedly expressed a desire to learn how to use AI correctly, effectively and ethically. Students expressed a desire to understand responsible AI use, plagiarism,

academic integrity, appropriate use of AI in research, the advantages and disadvantages of AI, and how to use AI without becoming overly dependent on it. Responses included concerns about using AI "responsibly," "ethically," "correctly," and in ways that would avoid plagiarism or disciplinary consequences. This finding highlights an important tension in students' engagement with AI, students are already using the technology extensively, but many remain uncertain about the academic rules and practices that should govern its use. Teaching presence therefore becomes increasingly important in AI-mediated learning. Rather than merely restricting AI use, lecturers may need to explicitly facilitate students' development of AI literacy by demonstrating appropriate prompting, verification of AI-generated information, acknowledgement of AI use, academic integrity and critical evaluation of AI outputs. The bootcamp can therefore be understood as an institutional form of teaching presence intended to bridge the gap between students' existing AI practices and responsible academic engagement.

4.4 Social Presence

The findings provide some indication of how students' engagement with AI relates to their broader learning environment. Students' responses suggest that AI was predominantly used as an individual academic support tool, particularly for understanding concepts, summarising information, brainstorming ideas and improving academic work, with limited evidence of its use for peer interaction or collaborative learning. However, students' expressed interest in learning how to use AI responsibly, ethically and appropriately within the university context reflects an awareness that AI use takes place within a wider academic community governed by shared expectations and practices. The academic bootcamp also provided a collective learning environment in which students were introduced to common understandings of responsible AI use. From a Community of Inquiry perspective, these findings suggest that social presence was less visible than cognitive and teaching presence, particularly in relation to peer interaction and collaborative knowledge construction. This highlights an opportunity for lecturers to incorporate collaborative AI-supported activities, such as students jointly evaluating AI-generated responses, discussing inaccuracies and biases, and comparing AI outputs with academic sources, thereby strengthening dialogue, peer learning and social presence within AI-mediated learning environments.

4.5 Integration of the Community of Inquiry Dimensions

Viewed collectively, the findings suggest that cognitive presence is currently the most visible dimension of students' engagement with AI, particularly through concept clarification, summarisation and brainstorming. Teaching presence emerges as an important enabling condition because students require structured guidance to transform frequent AI use into academically responsible and meaningful learning. Social presence is comparatively less evident and represents an area that could be deliberately strengthened through collaborative AI-supported learning activities. The findings therefore indicate that access to AI alone does not necessarily produce meaningful AI-mediated learning. Students may already be comfortable using AI, but effective educational integration requires lecturers and institutions to create learning environments in which students are encouraged to question, verify, discuss and critically evaluate AI-generated knowledge. From a CoI perspective, the educational value of AI is therefore likely to emerge not from the technology itself, but from the interaction between students' cognitive engagement, appropriate teaching guidance and opportunities for social and collaborative learning.

5. Conclusion

The study examined first-year students' perceptions and experiences of AI in higher education through the lens of the Community of Inquiry framework. The findings demonstrate that AI is already embedded in students' learning practices, with 86.5% of respondents having used AI tools for academic purposes prior to the bootcamp. Cognitive presence emerged most strongly, as students primarily used AI to understand concepts, summarise information and brainstorm ideas, suggesting that AI is increasingly functioning as a learning support tool for knowledge construction. Teaching presence was equally important, as students expressed a need for guidance on responsible, ethical and academically appropriate AI use, highlighting a distinction between being able to use AI and possessing the AI literacy necessary to use it critically. Social presence was less visible, with students predominantly engaging with AI individually rather than through collaborative learning activities. Overall, the findings indicate that meaningful AI-mediated learning requires an appropriate interaction between cognitive, teaching and social presence, consistent with the CoI understanding that these three dimensions are interdependent in creating meaningful educational experiences.

6. Recommendations

The study recommends that universities move beyond simply providing access to AI and adopt a structured approach to developing AI literacy among students from the beginning of their university studies. Academic orientation and first-year programmes should include practical training on effective prompting, critical evaluation and verification of AI-generated information, ethical AI use, academic integrity, acknowledgement of AI assistance, and the limitations and potential biases of generative AI. Lecturers should also intentionally integrate AI into learning activities that promote critical thinking rather than passive dependence, for example by requiring students to compare AI-generated responses with scholarly sources, identify inaccuracies and justify whether AI-generated information is credible. Furthermore, collaborative AI-supported activities should be introduced to strengthen social presence through peer discussion, group problem-solving and collective evaluation of AI outputs. Institutions should simultaneously provide lecturers with professional development on the pedagogically appropriate integration of AI so that teaching presence remains central to AI-mediated learning. Such an approach would position generative AI as a tool that supports, rather than replaces, human interaction, critical reflection and collaborative knowledge construction in higher education.

Ethics Declaration

The study was conducted in accordance with the ethical principles and institutional requirements applicable to research involving human participants. Ethical clearance was obtained from the Inter-Faculty Human Research Ethics Committee (Reference Number: IFHREC-26/08-0679). Prior to participation, students were informed about the purpose of the study, the research process, and their right to participate voluntarily or withdraw without consequence. Informed consent was obtained from all participants. Appropriate measures were also implemented to protect participants' anonymity, privacy, and confidentiality throughout the research process, while the collected data were handled and stored in accordance with applicable data protection requirements.

AI Declaration

Generative Artificial Intelligence, specifically ChatGPT (OpenAI), was used during the preparation of this paper as a supporting tool for language refinement, improving clarity of expression, and assisting with the organisation and development of ideas. Its use did not replace the authors' academic judgement, interpretation of the findings, or responsibility for the scholarly content of the paper. All arguments, interpretations, and final decisions regarding the study's content and structure remained the responsibility of the authors. The use of AI was undertaken in accordance with relevant institutional and ethical requirements.

References

Bearman, M. and Ajjawi, R., 2023. Learning to work with the black box: Pedagogy for a world with artificial intelligence. British Journal of Educational Technology, 54(5), pp.1160-1173.

Boretti, A., 2026. Hallucinations in generative AI: A threat to scholarly integrity and the urgent need for publisher-led academically supervised verification. Energy Research & Social Science, 136, p.104720.

Clark, V.L.P., 2019. Meaningful integration within mixed methods studies: Identifying why, what, when, and how. Contemporary educational psychology, 57, pp.106-111.

Cukurova, M., 2025. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. British Journal of Educational Technology, 56(2), pp.469-488.

Huang, C.W., Coleman, M., Gachago, D. and Van Belle, J.P., 2023, July. Using ChatGPT to encourage critical AI literacy skills and for assessment in higher education. In Annual conference of the southern african computer lecturers' association (pp. 105-118). Cham: Springer Nature Switzerland.

Jose, B., Cleetus, A., Joseph, B., Joseph, L., Jose, B. and John, A.K., 2025, August. Epistemic authority and generative AI in learning spaces: rethinking knowledge in the algorithmic age. In Frontiers in Education (Vol. 10, p. 1647687). Frontiers.

Kim, J., Lee, H. and Cho, Y.H., 2022. Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. Education and information technologies, 27(5), pp.6069-6104.

Mohebbi, A., 2025. Enabling learner independence and self-regulation in language education using AI tools: a systematic review. Cogent Education, 12(1), p.2433814.

Rana, V., Verhoeven, B. and Sharma, M., 2025. Generative AI in design thinking pedagogy: Enhancing creativity, critical thinking, and ethical reasoning in higher education. Journal of University Teaching and Learning Practice, 22(4), pp.1-22.

Valtonen, T., Leppänen, U., Hyypiä, M., Kokko, A., Manninen, J., Vartiainen, H., Sointu, E. and Hirsto, L., 2021. Learning environments preferred by university students: a shift toward informal and flexible learning environments. Learning Environments Research, 24(3), pp.371-388.

Yu, H., 2024. The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. *Heliyon*, *10*(2).

Non Academic Papers

Co-Designing AI Guidelines for Learners: Contexts, Principles, and Literacies

David Parsons[1], Kathryn MacCallum[2] and Stephen Powell[3]
[1] academyEX, Auckland, New Zealand
[2] University of Canterbury, Christchurch, New Zealand
[3] University of Waikato, Hamilton, New Zealand

david.parsons@academyex.com
kathryn.maccallum@canterbury.ac.nz
stephen.powell@waikato.ac.nz

Abstract: AI has become an integral component of learning contexts, so it is increasingly necessary for educators to provide context-specific guidelines to support learners in the responsible and effective use of AI. It is widely recognised that providing AI usage guidelines for learners is a contextual exercise, where the approach is unique to the type of learner and what they are studying, and as such, approaches need to be adapted. Not only is usage contextual, but so is the understanding and ability of students to apply these guidelines. As a result, co-design is recommended as a method to ensure that AI usage guidelines gain the necessary buy-in from all stakeholders and are designed in a supportive and comprehensible way. This paper argues that AI guidelines become meaningful only when a set of guiding principles is integrated with AI literacy and practice-based supports. This insight was developed from a co-design process involving a range of stakeholders from a predominantly online post-graduate educational institution in New Zealand. Over a period of nine months, students, staff, and external stakeholders engaged in an iterative process, using a range of methods including surveys, focus groups, and workshops, to design a set of effective guidelines around AI use. Specific insights included the difficulty of defining guidelines without first identifying a common set of principles, the value of aligning these principles to an AI literacy framework, and the need to supplement guidelines with further information and explanations of how guidelines may be applied in practice. The process also raised additional insights around the difficulties of defining guidelines through a consensus-building process - finding ways to make guidelines both discrete and comprehensive, and going beyond the skills aspects of AI literacy to encompass issues such as ethical practice and cultural awareness. This paper outlines the overall co-design process that was applied, the rationale for doing so, the methods used, the issues encountered and how they were resolved. It also provides the guidelines that were developed as an outcome of this process.

Keywords: AI Literacy; Co-Design; Guidelines; Postgraduate; Focus Group; Survey; Principles

1. Introduction

This article discusses the motivation and process for co-designing guidelines for the responsible use of AI in a postgraduate institution in New Zealand. A key assumption of this effort was that each educational context, subject, or discipline would interpret the appropriate uses of AI tools differently, taking a human-centred approach and shaped by the nature of the institution and the types of students who enrol (Fu & Weng, 2024). Therefore, it is unlikely that using generic guidelines from other contexts will provide an ideal framework for progress or ensure student buy-in to the process. It therefore follows that a collaborative and situated process, such as a co-design approach, is necessary to reflect the specific characteristics of the institution and its learning community with the development of guidelines involving a diverse range of stakeholders (Brandon et al, 2025). Applying a co-design process can result in transformative, rather than transactional, outcomes where students are more likely to follow the guidance provided (Ohio State University, n.d.). Students should not be bystanders in this process, as many are interested in opportunities to openly co-design their engagement with AI technologies (Cardona et al, 2023). Importantly, a set of guidelines openly co-designed with members of the student body is more likely to have success than one imposed from above by the institution.

Further, once created, guidelines must be actionable. This is not possible if learners lack the appropriate AI literacies to understand and follow them. Guidelines require specific literacies to reinforce learner agency with, rather than through, AI (Küçükuncular, 2025). AI guidelines and AI literacy are therefore fundamentally linked. Guidelines can support AI literacy development by turning abstract competencies into contextual applications (Sabatini et al, 2023), while AI literacy helps learners to interpret guidelines effectively. Guidance can then be operationalised through additional support materials (Funa & Gabay, 2025).

Good guidelines not only require transparency around use but should help learners understand educationally productive uses. The distinction between appropriate and inappropriate use should be explicit (Jin et al, 2025). Guidance should also focus less on integrity issues and more on the ethical and citizenship issues of AI use (Ghimire & Edwards, 2024). Centring values around fairness, transparency, cultural respect, disclosure, and

human-centred judgement helps learners understand that AI is a socio-technical issue, not simply a technical one (MacCallum, Parsons & Mohaghegh, 2026).

In summary, while many institutions create guidelines, there is often an assumption that learners have the appropriate literacies to interpret them, yet there is limited work examining how guidelines become actionable. Our research question was, therefore: How can AI guidelines be co-designed, so they become meaningful and usable within a specific educational context?

The contribution of this paper includes a process framework that captures how guidelines were scaffolded from principles and AI literacy and how they were expanded with practice support, providing insights into the complexities of co-designing AI guidance within a learning context.

2. Literature Review

The rapid emergence of GenAI tools has prompted many educational institutions to develop guidelines for AI use in teaching, learning, and assessment. However, many institutional guidelines overlook significant issues in AI tool use, privacy, and transparency (Ghimire & Edwards, 2024) as well as equitable access (Jin et al, 2025), and even where such guidelines exist, students may be uncertain about their institution's policies (Baek, Tate and Warschauer, 2024). In addition, successful implementations of such guidelines require supporting processes to develop relevant AI literacies in staff and students (Funa & Gabay, 2025). Therefore, this study was grounded in a set of understandings about what constitutes AI literacy, which were then used to support a co-design process.

2.1 AI Literacy

For students to meaningfully participate in designing AI guidelines, they require a certain level of AI literacy. For example, they need to be aware of issues like bias, data privacy, and reliability, so they can usefully contribute to discussions. The concept of AI literacy began to emerge in literature in 2016 (e.g., Kandlhofer et al., 2016). These and subsequent researchers proposed a range of concepts and definitions, many of which were integrated in Long and Magerko's (2020) literature review, which proposed five AI literacy themes linked to a set of competencies and design considerations. This work preceded the widespread use of generative AI but was part of a growing awareness that AI literacy is a broad, interdisciplinary theme (Fadel et al., 2024). From an educational perspective, several recent AI literacy frameworks have been developed to provide an organised way of thinking about how someone might scaffold their learning about AI in the age of generative AI. These include two frameworks from UNESCO, one for teachers (Miao & Cukurova, 2024) and another for students (Miao & Shiohira, 2024), and a framework for primary and secondary education by the OECD and the European Union (2026), among others. One such framework, which caters for multiple levels of AI literacy with a focus on equality AI literacy, is the SAIL framework (MacCallum, Parsons & Mohaghegh, 2024; 2026). This is the framework applied in this study.

2.2 Co-design

Co-design is a broad family of research approaches where people collaborate to carry out a design task, with variations in politics, practices, and epistemology (Zamenopoulos & Alexiou, 2018). Given this range of variables, co-designing guidelines not only requires student participation but also a process for how co-design is structured, implemented, and evaluated (Ambrosetti et al., 2025). Part of this process involves managing power dynamics, so students' contributions carry weight. There are several dimensions to this; for example, co-design should involve more than one cycle of engagement with students to ensure their voice is heard in the final product (Smith, F. et al, 2024). Further, Fielding (2001) proposes a 'students as researchers' model, based on the principles and values of partnership, positive orientation, centrality of learning, equity and authenticity, quality assurance, and ensuring that reflection leads to action. However, it should also be noted that effective facilitation, support, and guidance from academic staff were identified as the most important enablers of effective co-design in a study on assessment design, so staff need to take responsibility for the overall process and outcomes (Smith, A. et al, 2024). Similarly, Roschelle et al. (2006) note that co-design is not a fully democratic process, because accountability and decision-making are ultimately the responsibility of those leading the project. In this study, drawing from the recommendations in the literature, a small steering group managed the co-design process, with significant input from a range of staff, but students were included in multiple stages of the process, and importantly had the final say, as the last stage of the process consisted of student workshops.

3. Methods

The research design for this study was a co-design approach, with methods selected to suit the context. Co-design is a participatory design strategy that actively involves stakeholders, based on the premise that people affected by design decisions should help make those decisions (Zelenko, Gomez & Kelly, 2021). Many methods can be used, including workshops, focus groups, observations, and interviews (Iniesto, Littlejohn & Charitonos, 2022). The methods used within this study were a combination of a survey, focus groups and workshops, partly planned, and partly improvised to respond to the evolving context.

Co-design was chosen as a suitable way to actively involve members of a learning community in shaping AI guidance that met the community's needs. The strength of co-design in this context was in bringing together a diverse group of learners and educators to create guidance relevant to all members while providing insights into the AI literacies required for its successful implementation. Figure 1 provides an overview of the full co-design process.

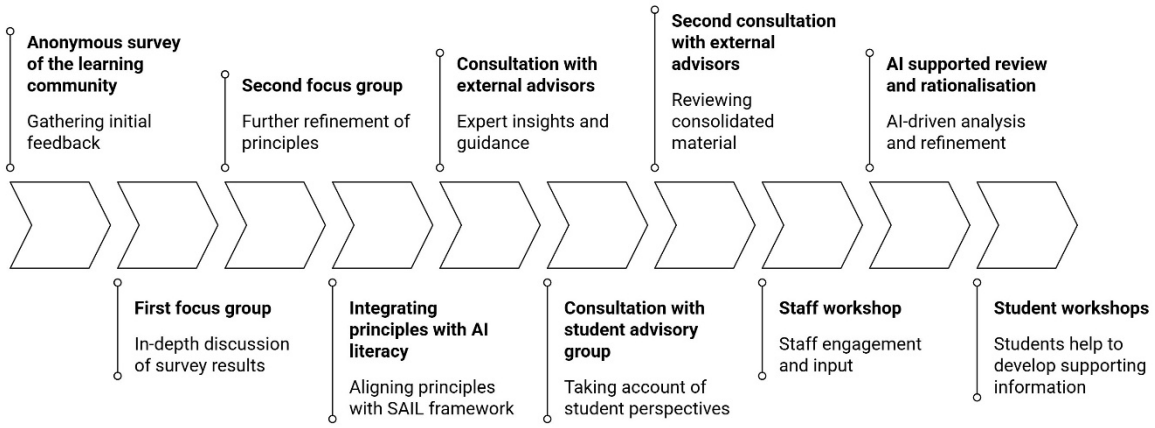


Figure 1: The co-design process

Table 1 shows the number of participants at each stage along with purposes and outputs. Although the number of survey responses (n=27) was small, the intent was not statistical generalisation but stakeholder-informed design. It also served as an invitation channel for focus group members. The two focus groups had the same members. The two student workshops were from different cohorts, mentored by the same five staff members.

Table 1: Participants, purposes, and outputs at each stage of the co-design process.

Stage	Participants	Purpose	Output
Survey	27	Baseline perceptions	Initial themes
Focus groups	12 (focus group 1) 12 (focus group 2)	Develop initial themes Expand the key ideas	Seven key ideas Ten principles, mapped to the SAIL framework
Stakeholders	21 (External stakeholders - two meetings) 8 (student advisory group)	Validate/refine	Eight refined principles, guideline development and structure of support materials Feedback on the level of detail and required support materials
Workshops	16 (Staff) 25 (Student group 1) + 5 staff mentors 27 (Student group 2) + 5 staff mentors	Develop final guidance	20 guidelines in five categories Explanations Example practices

4. Results

This section discusses the results of the co-design process, analysing the key findings under four themes.

4.1 Theme 1: Guidelines Emerge From Principles Rather Than Rules

Though we had expected to engage with creating guidelines early in the co-design process, the first focus group chose to engage in a more principles-based conversation. Rather than creating restrictions, they preferred to address the boundaries of AI use in learning, applying AI tools appropriately, and being transparent in AI usage. The group suggested that the goal should be to identify core principles that remain relevant across tools and contexts. From their discussion, we identified seven key ideas, some of which were principles-based (e.g., addressing ethical issues), while others were more practical and linked to theme 3.

> "*What are the key ideas about AI guidance that are essential?*" (from focus group 1 transcript)

The second focus group further developed the principles to include new ideas such as addressing digital divides and inclusion.

> "*Thinking, you know, a principles-based approach, rather than trying to come up with an exhaustive list to cover off every scenario*." (from focus group 2 transcript)

However, there were still some practical aspects in the proposed principles. These were factored out by the stakeholder groups, which created a set of 8 principles that provided the foundation for beginning to create specific guidelines. These principles stated that AI guidance should:

- Support AI literacy development
- Provide a common understanding of key concepts in AI
- Provide a strong foundation for the development of skills and competencies in AI
- Foster ethical awareness
- Embody cultural values
- Address digital divides and inclusion
- Preserve independent thinking
- Clearly define acceptable uses with practical examples

These principles were used as the structure for the guidelines that stakeholders, staff, and students began to expand with further explanations and example practices. During this process, it became apparent that the number of guidelines that had been proposed, at 27, was excessive. As workshop participants created further material, it came to light that there was significant overlap in the proposals, and we reduced the number of guidelines to 20, grouped in five categories derived from the original principles and AI literacies, but rationalised with a clearer focus.

4.2 Theme 2: AI Literacy Assists Learners in Interpreting Guidelines

Survey findings related to AI literacy suggested that, while many respondents had a broad grounding in key literacies, more development was needed in areas such as awareness and understanding of the potential of AI's broader capabilities and limitations, and some clearly needed further AI literacy development

> "*I feel very out of my depth with AI and often unsure*" (from survey responses)

Developing this theme, the first focus group suggested that AI literacy could be supported as part of the guideline support materials. The second focus group then began to map each principle to selected AI literacies from the SAIL framework to support coherence between the principles and AI literacy development.

> "*Yeah, I mean, that is one of the key literacies, isn't it, is knowing when not to use AI?*" (from focus group 2 transcript)

This pairing of principles and literacies proved an essential scaffold to the creation of guidelines, used by the stakeholder groups to start developing specific guidelines. Aligned to AI literacies, the principles provided a clearer structure for the participants to suggest guidelines that sat within the principles. In essence, the principles gave an overall structure to the process, but the alignment of multiple AI literacies to each principle gave a concrete set of ideas from which specific guidelines could be developed.

4.3 Theme 3: Practice Support is Required for Implementation

From the initial survey, it was clear that respondents wanted details about how to use AI appropriately in specific contexts and requested specific use cases to help them clarify how to use AI tools in practice, and how to report AI tool usage in their work. Similar ideas were explored in the first focus group, which agreed that guidance should support students with clear recommendations, strategies, and practical examples for using AI in their learning contexts, with exemplars for reflective practice. Stakeholder consultation provided more detailed ideas for what practice support would be required, and a range of possible support materials was suggested.

> "*Would it be useful to have a view of emerging uses to support [the] principles and/or provide examples throughout?*" (from external stakeholder transcript)

It became clear that each guideline would require further explanation, alongside examples of how to put that guideline into practice. Subsequently, the student advisory group emphasised that these materials should give students opportunities to develop their AI literacy through hands-on activities. So, the expectation that students would put guidelines into practice would be supported by specific tasks that they could perform to demonstrate their capabilities in following those guidelines.

> "*Progressing the development of specific use cases/exemplars (and how-to guides) so we can give students more specific guidance*." (from external stakeholder shared document comments)

At this point, the requirement was well established, and the workshops provided the opportunity for participants to record in detail what the additional explanations of each guideline would be and what example practices would be associated with them.

> "*e.g., If using Chat GPT, do a comparative analysis with other tools (Claude, Grok, etc.) and evaluate the differences in the output. Note gaps and limitations*." (from staff workshop notes)

4.4 Theme 4: Co-Design Revealed Limitations of Guideline-Centred Approaches

From the first focus group onwards, it became evident that a guidelines-first approach would not be an effective strategy. None of the participants felt able to suggest specific guidelines at this stage but wanted to focus on identifying core principles that remain relevant across tools and contexts. Rather than starting with restrictions, the group wanted to focus on ethical and educational values. This continued into the second focus group, where no specific guidelines were suggested.

> "*Principle-based rather than, you know, a list of X, Y, and Z, and I won't do this, and I will do that*." (from focus group 2 transcript)

This changed the direction of the subsequent stages, where the principles were refined further and aligned to specific AI literacies to provide the necessary scaffold for guidelines to be developed. This was a clear finding of our study, that a principles-first approach is needed, followed by alignment with AI literacies, with guideline development following once these are in place.

4.5 Results Summary

Across all themes, participants consistently identified a progression from principles to literacies and finally to guidelines supported by practical examples. These relationships informed the simple process framework shown in Figure 2. Principles, aligned with AI literacies, provided a frame within which guidelines could be created, which, like all scaffolds, began to fall away as we reached the latter part of the process, distilled into categories and guidelines.

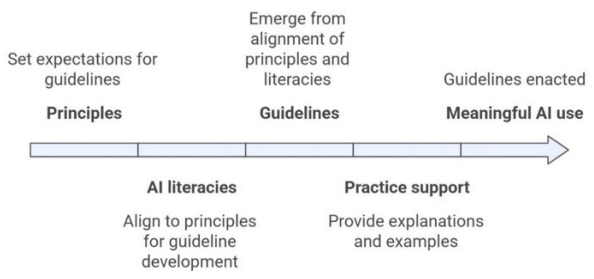


Figure 2: Framework of the co-design process

4.6 The Guidelines

The final set of 20 guidelines is organised into the following five categories:

Category 1: Understanding AI in Your Learning Context - Foundational literacy about what AI is and how it is used

Category 2: Using AI Purposefully in Learning - Practical, intentional, academically aligned use of AI

Category 3: Safe and Transparent Use of AI - Using AI with integrity, privacy, and safety

Category 4: Ethical Use of AI - Avoiding risks, misconduct, and negative impacts

Category 5: AI, Cultural Awareness, and Inclusion - Cultural perspectives, data sovereignty, and equity

As previously discussed, the co-design process revealed that the guidelines alone are not enough to ensure they are followed. As a result, each guideline in the material published to students consists of three parts:

- Guideline: a statement of the guideline itself
- Further Explanation: an explanation of what issues the guideline is addressing and why it matters
- Example Practice for Students: a concrete suggestion for how the guideline can be met/demonstrated

The full set of guidelines is listed below. The key focus of each guideline is highlighted in bold text. The further explanations and example practices for students are not included here.

Category 1. Understanding AI in Your Learning Context

1. Familiarise yourself with the **capabilities, strengths, and limitations of different AI tools**, and consider how these can affect their outputs and your learning.
2. Understand **how AI is defined and used in your learning context**, and stay informed about good practices shared by staff and students.
3. Stay informed about **how AI is affecting organisations and industries** related to your studies and reflect this understanding in your work.
4. Recognise that **AI outputs can vary and may be biased, inconsistent or inaccurate**, and apply critical judgment when interpreting or using them in your academic work.

Category 2. Using AI Purposefully in Learning

1. Use AI as one component of your overall digital literacy, and **consider whether AI is the best choice for a task** before using it for learning or assessment.
2. Decide where AI can add value to your learning goals and **choose tools that align with the requirements of your learning activity** or assessment.
3. When using AI for research, **verify the accuracy and credibility of any information AI** provides by triangulating its outputs with primary sources.
4. Review AI-supported work carefully to ensure the final product **reflects your own independent thinking, voice, and perspective**.

Category 3. Safe and Transparent Use of AI

1 **Follow any programme-specific guidance** or staff instructions on acceptable AI use provided in addition to these guidelines
2 Acknowledge and **describe your use of AI transparently and accurately**, citing the sources of any AI-created content, and how AI tools have been used.
3 Identify the privacy and security policies and risks of AI tools and **use safe practices** to protect yourself and your data.
4 Take opportunities to **provide reliable source materials** for AI tools to work from when it is safe, appropriate, and permitted to do so.

Category 4. Ethical Use of AI

1 **Check AI-generated content for possible bias** and seek to mitigate this through your own critical use.
2 **Never upload personal identifiable information** or commercially sensitive data from your research or workplace into open/public AI models.
3 **Avoid using AI to plagiarise**, mislead, or cause harm.

4 **Avoid unnecessary or excessive use of AI** resources to minimise environmental impacts.

Category 5. AI, Cultural Awareness, and Inclusion

1. Recognise that the cultural assumptions and value systems built into **AI systems may omit or misrepresent Indigenous knowledge** and verify and supplement such content with appropriate sources.
2. **Respect data sovereignty** and avoid uploading culturally sensitive or restricted material into AI systems.
3. Prompt AI tools to **explicitly consider cultural perspectives** and content in their responses.
4. **Use AI in ways that support inclusivity** and avoid tool or data use that might disadvantage others

5. Discussion

This study set out to co-design AI guidelines for a specific higher education context. However, a central insight was not simply the development of guidelines, but a clearer understanding of the conditions required for such guidance to be meaningful and usable in practice.

Our initial expectation was that co-design would lead directly to the development of guidelines. It quickly became evident that guidelines are difficult to construct in contexts where both understanding and practice are still developing. This suggests that guideline development in emerging technological contexts like AI may require a shift from a guideline-first approach to more principle-based approaches.

This aligns with critiques of policy-first approaches in digital contexts. Policy often assumes stable practices and shared understanding, but in rapidly evolving technological environments, it can be challenging for educators to take control of change (Selwyn, 2024). Instead, learners and educators operate within a "capability gap", where expectations of appropriate behaviour outpace individuals' ability to respond (Wang et al, 2026, p. 140). The co-design process made this tension visible. While AI literacies were intentionally integrated, they initially acted more as a scaffold for discussion than a direct pathway to guideline development. Participants frequently moved away from specifics and instead engaged in broader conversations about learning, ethics, and practice. Although these discussions did not immediately produce guidelines, they did surface how AI was being used within the institution, providing the necessary foundation from which guidance could later be derived.

A key insight emerging from this process is that guidelines are inherently interpretive rather than prescriptive. Participants emphasised the need for flexibility across contexts, tools, and disciplines, reinforcing that technology is inherently "unruly" and cannot be fully defined by a set of rules. Instead, effective engagement with it depends on social learning (Wynne, 1988). This interpretive nature places significant demands on learners and leads to a central finding of this study: AI guidelines are only actionable when accompanied by sufficiently supported AI literacy to interpret and apply them.

Across all stages of the co-design process, participants consistently identified the need for supporting materials, examples, and opportunities for practice, reinforcing that guidelines alone are insufficient. This emphasises the knowledge, critical thinking, and ethical awareness required to navigate complex digital environments (Miao & Shiohira, 2024).

Importantly, this insight emerged through the co-design process itself. Participants were able to articulate challenges and expectations but struggled to translate these into formalised guidelines without first engaging in discussions about capability, understanding, and practice. This highlights a key methodological contribution: co-design does not simply generate better guidelines; it reveals the limitations of guideline-centric approaches. In this case, it exposed that guidelines, when developed in isolation, are structurally incomplete.

As a result, this study emphasises the importance of a capability-informed approach, which can be conceptualised as a three-part structure:

- Guidelines – articulating principles, expectations, values, and boundaries
- AI literacy – enabling learners to interpret and apply these guidelines
- Practice-based supports – providing examples, activities, and opportunities for applied practice

This integrated model aligns with emerging perspectives in AI literacy, which emphasise not only understanding how AI works, but also developing the capacity to critically evaluate it, and use it responsibly (MacCallum, Parsons & Mohaghegh, 2026). The guidelines developed from this study - comprising a guideline statement,

explanation, and example of practice - reflect this insight. They acknowledge that understanding cannot be assumed, and that guidance must be accompanied by mechanisms that support interpretation and application.

Finally, this study highlights the value of co-design as a process for navigating emerging educational challenges. Co-designing AI guidelines is ineffective when constrained by a narrow, predefined focus. Instead, it requires openness to divergent, iterative, and sometimes messy conversations, where stakeholders explore not only what guidance should look like, but what is required to make it meaningful. In this process, researchers must follow the directions of stakeholder inquiry and translate diverse forms of input into coherent outputs, drawing on professional judgement to bridge the gap between discussion and design.

6. Conclusion, Limitations and Future Work

6.1 Limitations of the Study

The study is based in a single institution, engaging a small subset of the members of the stakeholder groups that may be impacted by changes to the institution's AI guidelines. Additionally, those stakeholders engaged in the co-design were self-selecting. It is possible that the reflections on co-design discussed in the paper are not representative of other learning communities inside or outside the institution. However, it is suggested that the processes and principles applied in this study may be informative for other institutions engaging in similar co-design processes with their own stakeholder groups. While the outcomes may be different, the process and underlying principles may be similar.

6.2 Future Work

The creation and application of AI guidelines is seen as an ongoing process that will need to be continually revisited as the features, capabilities, and challenges of AI tools continue to evolve. In addition, while guidelines support learners and staff, policies are also required by institutions for a range of reasons, but ultimately to manage risk, to ensure that AI is trustworthy and used for good (Ally & Mishra, 2024). Therefore, policy needs to be developed from the same principles as the guidelines.

Ethics Declaration

The research design was submitted to the academyEX institutional ethics committee and received approval on 17th April 2025 (approval number aex.staff.2025.02).

AI Declaration

Napkin was used to create the two figures. ChatGPT was used to provide a critical review of the draft article. After author revisions, it provided a further critical review that was used to help refine the submitted version.

References

Ally, M., and Mishra, S. (2024). "Policies for Artificial Intelligence in Higher Education: A Call for Action", *Canadian Journal of Learning and Technology*, Vol 50, No. 3. https://cjlt.ca/index.php/cjlt/article/view/28869/21191

Ambrosetti, É., Gaudin, C., Flandin, S., and Poizat G. (2025). "Students as co-designers in health professional education: a scoping review", *BMC Medical Education*, Vol 25, No. 645. https://doi.org/10.1186/s12909-025-07110-0

Baek, C., Tate, T., and Warschauer, M. (2024). ""ChatGPT seems too good to be true": College students' use and perceptions of generative AI", *Computers and Education: Artificial Intelligence*, Vol 7, No. 100294, https://doi.org/10.1016/j.caeai.2024.100294.

Brandon, E., Eaton, L., Gavin, D., and Papini, A. (2025). "In the Room Where It Happens: Generative AI Policy Creation in Higher Education", [online], EDUCAUSE Review. https://er.educause.edu/articles/2025/5/in-the-room-where-it-happens-generative-ai-policy-creation-in-higher-education

Cardona, M., Rodríguez, R., and Ishmael, K. (2023). "Artificial Intelligence and the Future of Teaching and Learning", [online], U.S. Department of Education. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Fadel, C., Taylor, R., and Black, A. (2024). "Interdisciplinary Themes: Embedding Contemporary Contexts within Content", [online], Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/Interdisciplinary-Themes.pdf

Fielding, M. (2001). "Students as Radical Agents of Change". *Journal of Educational Change*, Vol 2, pp 123–141. https://doi.org/10.1023/A:1017949213447

Fu, Y., and Weng, Z. (2024). "Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices", *Computers and Education: Artificial Intelligence*, Vol 7, No. 100306.

https://doi.org/10.1016/j.caeai.2024.100306.
Funa, A. A., and Gabay, R. A. E. (2025). "Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study", *Social Sciences & Humanities Open*, Vol 11, No. 101221. https://doi.org/10.1016/j.ssaho.2024.101221
Ghimire, A., and Edwards, J. (2024, July). "From guidelines to governance: A study of AI policies in education", In Proceedings International Conference on Artificial Intelligence in Education (pp 299-307). Springer Nature. https://doi.org/10.1007/978-3-031-64312-5_36
Iniesto, F., Littlejohn, A., and Charitonos, K. (2022). "A review of research with co-design methods in health education", *Open Education Studies,* Vol 4, pp 273–295. https://doi.org/10.1515/edu-2022-0017
Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). "Generative AI in higher education: A global perspective of institutional adoption policies and guidelines", *Computers and Education: Artificial Intelligence*, Vol 8, No. 100348. https://doi.org/10.1016/j.caeai.2024.100348
Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., and Huber, P. (2016). "Artificial intelligence and computer science in education: From kindergarten to university". IEEE frontiers in education conference. https://ieeexplore.ieee.org/abstract/document/7757570.
Küçükuncular, A. (2025). "Learning with, rather than through, AI: co-designing science education for critical AI literacy", *Frontiers in Education*, Vol 2025, No. 1716353. https://doi.org/10.3389/feduc.2025.1716353
Long, D., and Magerko, B. (2020). "What is AI literacy? Competencies and design considerations." In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1–16). https://doi.org/10.1145/3313831.3376727
MacCallum, K., Parsons, D., and Mohaghegh, M. (2024). "The Scaffolded AI Literacy (SAIL) Framework for Education: Preparing learners at all levels to engage constructively with Artificial Intelligence". *He Rourou*, Vol 1 No 1. https://doi.org/10.54474/herourou.v1i1.10835
MacCallum, K., Parsons, D., and Mohaghegh, M. (2026). "The Scaffolded AI literacy (SAIL) framework: Results of a Delphi study for equitable AI literacy framework design in education", *Computers and Education: Artificial Intelligence*, Vol 10, No. 100584. https://doi.org/10.1016/j.caeai.2026.100584
Miao, F., and Cukurova, M. (2024). "AI competency framework for teachers". [online] UNESCO. https://doi.org/10.54675/ZJTE2084
Miao, F., and Shiohira, K. (2024). "AI competency framework for students", [online], UNESCO. https://doi.org/10.54675/JKJB9835
OECD / European Union (2026). "Empowering learners for the age of AI: An AI literacy framework for primary and secondary education". [online], Paris: OECD Publishing. https://doi.org/10.1787/65cd27d4-en
Ohio State University. (n.d.). "AI Considerations for Teaching and Learning", [online], Ohio State University. https://teaching.resources.osu.edu/teaching-topics/ai-considerations-teaching-learning
Roschelle, J., Penuel, W., and Shechtman, N. (2006). "Co-design of Innovations with Teachers: Definition and Dynamics". In Barab, S. A., Hay, K. E., and Hickey, D. T. (Eds.), The International Conference of the Learning Sciences: Indiana University 2006. Proceedings of ICLS 2006, Volume 2 (pp 606-612). Bloomington, Indiana, USA: International Society of the Learning Sciences. https://dl.acm.org/doi/10.5555/1150034.1150122
Sabatini, J., Graesser, A. C., Hollander, J., and O'Reilly, T. (2023). "A framework of literacy development and how AI can transform theory and practice", *British Journal of Educational Technology*, Vol 54, No. 5, pp 1174-1203. https://doi.org/10.1111/bjet.13342
Selwyn, N. (2024). "On the Limits of Artificial Intelligence (AI) in Education", *Nordic Journal of Pedagogy and Criticism*, Vol 10, pp 3-14. https://pedagogikkogkritikk.no/index.php/ntpk/article/view/6062/9573
Smith, A., McConnell, L., Iyer, P., Allman-Farinelli, M., and Chen, J. (2024). "Co-designing assessment tasks with students in tertiary education: a scoping review of the literature", *Assessment & Evaluation in Higher Education*, Vol 50, No. 2, pp 199–218. https://doi.org/10.1080/02602938.2024.2376648
Smith, F., Sadek, M., Wan, E., Ito, A., and Mougenot, C. (2024). "Codesigning AI with End-Users: An AI Literacy Toolkit for Nontechnical Audiences", *Interacting with Computers*, Vol iwae029, https://doi.org/10.1093/iwc/iwae029
Wang, S., Alam, G. M., Skrbinjek, V., Bashir, K., and Lei, M. (2026). "Managing digital bridges: how online social capital translates digital literacy into higher education access", *Human Systems Management*, Vol 45, No. 2, pp 139-154. https://doi.org/10.1177/01672533251359152
Wynne, B. (1988). "Unruly Technology: Practical Rules, Impractical Discourses and Public Understanding", *Social Studies of Science*, Vol 18, pp 147 - 167. https://doi.org/10.1177/030631288018001006
Zamenopoulos, T., and Alexiou, K. (2018). "Co-design as collaborative research", [online], Bristol University/AHRC Connected Communities Programme. https://oro.open.ac.uk/58301/
Zelenko, O., Gomez, R., and Kelly, N. (2021). "Research Co-Design: Meaningful Collaboration in Research", In A. Blackler & E. Miller (Eds.), *How To Be A Design Academic: From Learning to Leading* (pp 227–244). CRC Press. https://doi.org/10.1201/9780429351693-20.

The Role of Academic Libraries in E-Learning Ecosystems in Higher Education: A Narrative Review

Edmont Pasipamire
Rosebank International, South Africa

epasipamire@rbi.ac.za

Abstract: As higher education continues its shift toward digitally mediated instruction, academic libraries have undergone a profound transformation, moving from traditional custodians of print collections to dynamic partners in e-learning. This narrative review explores the evolving contributions of academic libraries to the development and sustainability of online learning ecosystems, guided by Connectivist Learning Theory (Siemens, 2005). Drawing on peer-reviewed literature, institutional case studies, and policy documents published between 2013 and 2024, the review synthesizes emerging practices and conceptual frameworks that define the library's role in digital education. Findings reveal that libraries now play multifaceted roles: curating digital resources, embedding services into Learning Management Systems (LMS), delivering virtual information literacy instruction, and collaborating with faculty on course design. Libraries also lead in initiatives such as Open Educational Resources (OER), personalized learning support, and digital scholarship. However, persistent challenges including inadequate digital infrastructure, limited staff capacity, and marginalization in institutional strategy constrain broader impact. By organizing these developments thematically and theoretically, this review offers a comprehensive lens on how libraries are reshaping their identities within higher education. It reinforces the need for strategic investment, cross-departmental integration, and further research to fully leverage libraries as critical enablers of inclusive and sustainable digital learning.

Keywords: Academic Libraries, E-Learning, Digital Pedagogy, Higher Education, Narrative Review, Learning Ecosystems

1. Introduction and Background

Academic libraries have evolved from custodians of physical collections into dynamic facilitators of digital learning. The growth of e-learning has expanded their roles, positioning libraries as active partners in educational delivery, digital innovation, and student success (Steele, 2019; Brenton & Tury, 2021). A key aspect of this transformation is digital resource curation. Libraries provide e-books, scholarly databases, multimedia resources, research guides, and course-integrated materials to support virtual learning (Steele, 2019; Oyedokun, 2024). Emerging technologies, including artificial intelligence, virtual reality, and learning analytics, further enhance personalized and interactive services (Onunka et al., 2023; Bihan, 2022). Virtual reference services, including 24/7 chat, video conferencing, and asynchronous support, enable equitable research assistance for geographically dispersed learners (Steele, 2019; La Macchia, 2020). Similarly, information literacy instruction has shifted to online tutorials, interactive modules, and virtual workshops embedded in Learning Management Systems. These services require librarians to combine technological and pedagogical expertise while developing students' critical research and evaluation skills (Garner et al., 2024; Vogus, 2020). Collaboration with academic staff has also strengthened, with librarians contributing to curriculum development and co-creating digital learning materials such as video tutorials and modular resources (Owens et al., 2024; Makwana & Gadhavi, 2023; Sanches, 2018). Overall, libraries have become collaborative and pedagogically aligned partners in e-learning, supporting more responsive, inclusive, and digitally oriented education (Farmer, 2019).

1.1 Problem Statement

Despite the rapid expansion of e-learning in higher education and the transformation of academic libraries from custodians of print collections into active partners in digital pedagogy (Eyre et al., 2017), the literature examining this transformation remains fragmented across library science, educational technology, and higher education research. Studies identify various library functions, including digital resource curation, virtual reference, embedded librarianship, LMS integration, and support for Open Educational Resources (Abhinandan, 2026), but rarely synthesise these contributions into a coherent understanding of how academic libraries support e-learning ecosystems. Consequently, the contribution of academic libraries to the development and sustainability of e-learning ecosystems remains fragmented and insufficiently understood, despite their expanding pedagogical, technological, and information-support roles. This problem is further compounded by inconsistent theoretical grounding and considerable variation in institutional contexts, with research often examining individual library initiatives or cases rather than synthesising their broader contribution to e-learning (Bah et al., 2024). The lack of an integrated understanding makes it difficult for higher education institutions to recognise how academic libraries contribute to the development, continuity, and sustainability of e-learning environments. This narrative

literature review therefore synthesises existing research to examine the ways in which academic libraries have contributed to the development and sustainability of e-learning ecosystems in higher education, while identifying the key roles and challenges emerging from the literature.

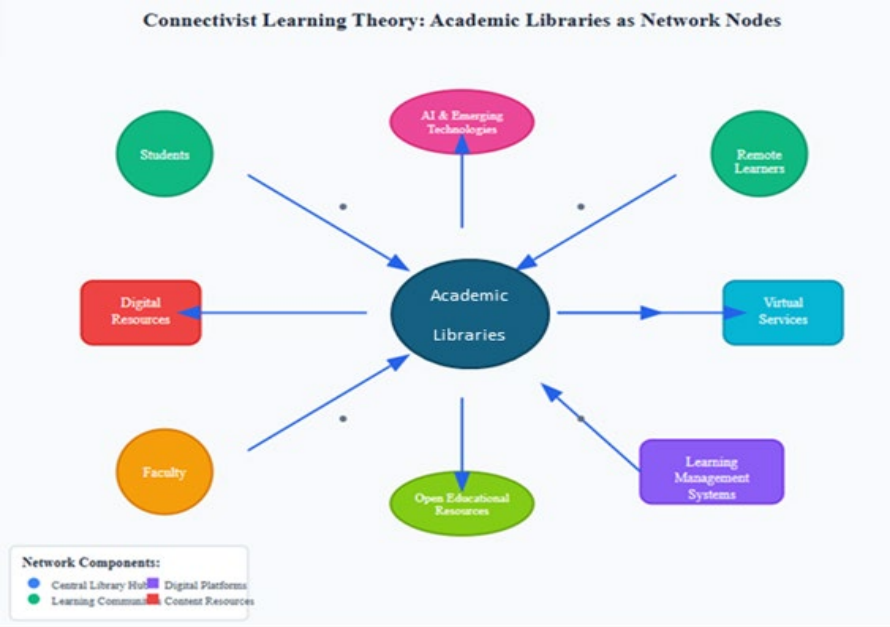
1.2 Research Question and Objectives

This narrative review is guided by the central research question: *In what ways have academic libraries contributed to the development and sustainability of e-learning ecosystems in higher education?* To address this question comprehensively, the review pursues several specific objectives:

- To identify and categorize the various roles that academic libraries have assumed within e-learning environments
- To explore and analyse the theoretical perspectives that underpin the role of libraries within higher education ecosystems
- To explore how academic libraries are integrated into higher education e-learning systems
- To identify and analyse the challenges that shape and limit library contributions to e-learning in higher education
- To present emerging trends and innovative practices in library-supported digital education.

1.3 Theoretical Framework

This review is grounded in Connectivist Learning Theory, proposed by Siemens (2005), which posits that learning occurs through the formation and navigation of networks comprising individuals, digital content, and technological platforms. Within this framework, academic libraries function as critical nodes in the e-learning ecosystem. They facilitate access to distributed knowledge, support the development of digital literacy skills, and enable learners to engage meaningfully with information in a connected environment. Connectivism emphasizes the learner's ability to identify, connect, and traverse diverse sources of knowledge. Academic libraries play a central role in this process by offering curated digital resources, guiding students in evaluating information, and fostering the competencies needed to thrive in digitally mediated learning spaces. This theoretical lens positions libraries not merely as information providers, but as integral facilitators of learning within dynamic, networked educational systems. Figure 1 below illustrates how academic libraries act as central nodes in the e-learning ecosystem, aligning with the core principles of Connectivist Learning Theory.



Source: Author, 2026

Figure 1: This illustrates Connectivist Learning Theory, with the academic library serving as a central network node connecting students, faculty, digital resources, virtual services, LMS platforms, OER and emerging technologies. It highlights the library's role in connecting learners to information, technology and learning support within the e-learning environment.

1.4 Methodology

This study employed a narrative literature review within an interpretivist paradigm to examine academic libraries' evolving contribution to e-learning in higher education. The approach supported the synthesis of diverse theoretical and empirical perspectives across library science, educational technology, and digital pedagogy (Green et al., 2006; Snyder, 2019). The review included peer-reviewed English-language journal articles published between 2015 and 2024 on academic libraries, e-learning, digital literacy, information literacy, and higher education. Relevant studies were thematically analysed through open coding, grouping similar concepts into broader categories. Iterative refinement produced five themes addressing library roles, theoretical foundations, integration practices, and challenges, which structured Section 2.

2. Overview of Literature

2.1 The Evolving Roles of Academic Libraries in E-Learning Ecosystems

As introduced in Section 1, academic libraries have moved from custodians of physical collections to active facilitators of digital learning, expanding into digital resource curation, virtual reference, information literacy instruction, and collaboration with academic staff (Steele, 2019; Brenton & Tury, 2021; Oyedokun, 2024; Onunka et al., 2023; Bihan, 2022; La Macchia, 2020; Garner et al., 2024; Vogus, 2020; Owens et al., 2024; Makwana & Gadhavi, 2023; Sanches, 2018; Farmer, 2019). These developments position libraries as increasingly collaborative and pedagogically integrated partners within higher education.

2.2 The Role of Libraries in Higher Education Ecosystems: Theoretical Perspectives

Academic libraries contribute to e-learning through several theoretical perspectives. Connectivist Learning Theory positions libraries as nodes in digital learning networks that provide curated information and support digital literacy and information navigation (Siemens, 2005; Detterbeck & Sciangula, 2017; Duke et al., 2013). Communities of Practice Theory views libraries as collaborative spaces that facilitate knowledge sharing and co-creation (Transue, 2013; Kim, 2015). From an Instructional Design perspective, libraries support teaching through digital learning objects, LMS integration, and collaboration with faculty to align information literacy with course objectives (Fonseca et al., 2024; Yukawa, 2010). Information Behaviour Theories, including TRA, TAM and TPB, highlight the importance of users' attitudes, intentions and behaviours in engaging with digital resources, informing the development of responsive, user-centred services (Efird et al., 2024; Salim et al., 2016; Kwanya, 2016). Collectively, these perspectives show how libraries support networked learning, collaboration, pedagogical integration and user-centred services in e-learning environments. Figure 2 integrates these theoretical perspectives and illustrates the multifaceted role of academic libraries in digital learning ecosystems.



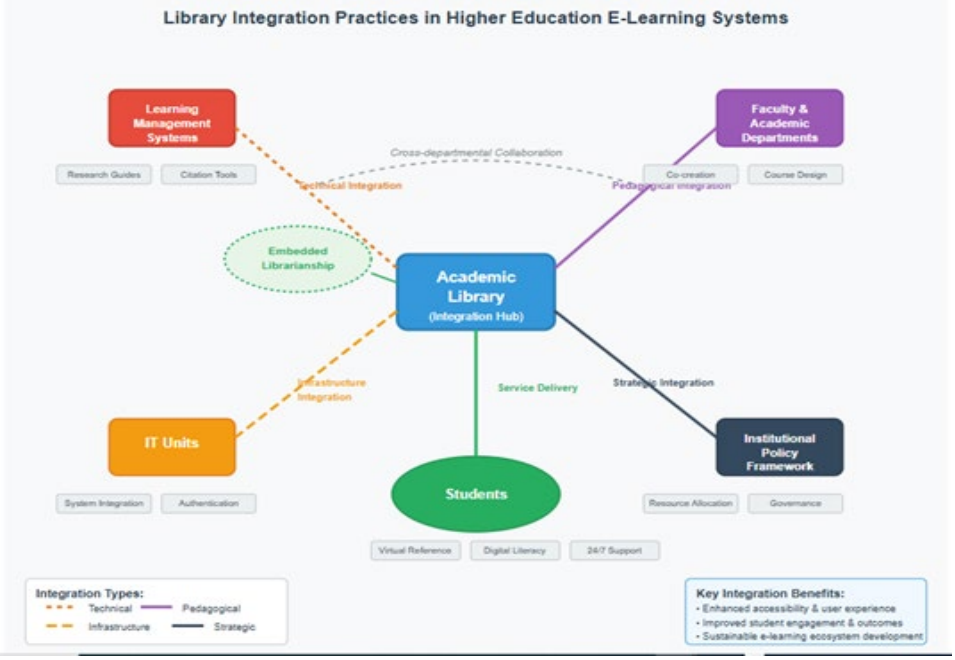
Source: Author, 2026

Figure 2: Illustrates how the four theoretical perspectives complement rather than compete with one another. Connectivism and Communities of Practice emphasise networked and collaborative learning, while

Instructional Design and Information Behaviour highlight applied, user-centred services. Together, they capture the library's pedagogical, technical, and social contributions to e-learning ecosystems.

2.3 Library Integration Practices in Higher Education E-Learning Systems

Academic libraries are increasingly integrated into e-learning through Learning Management Systems (LMS), embedded services, and collaboration with academic and technical staff. LMS integration allows research guides, citation tools, and course-specific resources to be placed directly within online courses, providing students with timely access to information without moving between platforms (Nel, 2020; Karplus, 2017). This approach makes library support more accessible and connects information-literacy development with students' immediate learning activities. Embedded librarianship further strengthens this integration by positioning librarians as partners in online teaching and learning. Librarians may contribute to course design, provide targeted research support, and participate in online learning activities, thereby extending library expertise into the curriculum (Kumar et al., 2024). Such integration also depends on collaboration with academic staff and IT units to align library services with curricular and technological requirements (Arroyo-Vázquez & Gómez-Hernández, 2020; Jank, 2013). The literature therefore suggests that effective e-learning integration requires libraries to work across institutional boundaries rather than operate as separate support services. At the institutional level, integrating library strategies into policies and planning can strengthen the sustainability of these practices by supporting coordinated governance, resource allocation, and service delivery (Kumar et al., 2024). However, limited awareness of library services among students and faculty remains a constraint, highlighting the need for ongoing training, evaluation, and outreach (Digital Collaborative Governance, 2022). Overall, the literature indicates a shift from standalone library services towards embedded, collaborative, and strategically aligned participation in e-learning. Figure 3 illustrates these integration practices and their relationship within higher education e-learning systems.



Source: Author, 2026

Figure 3: Positions the academic library as an integration hub linking four interdependent domains: Learning Management Systems (technical), faculty and academic departments (pedagogical), IT units (infrastructure), and institutional policy frameworks (strategic). The cross-departmental connections highlight that effective e-learning integration requires institution-wide coordination, governance, resource allocation, and aligned service delivery, as discussed in Section 2.3.

2.4 Challenges Framing Library Contributions to E-Learning

Academic libraries' contributions to e-learning are constrained by technological, financial, human-resource, institutional, and ethical challenges. These barriers shape the extent to which libraries can provide accessible

and integrated digital learning support. Legacy systems, inadequate network infrastructure, complex authentication, and poor mobile optimization limit seamless access to digital resources (Hu & Raman, 2024; Brenton & Tury, 2021). Rising digital-resource costs and inadequate funding for infrastructure and staff development further restrict institutional capacity (Satvik & Naik, 2024; Hu & Raman, 2024). These challenges are compounded by the digital divide, particularly among students with limited connectivity and devices (McAllister, 2018; Pettalongi et al., 2024). Effective digital transformation requires librarians with technological and pedagogical competencies, yet recruitment, retention, and competing traditional and digital responsibilities can contribute to professional fatigue (Frank et al., 2021; Barthorpe & Wright, 2017). Faculty resistance, departmental silos, and limited recognition of libraries' pedagogical value can also restrict collaboration and curriculum integration (Pathak, 2023; Lalitha et al., 2024; Din & Ali, 2024). AI, immersive technologies, and personalized learning tools create additional concerns regarding privacy, depersonalization, infrastructure requirements, and equitable access (Idowu, 2024; Lo & Vitale, 2024; Priya & Ramya, 2024). Therefore, technological advancement must be accompanied by ethical governance and attention to equitable access. Overall, these challenges do not diminish the importance of academic libraries; rather, they demonstrate that their contribution to e-learning depends on institutional capacity, technological readiness, staff competencies, collaboration, and responsible innovation.

2.5 Innovations and Emerging Trends in Library-Supported Digital Education

Library-supported digital education is increasingly personalised, interactive, and learner-centred, with AI, immersive technologies, gamification, OER, digital scholarship, personalised learning, and embedded librarianship reshaping library support. AI enables 24/7 reference assistance, personalised recommendations, and course-aligned content analysis (Fitria, 2023; Shankar et al., 2023), while VR and AR support experiential learning through virtual environments and simulations (Sarkar, 2023; Pope, 2018). Gamification enhances engagement in research and information literacy (Pope, 2018), while OER strengthens affordability and access through open-material development and adoption (Katz, 2019; Goodsett et al., 2015; Cassidy et al., 2016; Henke et al., 2024). Digital scholarship and learning analytics provide specialised tools, training, publishing platforms, and personalised support (Gisip et al., 2024; Giannaros et al., 2020; Priya & Ramya, 2024), including adaptive sequencing, learner profiling, and flexible pacing (Çullhaj, 2017). Embedded librarianship integrates librarians into online courses and LMSs to provide research and information-literacy support (Bonnand & Hansen, 2012; Tumbleson, 2016). Collectively, these developments position libraries as active partners in digital education, although sustainability requires investment, institutional alignment, inclusivity, and ethical technology use.

3. Discussion and Interpretation of Findings

The findings indicate that academic libraries are shifting from traditional resource providers to active participants in e-learning ecosystems. Their roles now extend to digital resource curation, virtual reference, information-literacy support, LMS integration, embedded librarianship, collaboration with academics, OER and emerging technologies. This suggests that effective e-learning depends not only on technological platforms but also on access to reliable information and professional expertise. A key interpretation is the shift from custodianship to collaboration and pedagogical engagement. Librarians increasingly support course delivery through online consultations, information-literacy instruction and embedded services (Steele, 2019; Oyedokun, 2024; Tumbleson, 2016). However, the depth of integration depends on institutional structures, staff capacity and recognition of library roles. Similarly, virtual reference, tutorials and LMS-embedded resources extend library support beyond physical spaces, but connectivity, devices and digital skills can limit equitable access (La Macchia, 2020; Vogus, 2020; Garner et al., 2024; Frank et al., 2021; Pettalongi et al., 2024). The findings also suggest that libraries should be viewed as components of institutional e-learning infrastructure, rather than peripheral support services. LMS integration, collaboration with IT units and involvement in curriculum development position libraries as partners in teaching and learning (Arroyo-Vázquez & Gómez-Hernández, 2020; Karplus, 2017). Nevertheless, departmental silos, faculty resistance and limited institutional recognition continue to constrain this role (Pathak, 2023; Lalitha et al., 2024; Din & Ali, 2024). Emerging technologies such as artificial intelligence, virtual and augmented reality, learning analytics and gamification provide opportunities to enhance library-supported learning. However, their value depends on educational relevance and responsible implementation. Concerns surrounding privacy, equity, infrastructure and staff capacity indicate that technology adoption should complement, rather than replace, human-centred library support (Fitria, 2023; Lo & Vitale, 2024; Priya & Ramya, 2024). These findings align with Connectivist Learning Theory, which views learning as occurring through networks of people, information and technologies (Siemens, 2005). Libraries function as

important nodes within these networks by connecting learners to curated information, digital resources and learning support. Collaboration among librarians, academics and learners further reflects the principles of Communities of Practice in digital learning environments (Ghosh et al., 2024; Transue, 2013; Kim, 2015; Yukawa, 2010). Overall, the findings position academic libraries as increasingly interconnected partners in e-learning, whose future effectiveness depends on institutional integration, adequate resources and strategic positioning within e-learning planning.

3.1 Limitations of Current Research

Research on library contributions to e-learning has several limitations. First, there is a geographic and institutional bias, as much of the existing literature originates from high-income countries, leaving limited evidence from the Global South (Din & Ali, 2024). Second, there is a lack of longitudinal research, with limited evidence on the sustained impact of library innovations on student outcomes over time (Gisip et al., 2024). Finally, inconsistent evaluation measures make it difficult to assess library contributions systematically, as many studies rely on descriptive or case-based evidence rather than standardised performance indicators (Frank et al., 2021).

4. Conclusions and Recommendations

This narrative review examined how academic libraries contribute to the development and sustainability of e-learning ecosystems in higher education. The literature indicates that academic libraries have moved beyond their traditional role as custodians of physical collections to become active partners in digitally mediated teaching and learning. Their contributions include curating digital resources, providing virtual reference and information-literacy services, integrating library resources into Learning Management Systems (LMS), supporting embedded librarianship, and collaborating with academic staff in course and curriculum development. The review further shows that libraries contribute to e-learning through their capacity to connect learners with digital knowledge, develop information and digital literacy, facilitate collaboration, and support learner-centred services. These contributions are consistent with Connectivist Learning Theory, which positions libraries as important nodes within networks of learners, information and digital technologies. Emerging practices, including artificial intelligence, virtual and augmented reality, Open Educational Resources (OER), gamification, learning analytics and personalised learning, are also expanding the potential of libraries to support more interactive and accessible digital learning environments. However, the sustainability of these contributions is dependent on the conditions within which academic libraries operate. The literature identifies technological and financial constraints, limited staff capacity, inadequate professional development, departmental silos, weak institutional recognition and inequalities in students' access to digital infrastructure as significant challenges. Therefore, strengthening the role of academic libraries in e-learning requires institutional support, strategic integration, adequate resources and continued professional development. Overall, the review suggests that academic libraries are increasingly important components of higher education e-learning ecosystems, but their potential can only be fully realised when their contributions are recognised and supported at institutional level.

4.1 Recommendations for Practice

- Strengthen embedded librarianship and LMS integration: Higher education institutions should systematically integrate librarians into online courses and Learning Management Systems to provide contextualised information-literacy instruction, research support and digital learning resources (Tumbleson, 2016; Bonnand & Hansen, 2012).
- Invest in digital infrastructure and staff development: Institutions should provide adequate technological infrastructure, resources and continuous professional development to enable librarians to deliver sustainable digital services and respond effectively to changing e-learning demands (Frank et al., 2021; Steele, 2019).
- Promote responsible adoption of emerging technologies and OER: Libraries should expand their support for OER and carefully evaluate emerging technologies such as AI, VR/AR and learning analytics, while addressing issues of accessibility, privacy, equity and ethical use (Goodsett et al., 2015; Katz, 2019; Fitria, 2023; Priya & Ramya, 2024).

4.2 Areas for Further Research

- Longitudinal evaluation of library-supported e-learning: Future research should examine the long-term relationship between library integration into e-learning environments and outcomes such as student engagement, academic performance, retention and information-literacy development.
- Comparative research across geographical and institutional contexts: Further studies should investigate library contributions to e-learning in the Global South and compare these with practices in the Global North, particularly in relation to infrastructure, funding, digital access and institutional support.
- Ethical and user-centred evaluation of emerging technologies: Future research should examine students' and staff members' perceptions of AI, learning analytics and immersive technologies in library services, with particular attention to privacy, equitable access, trust and the role of human support.

Ethics Declaration

This study did not require ethical clearance, as it did not involve human participants or personal data. All sources used were publicly available.

AI Declaration

AI tools were used in a limited capacity to support language refinement. The intellectual content, analysis, and conclusions are the sole responsibility of the author.

References

Abhinandan, M. (2026). Role of academic libraries in lifelong learning beyond formal education. *Shodh Manjusha: An International Multidisciplinary Journal*. https://doi.org/10.70388/sm250197

Arroyo-Vázquez, N., & Gómez-Hernández, J.-A. (2020). La biblioteca integrada en la enseñanza universitaria online: situación en España // Integrating libraries into online university learning: the current situation in Spain. http://eprints.rclis.org/40141/

Bah, U., Mustafa, A., & Zool, M. (2024). From books to bytes: A systematic review of the role of information and communication technology (ICT) in academic library administration. *Jurnal Intelek*. https://doi.org/10.24191/ji.v19i2.26491

Barthorpe, G., & Wright, E. (2017). Changing Organizational Capacity and Building Staff Capability. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2132&context=iatul

Bihan, B. (2022). Developing and incorporating impactful library research guides for online and hybrid learners (pp. 103-119). Edward Elgar Publishing eBooks. https://doi.org/10.4337/9781800888883.00019

Bonnand, S., & Hansen, M. A. (2012). Embedded Librarians: Delivering Synchronous Library Instruction and Research Assistance to Meet Needs of Distance Students and Faculty (pp. 326-339). IGI Global. https://doi.org/10.4018/978-1-61350-441-3.ch015

Brenton, S., & Tury, S. (2021). Acceleration of digital learning and what it means for libraries (pp. 79-90). Chandos Publishing. https://doi.org/10.1016/B978-0-323-88493-8.00014-8

Cassidy, T., Reinauer, O., & Walz, A. (2016). In the Highways and Hedges: Library Support for OER Adoption Efforts at Higher Education Institutions across Virginia. https://doi.org/10.5703/1288284316313

Çullhaj, S. (2017). Key Features of Personalized Learning. European Journal of Multidisciplinary Studies, 6(2), 130. https://doi.org/10.26417/EJMS.V6I2.P130-132

Detterbeck, K., & Sciangula, M. (2017). Innovation Through Collaboration: Using an Open-Source Learning Management System to Enhance Library Instruction and Student Learning (pp.221-238). Chandos Publishing. https://doi.org/10.1016/B978-0-08-100598-9.00012-X

Digital Collaborative Governance of Library in Developing E-Library in the Technical Implementing Unit of Universitas Sebelas Maret's Library. (2022). Journal of Social and Political Sciences, 5(4). https://doi.org/10.31014/aior.1991.05.04.381

Din, S., & Ali, S. (2024). New library technologies and trends in 2025. International Journal for Multidisciplinary Research, 6(6). https://doi.org/10.36948/ijfmr.2024.v06i06.31169

Duke, B., Harper, G., & Johnston, M. (2013). Connectivism as a Digital Age Learning Theory. https://www.hetl.org/wp-content/uploads/2013/09/HETLReview2013SpecialIssueArticle1.pdf

Efird, L., Marion, E., Willis, D. J., Gierisch, J. M., & Corsino, L. (2024). Designing and Implementing a Community-Engaged Research e-Library: A Case Study for Adapting Academic Library Information Infrastructure to Respond to Stakeholder Needs. College & Research Libraries, 85(7). https://doi.org/10.5860/crl.85.7.994

Eyre, J. R., Maclachlan, J. C., & Williford, C. (2017). A splendid torch: Learning and teaching in today's academic libraries. *Lincoln (University of Nebraska)*. https://doi.org/10.5281/zenodo.7805874

Farmer, L. S. J. (2019). The Role of Librarians in Blended Courses (pp. 122-138). IGI Global. https://doi.org/10.4018/978-1-5225-5557-5.CH007

Fitria, T. N. (2023). Augmented Reality (AR) and Virtual Reality (VR) Technology in Education: Media of Teaching and Learning: A Review. International Journal of Computer and Information System, 4(1), 14-25. https://doi.org/10.29040/ijcis.v4i1.102

Fonseca, A. P., Zegers, C., & Firoozi, S. (2024). Collaborative Learning Ecosystems: Enhancing Communities of Practice in Digital Spaces, 1(2). https://doi.org/10.56830/ijsol12202405

Frank, J., Salsbury, M., McKelvey, H., & McLain, R. (2021). Digital Equity & Inclusion Strategies for Libraries: Promoting Student Success for All Learners, 5(3). https://jps.library.utoronto.ca/index.php/ijidi/article/view/36190

Garner, B., Snyder, T. L., Wise, J., & Walters, R. (2024). From tradition to transformation: Academic librarians as key players in online education. https://doi.org/10.69554/oqxc4912

Ghosh, S., Sarkar, S. K., Roy, P., Roy, B., & Podder, A. (2024). Role of Libraries in Promoting Digital Literacy and Information Fluency. Advances in Library and Information Science (ALIS) Book Series, 77-108. https://doi.org/10.4018/979-8-3693-3053-1.ch005

Giannaros, A., Kotsopoulos, K. I., Tsolis, D., & Pavlidis, G. (2020). Creating a Personalised Experience for Libraries' Visitors (pp. 491-498). Springer, Cham. https://doi.org/10.1007/978-3-030-36126-6_55

Gisip, J., Ibrahim, N., Ratim, S., & Ghani, F. S. A. (2024). Open educational resources (OER) in e-learning for higher education. International Journal on E-Learning and Higher Education, 19(2), 449-467. https://doi.org/10.24191/ijelhe.v19n2.19227

Goodsett, M., Miles, M., & Loomis, B. (2015). Developing an Open Educational Resource: Leading Campus OER Initiatives through Library-Faculty Collaboration. https://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1137&context=msl_facpu

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5*(3), 101–117. https://doi.org/10.1016/S0899-3467(07)60142-6

Henke, N. A., Anthoney, M., & Pierce, J. B. (2024). Librarians' Roles in OER Authoring and Open Pedagogy: Lessons from the Field. Internet Reference Services Quarterly, 1-10. https://doi.org/10.1080/10875301.2024.2443851

Hu, K. C., & Raman, A. (2024). Systematic literature review on the holistic integration of e-learning in universities: Policy, human, financial, and technical perspectives. Contemporary Educational Technology, 16(2), ep497. https://doi.org/10.30935/cedtech/14287

Idowu, E. (2024). Personalized Learning: Tailoring Instruction to Individual Student Needs. https://doi.org/10.20944/preprints202411.0863.v1

Jank, D. A. (2013). Characteristics of Collaborative Partnerships in Library Technology and Web Service Initiatives: Emergent Patterns and Lessons Learned (Vol. 36). Emerald Group Publishing Limited. https://doi.org/10.1108/S0065-2830(2013)0000036004

Karplus, S. S. (2017). Integrating Academic Library Resources and Learning Management Systems: The Library Blackboard Site. Education Libraries, 29(1), 5-11. https://doi.org/10.26443/EL.V29I1.219

Katz, S. (2019). Leveraging Library Expertise in Support of Institutional Goals: A Case Study of an OER Initiative at Lehman College. https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1300&context=le_pubs

Kim, J.-A. (2015). Integrating Communities of Practice into Library Services. Collaborative Librarianship, 7(2), 47-55. https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1067&context=collaborativelibrarianship

Kumar, N. S., Antoniraj, S., Jayanthi, S., Mirdula, S., Selvaraj, S., Rajkumar, N., & Senthilkumar, K. R. (2024). Educational Technology and Libraries Supporting Online Learning (pp. 209-237). SAGE Publishing. https://doi.org/10.4018/979-8-3693-2782-1.ch012

Kwanya, T. (2016). Information Seeking Behaviour in Digital Library Contexts (pp. 1384-1408). IGI Global. https://doi.org/10.4018/978-1-5225-0296-8.CH001

La Macchia, K. (2020). An academic digital library advancing an equitable and inclusive educational environment. 37(1), 18–27. https://doi.org/10.1108/DLP-07-2020-0062

Lalitha, B., Ramalakshmi, K., Gunasekaran, H., Parthasarathy, M., Saminasri, P., & Rajkumar, N. (2024). Anticipating AI Impact on Library Services (pp. 195-213). SAGE Publishing. https://doi.org/10.4018/979-8-3693-5593-0.ch014

Lo, L. S., & Vitale, C. H. (2024). Evolving AI Strategies in Libraries: Insights from Two Polls of ARL Member Representatives over Nine Months. https://doi.org/10.29242/report.aipolls2023

Makwana, P. N., & Gadhavi, G. G. (2023). The Evolving Role of Librarians in Digital Humanities. International Journal of Science and Research, 12(10), 830-832. https://doi.org/10.21275/sr231009164826

McAllister, C. R. (2018). Beyond Cost Per Use: Exploring Multivariable E-Resource Assessment. https://doi.org/10.5703/1288284316690

Nel, M. (2020). Information behaviour and information practices of academic librarians: a scoping review to guide studies on their learning in practice. https://doi.org/10.47989/IRISIC2020

Onunka, O., Onunka, T., Fawole, A. A., Adeleke, I. J., & Daraojimba, C. (2023). Library and information services in the digital age: opportunities and challenges. Acta Informatica Malaysia, 7(2), 113-121. https://doi.org/10.26480/aim.01.2023.113.121

Owens, E., Arce, V., Del Bosque, D., Fowler, R., & Sheffield, S. (2024). Beyond the Library: The Role of Academic Libraries' Chat Reference in Answering Campus Questions. The Reference Librarian. https://doi.org/10.1080/02763877.2024.2334216

Oyedokun, T. T. (2024). Navigating the dynamics of present-day academic libraries: An in-depth analysis of strategies, challenges, and emerging trends. IFLA Journal. https://doi.org/10.1177/03400352241291907

Pathak, N. K. (2023). (N.E.P. 2020) role of Libraries in the Field of higher Education. Techno Learn: An International Journal of Educational Technology. https://doi.org/10.30954/2231-4105.01.2023.2

Pettalongi, S. S., Londol, M. M., & Umboh, S. E. (2024). Disparities in Digital Education: Socioeconomic Barriers to Accessing Online Learning Resources. International Journal of Social and Human, 1(3), 181-189. https://doi.org/10.59613/be6gdv98

Pope, H. (2018). Chapter 2. Incorporating Virtual and Augmented Reality in Libraries. Library Technology Reports, 54(6), 8-11. https://journals.ala.org/index.php/ltr/article/view/6779

Priya, S., & Ramya, R. (2024). Future Trends and Emerging Technologies in AI and Libraries (pp.245-271). SAGE Publishing. https://doi.org/10.4018/979-8-3693-1573-6.ch010

Salim, F., Saigar, B., Armoham, P. K., Gobalakrishnan, S., Jap, M. Y. P., & Lim, N. A. A. (2016). Students' information-seeking intention in academic digital libraries (pp. 259-273). IGI Global. https://doi.org/10.4018/978-1-5225-0746-8.CH017

Sanches, T. (2018). Changing Roles for Research and Information Skills Development: Librarians as Teachers, Researchers as Learners (pp.462-471). Springer, Cham. https://doi.org/10.1007/978-3-030-13472-3_44

Sarkar, T. D. (2023). Augmented reality applications and the future library. Library Hi Tech News. https://doi.org/10.1108/lhtn-07-2023-0129

Satvik, B., & Naik, K. (2024). Barriers to the Adoption of E-Learning in Higher Education Institutes (HEI). International Journal of E-Services and Mobile Applications, 17(1), 1-19. https://doi.org/10.4018/ijesma.365667

Shankar, A. U., Tewari, V., Rahman, M., Mishra, A., & Bajaj, K. K. (2023). Impact of Virtual Reality (Vr) and Augmented Reality (Ar) in Education. https://doi.org/10.52783/tjjpt.v44.i4.1014

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. elearnspace. https://masterlereneninnoverenrotterdam.files.wordpress.com/2012/05/elearnspace-org-elearnspace-connectivism-a-learning-theory-for-the-digital-age.pdf

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Steele, C. (2019). Staffing the digital library of the 21st century: Professional Development and Continuing Education. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2132&context=iatul

Transue, B. (2013). Connectivism and Information Literacy: Moving From Learning Theory to Pedagogical Practice. Public Services Quarterly, 9(3), 185-195. https://doi.org/10.1080/15228959.2013.815501

Tumbleson, B. E. (2016). Collaborating in Research: Embedded Librarianship in the Learning Management System. The Reference Librarian, 57(3), 224-234. https://doi.org/10.1080/02763877.2015.1134376

Vogus, B. (2020). Examining virtual reference services in academic libraries. Public Services Quarterly, 16(4), 249-253. https://doi.org/10.1080/15228959.2020.1818664

Yukawa, J. (2010). Communities of Practice for Blended Learning: Toward an Integrated Model for LIS Education. Journal of Education for Library and Information Science, 51(2), 54-75. https://www.questia.com/library/journal/1P3-2012698611/communities-of-practice-for-blended-learning-toward

Towards Human-Centered Learning: AI-Supported Skill Management in a Large Automotive Organization

Svenja Pieczurczyk and Paul Drews
Institute of Information Systems, Leuphana University Lüneburg, Lüneburg, Germany

svenja.pieczurczyk@stud.leuphana.de
paul.drews@leuphana.de

Abstract: The use of AI in skill management has the potential to support personalized e-learning processes by identifying future skills, revealing skill gaps, and linking them to learning opportunities. At the same time, designing an AI-supported skill management system that is leveraging these potentials presents organizations with challenges. To investigate this design process, we conducted a qualitative case study based on 17 interviews and observations carried out during the pilot implementation of such a system. The case illustrates that AI-supported skill management does not automatically enable human-centered learning, but depends on balancing AI potentials, skill governance processes, and individual learning experiences. In this context, AI can be used to provide orientation in identifying future skills and deriving personalized learning impulses. This is accompanied by a reconfiguration of roles, responsibilities, and control, which are distributed across a shared sociotechnical governance structure.

Keywords: AI-Supported Skill Management, Sociotechnical Governance, Human-Centered Learning

1. Introduction

Digital transformation is reshaping today's work environment through increased speed, unpredictability, and technological advancement, resulting in rapidly changing qualification requirements within organizations. New skills emerge quickly, while existing skills lose relevance or become obsolete. This phenomenon, known as skill instability, confronts organizations with the challenge of identifying relevant skills and systematically developing them (Kusmin et al. 2024). One proposed response to address this challenge is the use of artificial intelligence (AI), which is expected to provide dynamic, data-driven, and personalized skill management solutions (Talodhikar and Farooqui 2025), thereby enabling tailored qualification measures (Freise et al. 2025).

Previous research has examined the potential of AI in workforce development and modern approaches to skill management, it remains unclear how the shift towards AI-supported skill management can be designed and what role humans will play in learning systems. Against this background, this study addresses the following research question: *How is AI-supported skill management configured during a pilot implementation in a large organization to support employee learning?* To explore this question, we conducted a qualitative case study within a large German automotive organization. The study is based on data collected through 17 interviews and observations carried out during the pilot implementation of an AI-supported skill management system.

The findings are structured along a sociotechnical perspective that considers the interplay between technological tools and platforms, organizational processes, and human-centered learning as an organizational phenomenon (Orlikowski 1992). The results demonstrate that AI-supported skill management does not automatically lead to human-centered learning. Instead, we describe the reconfiguration of a sociotechnical governance structure as a central aspect of this process by showing which technological infrastructure is required, how responsibilities should be redistributed and coordinated across actors, and under which conditions human-centered learning can emerge.

2. Theoretical Background

2.1 Skill Management in an Era of Skill Instability

The changing skill requirements for specific tasks and job profiles have long been regarded as a challenge for both business practice and academic research. In the past, the discussion was shaped by the concept of skills obsolescence. This term describes the depreciation of existing knowledge and thus the loss of the economic value of skills due to technological causes (de Grip and van Loo 2002). With the emergence of AI, both the dynamics of this process and the uncertainty surrounding skills have increased significantly (Meier and Seufert 2022). As a result, the focus of the discussion shifted toward the volatility and rapid transformation of skill requirements. Kusmin et al. (2024) move beyond the phenomenon of skills obsolescence and more

comprehensively describe this current skills instability as the continuous emergence, transformation, and obsolescence of skills caused by technological and labor-market-related changes.

As job requirements became increasingly dynamic, organizations' need for greater agility in describing and deploying their workforce also increased. At the same time, reskilling and upskilling became central measures for employee development. Consequently, organizational focus gradually shifted from traditional job-based approaches toward skill-based approaches (Collings and McMackin 2025). Modern skill management, which enables the design of a future-oriented learning ecosystem, is considered the foundation for structural workforce transformations (Darmawan 2025).

A key aspect of this development is that employees and their individual skills are placed at the center, rather than relying on rigid and hierarchically organized job descriptions (Collings and McMackin 2025). To make this manageable and transparent, systematic skill management is required, enabling the definition and description of required skills, their structuring, as well as their evaluation and validation (Meier and Seufert 2022). In this context, a skill management process should be implemented that is as dynamic and holistic as possible and that promotes the closing of individual skill gaps through skill development initiatives.

At the same time, aspects such as employees' motivation for lifelong learning, as well as organizational framework conditions such as learning culture and leadership, and the alignment of competencies with business goals, should be taken into account. Ultimately, skill management is intended to function as a navigator for the workforce through continuous technological change and to ensure that employees' individual needs are precisely aligned with market requirements so that emerging gaps can continuously be closed (Braun et al. 2025).

2.2 AI-Supported Skill Management as an Enabler for E-Learning

The use of learning technologies and AI for the analysis and management of skills is considered a core aspect of modern skill management (Braun et al. 2025). Data-driven models for individual skill profiles are intended to serve as a key factor in managing the complexity of skill requirements (Talodhikar and Farooqui 2025). Methods such as skills inference can be used to enable an up-to-date derivation and assessment of employees' implicit skills through the analysis of large-scale data streams, without requiring self- or external assessments. Furthermore, AI promises particularly detailed insights into employees' skill gaps, the quantification and analysis of skills, and the scaling of upskilling and reskilling measures (Tona et al. 2025). So-called SkillsTech solutions aim to integrate skill management and human resource development by, for example, using defined skill taxonomies as a reference point in digital learning experience systems. In this way, enablement of e-learning is created, as it allows a matching of employees' individual skill gaps with available learning content (Braun et al. 2025). According to Freise et al. (2025), AI can support the creation of personalized training plans that align with employees' individual career goals. Skill management is thereby closely linked to e-learning, as it reflects dynamic skill requirements and enables scalable, flexible, and personalized training.

2.3 A Sociotechnical Perspective on Human-Centered Learning

Despite the considerable potential of technological solutions, Meier and Seufert (2022) reveal a significant gap between the capabilities of modern SkillsTech solutions and their actual implementation. They further question whether AI is currently capable of supporting sustainable workforce transformation and employee development. Insufficient data availability, the complexity of human behavior, ethical concerns, as well as negative employee reactions, distrust, or lack of acceptance may represent inhibiting factors (Tambe et al. 2019).

From a sociotechnical perspective, organizational effectiveness is highest when social and technical systems interact harmoniously (Ang et al. 2025). In this context, learning is not understood solely as a technologically driven process, but rather as an active process of appropriation in which employees adapt technologies and integrate them into their work and learning practices (Orlikowski 1992).

To bridge the gap between technological possibilities and actual organizational value (Tian et al. 2023), increasing attention is being directed toward human-centeredness (Tona et al. 2025). In this approach, the individual needs of employees are placed at the center instead of offering standardized learning programs. In this way, employees are enabled to shape their own development more actively through self-determination, autonomy, and empowerment (Freise et al. 2025).

3. Methodology

Our case study examined a large automotive corporation based in Germany that, in response to a challenging business environment, aims to transform its workforce. The objective is to strengthen workforce adaptability through more suitable and future-oriented qualification. To achieve this, AI is intended to be used for process optimization. In line with this objective, the HR department developed an AI-supported skill management approach designed to make individual employee skill profiles visible and enable systematic gap analyses. The aim of the pilot project was therefore to examine how market-related skill requirements can be translated into an organizational skills taxonomy, how employees' existing skills can be made visible, and how the resulting skill gaps can be linked to targeted learning and development opportunities.

Guided by the objective of understanding the impact of AI on human learning in the context of organizational skill management, we conducted a qualitative case study (Yin 2018) to investigate that pilot implementation.

The data is based on a multi-perspective design and includes 17 semi-structured interviews (Mayring 2014), observational data from approximately 60 meetings (weekly and biweekly) carried out during a nine-month pilot phase as well as internal and external document material. The interviews were conducted between August 2024 and June 2025 and lasted between 33 and 60 minutes (M = 51.26 minutes). The semi-structured interview guide covered questions on continuous competency reassessment, including processes, tools, stakeholders, information sources, barriers, and enabling factors, as well as questions on current and future IT competencies and short-, medium-, and long-term approaches to competency development in the context of AI-supported skill management. This project-internal sample was well suited to the exploratory focus of the study, as the participants possessed detailed knowledge of the processes involved in the design and implementation of the pilot project. Reliance on interview data alone was reduced by triangulating the interview findings with observational data (Yin 2018).

Participants were selected based on their key roles within the pilot project, as they were directly involved in defining the skill infrastructure and integrating skill data into learning systems. These include HR digitalization specialists (n = 5, ID: T), stakeholders responsible for skill management in HR and IT (n = 6, ID: O), and learning experience designers (n = 6, ID: L) responsible for learning experience. All interviews were audio-recorded, transcribed, anonymized, and included in the analysis.

Data analysis was conducted using qualitative content analysis according to Mayring (2014). The aim was a systematic reduction of the material to identify central patterns of meaning. Significant text segments were coded and iteratively summarized into categories. The resulting codes were assigned to three overarching main categories, which were contextualized within sociotechnical theory (Orlikowski 1992) along the technological, organizational, and human-centered learning dimensions. The development of subcategories was deliberately omitted, as the main categories provided a sufficiently differentiated structure to address the research question. The analysis process was structurally supported using Word and Excel (Ose 2016).

4. Findings

4.1 Building an AI-Supported Skill Infrastructure

Our study identifies three core elements of the emerging skill infrastructure that appeared central in the pilot implementation of AI-supported skill management.

4.1.1 Development of an AI-Supported Skill Library

As a central infrastructural component of AI-supported skill management, the organization implemented an AI-generated skill library based on large-scale benchmark workforce data. It serves as the foundation for aggregating and standardizing skill data and includes AI-identified skills. It establishes the basis for the organization’s internal skill taxonomy, which was integrated into existing Human Resource Management (HRM) and learning experience systems.

4.1.2 AI-Supported Skill Identification and Bundling

*T*he described infrastructure is complemented by an intended enabling mechanism that not only identifies relevant skills, but also assigns synonyms to them and clusters them conceptually. This continuous process of pre-structuring and simplification makes skill data more precise and manageable. “AI is supposed to help us

identify the right skills and match them with the right trainings. AI is supposed to help us assign synonyms to skills and create bundles of skills. Another aspect is that we want to use AI to derive skills automatically." (T2)

4.1.3 Integration of AI-Supported Skill Infrastructure into HRM and Learning Experience Platform (LXP):

The identified skill data were integrated into the core functionalities of existing HRM systems which serve as a single source of truth for the functional areas of skill management and learning. A key aspect was the integration of skill data into the organization's LXP. "The skill catalog as such serves different functions within our large HRM system […] for example in combination with our LXP platform." (O6)

In this context, skill data support the platform's recommendation mechanisms, as they are tagged to training content, thereby providing orientation in individual learning processes, in the assessment of one's own skills and the identification and closing of skill gaps.

4.2 Distributed Skill Governance

In the pilot context, the availability of a new skill infrastructure appeared to redistribute aspects of skill management governance. Two central aspects of this process were the emergence of new roles and coordination between them. Our case study therefore illustrates how these distributed responsibilities were aligned in practice.

4.2.1 Distributed Roles and Responsibilities

Within this context, HR assumes responsibility for the central structuring of the skill taxonomy and for steering the operationalization of skill management. "We now have the data […] This means that in the future we will manage much more centrally." (O6) One of these steering functions consists of regularly updating the skill taxonomy through structured review cycles (three to four times per year), during which AI-generated skill data are validated by HR and integrated into existing HRM and learning systems. "We use review cycles for the job and skill catalog. We plan to perform updates three to four times a year." (O3)

Line managers are responsible for applying the taxonomy and specifying the requirements of individual positions. In dialogue with employees, for example during annual performance reviews, they define relevant skills and required proficiency levels within the respective organizational context, enabling them to identify and document skill gaps.

Employees constitute an active part of this dialogue and independently complement their individual skill profiles as a basis for setting their personal learning goals and creating individual qualification pathways.

4.2.2 Coordinating AI-Supported Skill Management Across Organizational Actors

Distributed responsibilities are aligned in practice through a shared AI-supported skill infrastructure that underpins a multi-level skill governance approach. Within this process, coordination is achieved through recurring review cycles as a governance mechanism, the integration of the skill infrastructure into HRM and LXP systems as a technical mechanism, and continuous dialogue between HR, line managers, and employees as a social mechanism. The AI-supported skill infrastructure serves as a blueprint that constitutes a higher-level, organization-wide reference structure standardized and embedded within HR systems.

The organization aims to establish a bottom-up approach in which employees keep the system responsive to actual work practice. In this sense, employees are not treated as passive objects of skill assessment, but rather as active contributors shaping their own learning pathways. "We need to create a framework in which people can independently adapt their profiles according to the skills they actually need. We may sit next to each other in the same office with the same job description, yet you may still perform different tasks than I do and therefore require a different skill profile. This individualization is therefore extremely important. In this way, we can achieve this level of individualization, and every employee can also independently determine their own skills." (O2)

4.3 Conditions for Human-Centered Learning

The pilot implementation revealed four conditions that appeared to shape whether AI-supported skill management could become human-centered in practice.

4.3.1 Perceived Individual Value

The findings suggest that in this pilot context, human-centered learning was closely linked to employees' willingness to participate in skill profiling and disclose skill data.

For this reason, HR actors emphasized the need to create visible incentives and recognizable individual benefits for maintaining one's own skill profile. These benefits may include improved training recommendations, broader career opportunities, or access to new projects. However, the pilot also indicates that such benefits are currently limited, as the skill-based logic has thus far mainly been integrated into the learning environment rather than into the broader HR architecture. Consequently, human-centeredness does not emerge automatically through personalization alone, but rather depends on whether employees perceive genuine value in engaging with the system. "We will have to establish processes in which employees realize that maintaining their skills actually benefits them, whether in terms of career development, better training recommendations, new career paths, or exciting project opportunities that are suggested to them. […] A major challenge is gaining people's acceptance to disclose their data and enabling them to recognize a benefit in doing so." (O3).

4.3.2 Dialogical skill Gap Analysis

*T*he findings suggest that employee development remains a socially negotiated process rather than a purely data-driven one. Employees articulate their goals and interests during performance and development reviews, in which skill gaps are jointly interpreted together with their line manager. "Exactly, whereby it is important to us that the employee remains in the driver's seat, within our group this is very important. Employees must clearly articulate what their goals are and what they want to achieve. The plan for the qualification pathway is then developed together." (O5) Skill assessments and development decisions are ultimately embedded in dialogue and cannot be fully delegated to the system. This indicates that human-centered learning in AI-supported skill management is characterized by a hybrid logic in which algorithmic recommendations and interpersonal sensemaking complement one another. "In the end, it will always remain with the manager and the employee to assess skills. This will always take place in a personal conversation. That is irreplaceable." (O3)

4.3.3 Enablement for Skill Management Participation

In the pilot context, human-centered learning in AI-supported environments appeared to require active enablement. HR and business units described it as their responsibility to support employees in using learning platforms and to develop the capabilities required for effective action in increasingly data-driven learning environments. This includes not only familiarity with the learning platform itself, but also broader data-related skills. "It is very important to empower people to engage with LXP platforms." (T5) As digital systems make learning needs, development opportunities, and skill gaps more visible, employees are increasingly expected to interpret information, operate digital tools, and make decisions based on data. Human-centered learning therefore also creates a new demand for digital and data literacy. "Everything is moving toward data usage and data-based decision-making. Through digitalization, many things become more transparent. At the same time, I have much more information and can analyze things more effectively and comprehensively. However, this also requires being capable of using tools, thinking in terms of data, and making data-based decisions [...]. In principle, every employee should become a small data scientist." (T3) The findings suggest that organizations need to actively support this transition through awareness-building and empowerment measures if AI-supported skill management is intended to empower rather than burden learners.

4.3.4 Access to Personalized Learning Opportunities

To address emerging qualification needs, HR uses individual skill profiles to offer data-driven, tailored, and future-oriented training opportunities. "Employees can formulate their own goals accordingly, and we then, of course, use this information to determine which training programs we need to provide." (O4) Employees can autonomously choose from options provided through the learning experience platform.

5. Discussion

5.1 Reconfiguring Sociotechnical Governance for Human-Centered Learning

Within the context of the AI-supported skill management approach, we identified a socio-cultural shift in skill-related responsibilities in our case study. Rather than remaining primarily embedded within HR, skill management appeared to become a sociotechnical structure involving HR actors, AI solutions, digital systems,

managers, and employees, thereby transforming who contributes to skill identification, how development needs are interpreted, and where responsibility for learning and employability is situated. Meier and Seufert (2022) highlight the associated need for coordination and collaboration. Our work builds on this by showing how the responsibility is configured in practice, and how core elements of the skill infrastructure, new roles, and coordination processes interact, as well as under which key conditions they can foster human-centered learning.

However, this reconfiguration of responsibility does not necessarily imply a loss of relevance for HR. Rather, the findings of our study indicate a transformation of the HR function from an administratively oriented role toward a more strategic and advisory one. In line with Darmawan (2025), our case study suggests that HR increasingly acts as an orchestrator of skill-related infrastructures. In this role, HR provides access to future-oriented skills, translates external skill demands into the internal skill taxonomy, and establishes conditions for systematic engagement with development needs.

However, its control is exercised less through the direct specification of development activities and more through the design of guiding frameworks, such as the skill taxonomy and digital platforms, which shape how skills become visible and actionable within the organization.

At the same time, our findings point to the importance of line managers as interpretative intermediaries within skill management. While AI-generated skill data provided by HR offers strategic orientation, it remains too abstract to be fully understood by all employees or to directly guide individual development in practice. Line managers contextualize this skill data by relating it to specific roles, tasks, and concrete job requirements. In doing so, they identify skill gaps and translate the abstract skill taxonomy into situated developmental priorities. This is consistent with the work of Braun et al. (2025), which emphasizes the importance of managerial involvement in skill development and skill gap closure. Our findings reinforce this perspective by showing that line managers also function as a critical interface between data-driven skill intelligence and the lived realities of employees' work practices.

Employees, in turn, emerge as active actors in the dialogue on skill requirements and as co-producers of skill visibility, rather than passive recipients of development measures. However, the findings also suggest that enhanced agency is accompanied by new forms of responsibility that need to be actively developed. It is already established that the effectiveness of skill management systems is particularly high when employees are willing to disclose their skills and engage with e-learning platforms (Tona et al. 2025). Responsibility for employability and development appeared to be partially shifted toward the individual.

While Tona et al. (2025) describe the pursuit of objectivity in employees' self-assessment of skills, our pilot study suggests that subjectivity in AI-supported skill management can be of particular value. Rather than being a problem to be eliminated, this individuality represents a valuable resource for human-centered learning, as it enables the interpretation of development needs in relation to actual work practices, personal goals, and social dialogue. In line with Braun et al. (2025), our study further shows that the value of this subjectivity lies not only in its relevance for employees, but also in its contribution to organizational learning, as interpersonal dialogue creates meaningful opportunities to interpret skill gaps and shape development pathways.

In summary, the case suggests that the pilot implementation of AI-supported skill management was associated with a decentralization of skill governance, in which HR was no longer solely responsible. On the one hand, relevant skills are increasingly made visible within the organization through AI, data infrastructures, and recommendation systems. On the other hand, interpretative and development-related decisions remain embedded in the dialogue and interaction between line managers and employees. Our study therefore illustrates a shift from centralized administrative skill management towards a more distributed form of sociotechnical governance, in which agency and control are not opposing forces but are continuously renegotiated between human and technological actors. In this context, AI-supported skill management should therefore not be understood as a replacement for human judgment, but as an enabler of more human-centered learning processes.

5.2 Implications, Limitations and Future Research

This study contributes to the theoretical debate on AI in skill management by illustrating, through a pilot case, that AI-supported skill management constitutes a complex and ambivalent field rather than a purely technological advancement. The findings are situated within the tension between technological enthusiasm and practical skepticism that characterizes the implementation of such systems in organizational contexts. By providing empirically grounded insights into the integration of AI-generated skills into organizational learning

platforms, the study illustrates how AI-supported skill management can serve as a foundation for learning recommendation mechanisms, while remaining dependent on the interplay between technological, organizational, and human actors.

The case illustrates that such systems have the potential to enhance personalized learning, while also emphasizing the importance of aligning future-oriented skills with organizational structures and human-centered learning practices. For practitioners, it is not sufficient to merely introduce the technology or define technical implementation conditions. Rather, AI-supported skill management must be designed in a way that actively enables the interplay between technological systems, organizational structures, and human actors.

Our contribution, while insightful, faces several limitations. (1) The sample size of 17 participants entails a risk of bias and limits the representation of diverse perspectives, the depth of insight, and the generalizability of the findings. (2) Generalizability is further constrained, as the study was conducted within a single organization. However, given the organization's global scale, the findings still provide insights into core aspects of implementing AI-supported skill management that may be relevant for other organizations. (3) The study does not directly capture employees' perspective in detail through interviews. Nevertheless, the selected sample provides a relevant overview, as participants were actively involved in shaping the process.

Future research could address these limitations in several ways. First, investigating the employee perspective in AI-supported skill management systems. In particular, it may be useful to distinguish between multiple forms of employee involvement, ranging from active co-creation of skill profiles and governance contributions to the use of system-generated learning recommendations. This dual role highlights employees both as active contributors to and users of AI-supported skill ecosystems. Second, longitudinal studies should be conducted to investigate the long-term effects of AI-supported skill management on employability and individual learning activities. In addition, the technological and organizational maturity of AI-supported skill management should be considered over a longer period of time. Finally, we believe that future research is needed to confirm our assumptions by examining other organizations in various industries and including further experiences and opinions through a larger number of interviews.

6. Conclusion

AI has the potential to modernize skill management in addressing skill instability, redefining humans' role in future learning and skill management systems. Our study suggests that leveraging this potential requires an understanding of the interplay between technological, organizational, and human factors. In this regard, we identify sociotechnical structures that appeared relevant for fostering human-centered learning in the pilot context. These prerequisites include the organizational and social embedding of AI, a clearly recognizable individual value for employees in disclosing and maintaining their skill profiles, the preservation of employee agency, targeted enablement for the use of digital learning environments, and the central role of personal dialogue in identifying and addressing skill gaps. Overall, the findings indicate a shift toward more human-centered learning approaches, in which organizations tailor learning opportunities based on individual competence profiles and development needs. In this context, AI does not replace human judgment but provides an orientation framework for e-learning.

Ethics Declaration

All interviews were anonymized, voluntary, and conducted with informed consent, and no sensitive personal data were collected. The study complies with ethical standards and data protection regulations.

The results, opinions and conclusions expressed in this study are not necessarily those of Volkswagen Aktiengesellschaft.

AI Declaration

The authors declare that they have used AI for proofreading and to enhance the readability of this manuscript. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

Ang, K. C., Sankaran, S. and Liu, D. (2025) “Advancing sociotechnical systems theory: New principles for human-robot team design and development”, Applied Ergonomics, 129, Article 104604. doi:10.1016/j.apergo.2025.104604.

Braun, G., Räty, P., Bokinge, M., Rikala, P., Hämäläinen, R., Syberfeldt, A. and Stahre, J. (2025) “The skill bridge–A global qualitative analysis of skill gap management”, Journal of Environmental Management, 395, Article 127738. doi:10.1016/j.jenvman.2025.127738.

Collings, D. G. and McMackin, J. (2025) “Skills-first HR: A key enabler of future global strategy”, Organizational Dynamics, 54, Article 101140.

Darmawan, D. (2025) “Job redesign, automation, and digital workforce competencies in the twenty-first century”, Bulletin of Science, Technology and Society, 4(3), pp. 103–114.

de Grip, A. and van Loo, J. B. (2002) “The economics of skills obsolescence: A review”, in de Grip, A., van Loo, J. B. and Mayhew, K. (eds.) The economics of skills obsolescence: Theoretical innovations and empirical applications. Oxford: Elsevier Science, pp. 1–26.

Freise, L. R., Ritz, E., Bretschneider, U., Rietsche, R., Beitinger, G. and Leimeister, J. M. (2025) “How Siemens empowered workforce re- and upskilling through digital learning”, MIS Quarterly Executive, 24, Article 4. doi:10.17705/2msqe.00118.

Kusmin, K. L., Normak, P. and Ley, T. (2024) “A methodology for planning, implementation and evaluation of skills intelligence management: Results of a design science project in technology organisations”, Frontiers in Artificial Intelligence, 7, Article 1424924. doi:10.3389/frai.2024.1424924.

Mayring, P. (2014) Qualitative content analysis: Theoretical foundation, basic procedures and software solution. Klagenfurt.

Meier, C. and Seufert, S. (2022) “AI-supported systems for integrated skills-management and skills-development”, in AI education in the context of work. Cham: Springer International Publishing, pp. 3–25.

Orlikowski, W. J. (1992) “The duality of technology: Rethinking the concept of technology in organizations”, Organization Science, 3(3), pp. 398–427. doi:10.1287/orsc.3.3.398.

Ose, S. O. (2016) “Using Excel and Word to structure qualitative data”, Journal of Applied Social Science, 10(2), pp. 147–162. doi:10.1177/1936724416664948.

Tambe, P., Cappelli, P. and Yakubovich, V. (2019) “AI in human resources management: Challenges and a path forward”, California Management Review, 61(4), pp. 15–42.

Talodhikar, S. and Farooqui, S. (2025) “Utilizing artificial intelligence for competency mapping and personalised skill development in IT organizations”, Journal of Information Systems Engineering and Management, 10(15s).

Tian, Y., Claudy, M., Xu, D. J. and Liao, S. S. (2023) “Exploring the role of AI explanations in delivering rejection messages: A comparative analysis of organizational justice perceptions between HR and AI”, in Proceedings of the 44th International Conference on Information Systems (ICIS 2023). Hyderabad, India: Association for Information Systems.

Tona, O., Leidner, D. E., van der Meulen, N., Wixom, B., Nunes, J. and Shagam, D. (2025) “The deployment of AI to infer employee skills: Insights from Johnson and Johnson’s digital-first workforce initiative”, Information Systems Journal, 35(6), pp. 4290–4320. doi:10.1111/isj.12594.

Yin, R. K. (2018) Case study research and applications. 6th edn. Thousand Oaks, CA: Sage Publications.

Curriculum Gap Revisited: LLMs, Adaptive Expertise and PBL for IS

Paul Pierce
Department of Informatics, School of Economics and Management, Lund University, Sweden

paul.pierce@ics.lu.se

Abstract: The Information Systems (IS) field has long faced a misalignment between curriculum guidelines and labour-market expectations. Drawing on prior studies comparing Swedish IS job advertisements with the IS 2010 Curriculum Guidelines and the IS2020 Competency Model (Steen and Pierce, 2020; Steen and Pierce, 2023), this paper argues that the mainstreaming of large language models (LLMs) since 2022 has intensified this gap. IS education remains largely organised around stable, codified competencies; the paper argues that, in the post-2022 context, employers will increasingly value adaptive expertise, critical AI literacy and robust transfer across contexts — a claim the planned empirical phase is designed to test. The Stanford AI Index from 2026 places this concern in wider context, reporting that formal education is lagging behind AI and that employment among software developers aged 22 to 25 fell nearly 20% from 2024, suggesting pressure on the entry-level roles for which IS programmes traditionally prepare graduates (Sajadieh et al, 2026). Against this backdrop, the paper makes a conceptual and programmatic case for problem-based learning (PBL), reconceived through Savin-Baden's (2014) new constellations framework and Barnett's (2017) feasible utopias, as a pedagogical architecture for developing adaptive expertise. Rather than proposing another curriculum revision, I argue for rethinking how IS education is designed and delivered, with LLMs functioning not as answer engines but as co-inquirers that stretch students' reasoning, judgment and metacognitive capacity. The paper engages the debate over minimally guided instruction (Kirschner, Sweller and Clark, 2006; Hmelo-Silver, Duncan and Chinn, 2007) and concludes with a research agenda comprising a regional employer survey and follow-up interviews. Prior content analysis of 12,358 Swedish IS job advertisements showed that interpersonal dispositions appeared in 97.3% of postings and enterprise architecture in 15.1%, areas current IS frameworks under-weight. A planned next phase will survey approximately one hundred employers in the Öresund cross-border region and conduct in-depth interviews with recruiters, IS managers and recent IS graduates, probing the reasoning behind employer expectations that the prior job-advertisement data could not reveal. The paper contributes a structural argument and a planned empirical programme to the IS education community.

Keywords: IS Education, Curriculum Gap, Problem-Based Learning, Large Language Models, Adaptive Expertise, Learning Ecologies

1. Introduction

The Information Systems field has always had an uneasy relationship with its own curriculum. This is not unique to IS; most applied disciplines struggle to reconcile the pace of industrial change with the rhythms of curriculum revision, but it is more acute in IS than in most. The community has responded with two major frameworks over the past sixteen years: the IS 2010 Curriculum Guidelines (Topi et al, 2010) and the IS2020 Competency Model (Leidig et al, 2022). Both were serious efforts to define what an IS graduate should know and be able to do. And yet, as my late colleague Odd Steen and I argued in a series of empirical studies of Swedish IS job advertisements, neither fully captured what Swedish employers were asking for (Steen and Pierce, 2020; Steen and Pierce, 2023). ICT System Administration was more prominent in employer expectations than in either framework. Enterprise Architecture, downgraded in IS2020 to a graduate-level concern, remained a meaningful presence in undergraduate job ads. Most tellingly, the interpersonal and communicative dispositions employers emphasised most, namely cooperation, communication and language proficiency, were treated with more ambivalence in the guidelines than the evidence warranted.

These gaps suggest a structural limitation in how the IS field conceptualises what education is for: an implicit assumption that the goal is to produce graduates who possess a defined set of competencies, rather than graduates capable of learning, adapting and exercising judgment in conditions that cannot be fully anticipated at the time of study. This distinction has long been made in the educational literature through concepts such as adaptive expertise, transfer and supercomplexity. Contemporary accounts of adaptive expertise and adaptive performance emphasise professionals' capacity to respond to novel and changing situations, rather than merely reproduce established routines (Pelgrim et al, 2022). Work on epistemic fluency similarly frames professional education as preparation for knowledgeable action across multiple forms of expertise, tools and contexts (Markauskaite and Goodyear, 2017). Meanwhile, the literature on wicked competencies highlights the difficulty, but necessity, of cultivating evaluative judgment, communication, collaboration and other complex graduate capabilities that resist simple curricular codification (Sridharan and Kinshuk, 2024).

Since the release of ChatGPT in November 2022 and the rapid proliferation of LLMs as everyday professional tools, the IS competency landscape has undergone a shift qualitatively different from the incremental changes

that typically prompt curriculum revision. This is not simply a matter of adding AI literacy or prompt engineering to the skill list; it is a recognition that LLMs have begun to automate precisely the codified, procedural competencies IS education has historically centred on. Routine programming, information retrieval, basic data analysis and system documentation are not disappearing from the IS professional's world, but they are becoming tasks that a capable professional manages and interrogates rather than performs from scratch. What this creates is a profound elevation of what might be called the judgment premium: the value placed on the capacity to evaluate, contextualise, reframe and critically engage with AI-generated outputs. The Stanford AI Index 2026 reports productivity gains of 14% to 26% in structured, process-driven tasks while tasks requiring more judgment showed weaker or negative effects (Sajadieh et al, 2026). Employment among software developers aged 22 to 25 fell nearly 20% from 2024, while older cohorts have not experienced the same decline. Correlation is not causation, but the directional signal is difficult to ignore.

This paper responds to that challenge as the conceptual foundation of a broader research programme. I draw on the prior job-advertisement studies to establish the curriculum–employer gap as it existed prior to the LLM inflection point, then argue that the post-2022 context has deepened that gap in ways no further iteration of the IS2020 framework is likely to address. The main contribution is a conceptual argument for problem-based learning, reconceived through Savin-Baden's account of new PBL constellations and Barnett's account of the ecological university as a feasible utopia, as a pedagogical architecture well suited to developing adaptive expertise.

The paper focuses on IS education at the undergraduate level, where curriculum guidelines are most prescriptive and the misalignment most consequential. The arguments are applicable more broadly to information system education. The learning ecologies perspective adopted here situates education not only within the classroom, but across the wider contexts, relationships, practices and institutional arrangements through which professional capability develops (Jackson, 2013; Barnett and Jackson, 2019).

2. The Curriculum-Employer Gap: What We Already Know

The empirical foundation for the present argument is the studies Odd Steen and I conducted between 2019 and 2023, which examined the relationship between IS curriculum guidelines and Swedish employer expectations. The first study (Steen and Pierce, 2020) examined the content and structure of nineteen Swedish undergraduate IS programmes alongside their Australian and US counterparts, revealing a distinctive national profile: stronger emphasis on enterprise architecture and application development than either comparator framework prescribed. The second study (Steen and Pierce, 2023) moved the lens from curriculum documents to the labour market itself, conducting content analysis of 12,358 IS job advertisements drawn from Platsbanken, the Swedish Public Employment Service's open-data job database, with publication dates spanning 2017 to early 2023.

Four findings are most relevant here. First, ICT System Administration was substantially more prevalent in the ads than either the IS 2010 Guidelines or the IS2020 Competency Model acknowledged: in the top ten occupations specifying an IS education background, three fell within this category, including the second most frequent overall (System Manager). Neither framework treats this as a significant competency domain. Second, the three most prominently demanded personal dispositions, being communicative, demonstrating language proficiency in Swedish and/or English and being cooperative, appeared in almost all (97.3%) of the IS-education-specifying ads. The IS2020 Competency Model is considerably more preoccupied with technical competency than with these dispositions, which I would argue treat as supplementary something employers regard as constitutive. Third, Enterprise Architecture, downgraded in IS2020 to a graduate-level concern, was mentioned in 15.1% of IS-education-specifying ads at roughly equivalent rates for undergraduate and graduate-level candidates. This reflects a broader tendency of the IS2020 framework to follow a US perspective; six of the eleven members of the IS2020 Task Force were from US institutions, and the primary data source for occupational analysis was the AIS Job Index, a US-specific dataset. Fourth, and most pertinent to the present paper, was the finding around programming. IS2020 re-established programming as a core competency, a move with considerable support in the literature given its demotion from the IS 2010 framework. Our evidence offered only partial support: while programming skills were present in the ads, the depth and specificity called for in IS2020 were not consistently reflected in employer expectations. More significantly, the original paper flagged that the rapid development of generative AI was already complicating the picture: recent work shows that LLM-based tools now provide code suggestions, explanations and debugging support in programming education, while also raising questions about how curricula can preserve algorithmic thinking, error detection and independent problem-solving as routine coding tasks become increasingly automated (Kohen-Vacs, Usher and Jansen, 2025). These prior findings establish the curriculum-employer gap as it stood up to early 2023; they do

not yet establish how the gap has shifted under LLM pressure, which is the present paper's conceptual focus and the question the planned next phase is designed to answer.

3. The LLM Inflection Point: From Competency Gap to Competency Crisis

The arrival of ChatGPT in November 2022 represents one of those rare moments when accumulated incremental pressures produce a qualitative shift in an applied field. I am not making a claim about technological singularity or artificial general intelligence; I am making a more modest but still significant claim about the nature of the IS professional's work and, by extension, the nature of IS education.

LLMs are extraordinarily effective at producing fluent, contextually plausible text in response to a prompt, including code. They perform well on tasks defined by a well-specified input-output relationship and drawing on patterns abundantly present in training data. What is instructive is what the Stanford AI Index 2026 calls the 'jagged frontier' of AI capability: frontier models now exceed human performance on structured benchmarks, from mathematically very complex problems to advanced coding competitions, while achieving only 50.1% accuracy in telling time from an analog clock (Sajadieh et al, 2026). This uneven profile maps with remarkable precision onto the routine/adaptive expertise distinction. LLMs are considerably less reliable at tasks requiring genuine novelty, deep contextual understanding, rigorous causal reasoning or what Schön (1983) called reflection-in-action: the situated judgment of assessing what a particular situation actually calls for, weighing factors that are not fully specified. The distinction between the production of plausible outputs and the exercise of genuine judgment is the central educational issue raised by LLMs.

This distinction has profound implications for IS education. The competencies foregrounded in IS 2010 and IS2020, specifically programming, data management, systems analysis and IT infrastructure, are largely competencies of codified knowledge: things that can be described, specified, practised and now, to a significant degree, automated. If IS graduates will work in environments where LLMs are routine professional tools, we need to ask what aspects of IS practice remain genuinely dependent on human expertise, judgment and the capacity for continuous learning.

The answer lies in two interconnected areas. The first is what Hatano and Inagaki (1986) called adaptive expertise: the capacity to handle novel problems that do not fit established routines, to recognise when standard approaches are inadequate and to generate new approaches through deep understanding and creative reasoning. Adaptive expertise is contrasted with routine expertise: high competence at familiar problem types, developed through repeated practice. Routine expertise is exactly what LLMs are increasingly good at providing. Adaptive expertise, which requires understanding why and being able to reframe and transfer that understanding, remains less securely supported by current LLM capabilities than routine performance. The second is what the learning ecologies literature frames as transfer across contexts (Jackson, 2013): the capacity to apply knowledge not just within the domain where it was acquired but across different professional and institutional settings, different kinds of problems and different sociotechnical arrangements. This boundary-crossing inquiry (Engeström, 2001) is the hallmark of what Barnett (2000) described as the graduate capable of functioning in supercomplexity, the condition in which not just the answers but the questions themselves are contested.

Looking at the IS2020 framework through this lens, one is struck by its structural bet on codification. The competency model is built around specifiable, teachable skill areas: application development, data and information management, IT infrastructure, IS management and strategy. These are genuinely important. But the framework's underlying logic is one of competency acquisition. This logic was already showing strain in the pre-LLM context, as our job ads study suggested. In the post-LLM context, it is structurally inadequate, not because the competencies identified are wrong, but because the conditions under which IS professionals must exercise those competencies have changed in a way that elevates a different order of capacity above them.

There is a longer-term learning dimension. The Stanford AI Index reports early but troubling evidence that heavy AI reliance may carry learning penalties that slow skill development (Sajadieh et al, 2026). The research is preliminary, but the directional concern is consistent with what educational theorists would predict. If students habitually outsource the generative and analytical stages of cognitive work to an LLM, they forgo the productive struggle through which adaptive expertise develops. The wider context the AI Index describes is one of formal education lagging behind AI: four in five U.S. high school and college students report using AI for schoolwork, while only around half of middle and high schools have AI policies and just 6% of teachers say those policies are clear (Sajadieh et al, 2026). This is not merely an administrative gap; it reflects pedagogical uncertainty about

what education should accomplish when the tools available to learners differ fundamentally from those available to the designers of their curricula.

There is also an ethical dimension. The use of LLMs raises questions about authorship, the distribution of cognitive labour, power and surveillance, and the environmental costs of large-scale AI computation (Kasneci et al, 2023). These are constitutive of what it means to be a responsible IS professional today. An IS education designed around adaptive expertise and critical AI literacy is inherently better positioned to address these dimensions than one designed around technical competency acquisition, precisely because it keeps judgment, including ethical judgment, at the centre of the educational enterprise.

4. The Pedagogical Response: Problem-Based Learning Reconceived

Having established the prior curriculum-employer gap and argued that LLMs have deepened it, I now want to make the positive case for problem-based learning as the pedagogical response adequate to the present moment. PBL is not a silver bullet, and its claims have been seriously contested. Kirschner, Sweller and Clark (2006) mounted an influential critique of minimally guided instruction, arguing on cognitive-load grounds that novice learners benefit more from worked examples and direct instruction than from open-ended inquiry. The response from Hmelo-Silver, Duncan and Chinn (2007) was that this critique conflated PBL with discovery learning and underestimated the role of scaffolding in well-designed PBL environments, which manages cognitive load while preserving inquiry. I take the Hmelo-Silver position seriously: the version of PBL I am advocating is heavily scaffolded, not free-form, and the role of the educator is to design problem environments in which inquiry is structured rather than abandoned to chance. As Savin-Baden (2014) observed, 'problem-based learning' covers a considerable range of pedagogical practices. What I am arguing for is not PBL as a generic method but PBL as theorised through what Savin-Baden (2014) calls new constellations: pedagogical arrangements explicitly designed to develop the capacity to engage with ill-structured problems, to work across disciplinary and contextual boundaries and to exercise judgment under uncertainty.

Savin-Baden's account is built around the idea that the most valuable learning happens at the edges of competence: when students encounter problems that do not yield to existing frameworks. This is well suited to the LLM context for a paradoxical reason: LLMs are very good at producing answers to questions that have answers. They are less useful when the question is genuinely open, when the task is not to retrieve information but to exercise judgment about what the problem actually is, what knowledge is relevant and what would count as an adequate solution.

Consider what this means for IS education. A systems analysis task that can be completed by prompting an LLM is unlikely to develop adaptive expertise. The educational value lies instead in tasks where students must interpret conflicting stakeholder interests, work with incomplete information, navigate value conflicts and recognise that technical choices have social and organisational consequences. These are the conditions under which professional judgment is formed. If IS education is to remain meaningful in an LLM-augmented world, it must create learning experiences that students cannot simply delegate to the machine.

Barnett's framework is particularly generative here. His account of supercomplexity, the condition in which not just practical problems but the frameworks for understanding them are multiple, contested and in tension, is both diagnosis and prescription. A graduate educated for supercomplexity has developed what Barnett calls the will and the capacity to engage: to act in uncertainty, to tolerate the discomfort of not knowing and to bring critical reflection to bear on the frameworks shaping their own understanding. This is precisely the disposition IS education for the LLM era needs to cultivate.

The learning ecologies perspective adds a further dimension. The idea that learning happens within and across interconnected systems of practice (Jackson, 2013; Barnett and Jackson, 2019) is directly relevant to the IS context, where the boundaries between educational and professional practice are particularly permeable. An IS graduate who has learned to exercise judgment within the bounded context of a university course but lacks the capacity to transfer that judgment to the different problems they encounter in practice has not yet learned what matters most. PBL designed around problems that cross institutional and disciplinary boundaries (Engeström, 2001) is better positioned than lecture-based or competency-focused instruction to develop this transferable capacity.

The dominant framing of LLMs in education positions them either as threats to academic integrity or as tutoring tools. Both share an assumption: the student is the learner and the LLM is the resource. What I am proposing is more demanding: in a PBL context designed for IS education, LLMs can function as co-inquirers, partners in the

inquiry process that students work with and against in developing their own judgment. This draws on a tradition in educational theory that sees thinking as fundamentally dialogical (Wegerif, 2007). What is novel is the character of the LLM as a dialogical partner: extraordinarily knowledgeable, fluent and tireless, always producing something plausible. The educational challenge is that plausibility and correctness are not the same thing, and learning to distinguish them is a form of reasoning that must be actively cultivated. This framing receives some empirical grounding: Anthropic (2025) reports that students using Claude often engage in cognitively consequential work, including creating and improving educational content and obtaining technical explanations or solutions. The sample is self-selected, with computer-science students overrepresented, but the headline pattern is suggestive: left to their own devices, students tend to use LLMs in the register of higher-order thinking. What PBL can add is a structured pedagogical architecture that makes the critical and reflective dimensions of this engagement explicit, sustained and subject to assessment.

5. Towards a Reconceptualised IS-PBL Framework

What follows is a set of design principles rather than a finished architecture; the empirical research programme described in the next section is necessary to test and refine them. The aim, to use Barnett's (2017) formulation, is a feasible utopia: ambitious enough to stretch what universities might do, grounded enough to be practicable. Each principle below operationalises one conceptual move from earlier sections: problem design and inquiry address adaptive expertise; transfer addresses cross-context capability; the ecological-ethical principle addresses ethics.

First, problem design. Problems in an LLM-era IS-PBL curriculum need to be built around genuinely wicked problems (Rittel and Webber, 1973): situations in which goals, constraints and relevant knowledge are not fully specifiable in advance, and in which multiple reasonable approaches exist, none straightforwardly correct. A problem that can be solved by prompting an LLM effectively has the wrong structure for educational purposes. This suggests a shift away from scenarios eliciting the application of known methods and towards scenarios built around the analysis of real organisational situations.

Second, the process of inquiry and the role of LLMs within it. LLMs should be integrated into IS-PBL not as tools for task completion but as generative interlocutors whose outputs are always provisional. Assessment cannot rest primarily on the quality of a student's final product, since it may be substantially LLM-assisted, but on the quality of the reasoning process, the capacity to evaluate and critique AI-generated inputs, and the ability to articulate and defend the judgments made. Process portfolios, reflective journals and structured peer critique become more important relative to conventional written assignments.

Third, transfer. A reconceptualised IS-PBL framework must be designed explicitly around the development of transfer capacity: the ability to take learning from one context and apply it, with appropriate adaptation, to another. If students are assessed only on performance within the bounded context of a course, we cannot know whether they have developed the flexible understanding adaptive expertise requires. Assessment that requires demonstrated transfer, and curricular arrangements that create continuity between educational and professional problem spaces, including internships and project partnerships, are important complements to classroom-based PBL.

Fourth, what Barnett (2017) calls the ecological university: the institution that takes seriously its relationships with broader social, environmental and epistemic contexts. An IS-PBL framework for the LLM era needs to foreground the ethical dimensions of AI use, including authorship, power, surveillance and environmental cost, not as a separate module but as an integral dimension of every problem scenario. The goal is IS graduates with sufficient stability of professional identity to engage with complex, contested situations without losing analytical rigour or ethical compass.

Together these principles constitute a departure from the competency-acquisition logic of IS2020. They do not reject the specific competencies IS2020 identifies, but situate competency development within a broader architecture that treats adaptive expertise, critical judgment and ethical reasoning as primary educational goals. This is empirically grounded: the Swedish job ads studies already suggested employers valued the dispositional and relational dimensions of IS graduate capability at least as highly as the technical ones; the LLM context makes this weighting more pronounced.

6. Discussion and Research Agenda

The argument developed here is conceptual and programmatic: it builds on prior empirical work, argues that the LLM context makes the curriculum-employer problem more urgent, and proposes PBL as a response. That empirical work is the task of the research programme of which this paper is the first instalment. Three claims separate what prior evidence has established from what the next phase will test: that the gap documented in 2017-2023 (Steen and Pierce, 2023) persists post-2022; that its character has shifted toward adaptive expertise and critical AI literacy; and that employers and recent graduates can articulate this shift in ways job-advertisement data alone cannot capture.

The next phase will be a survey of approximately one hundred organisations in the Öresund region, the cross-border metropolitan area encompassing Malmö and Copenhagen, designed to update the job ads evidence for the post-ChatGPT period and to gather direct employer perspectives on competencies they now regard as essential. The survey has two connected purposes: to test whether competency priorities have shifted since 2023 as predicted, and to assess whether employers are rethinking how they recruit and onboard IS graduates in light of LLM capabilities. The survey will be followed by in-depth interviews with a smaller number of key informants: recruiters, IS managers and recent IS graduates working in the region. The survey will use a paired-item design contrasting employer priorities for codified competencies (programming, data management) with priorities for adaptive capabilities (judgement, transfer, critical AI literacy), sampled from Platsbanken-registered employers and compared with the 2017-2023 ads data. The interviews are essential to this design: survey data and job-advertisement coding can reveal what employers prioritise, but only direct conversation with recruiters, IS managers and recent graduates can surface the reasoning, lived experience and tacit weightings that distinguish a post-LLM understanding of IS practice from its pre-LLM counterpart.

7. Conclusion

This paper has argued that IS education was already misaligned with employer expectations before the LLM inflection point, particularly in occupational coverage, personal dispositions, enterprise architecture and programming emphasis (Steen and Pierce, 2020, 2023). Since 2022, this misalignment has deepened qualitatively. LLMs have increased the value of adaptive expertise, critical judgment and knowledge transfer at the same time as IS education remains strongly shaped by codified competency frameworks. The Stanford AI Index 2026 gives this argument macro-level support, suggesting pressure on the entry-level roles IS programmes have traditionally prepared graduates to enter (Sajadieh et al, 2026).

In response, the paper has made a conceptual and programmatic case for PBL, reconceived through Savin-Baden's new constellations and Barnett's feasible utopias, as a pedagogical architecture for developing IS graduates capable of working critically and adaptively in LLM-augmented professional contexts. The argument engages the Kirschner-Sweller-Clark critique by advocating a scaffolded PBL design that preserves inquiry while managing cognitive load. This is an argument rather than a proof; the framework requires empirical testing and refinement. The research programme outlined here begins that work. The curriculum gap is not primarily a problem of content, but of educational design: what IS education is for, and how it can create the conditions for genuine learning rather than merely competency acquisition.

Ethics Declaration

This paper is conceptual and draws on already-published empirical work; no new human-subjects research was undertaken for it. No ethical clearance was therefore required.

AI Declaration

An AI assistant was used during the development of this paper to support structuring, prose review, text shortening and reference checking. All substantive arguments, interpretations and conclusions are the author's own. The author reviewed and verified all references and accepts responsibility for the final text.

References

Anthropic (2025) 'Anthropic education report: How university students use Claude', Anthropic. Available at: https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude (Accessed: 2 June 2026).

Barnett, R. (2000) *Realizing the University in an Age of Supercomplexity*. Buckingham: Society for Research into Higher Education and Open University Press.

Barnett, R. (2017) *The Ecological University: A Feasible Utopia*. London: Routledge.

Barnett, R. and Jackson, N. (eds.) (2019) *Ecologies for Learning and Practice: Emerging Ideas, Sightings and Possibilities*. London: Routledge.

Engeström, Y. (2001) 'Expansive learning at work: Toward an activity theoretical reconceptualization', *Journal of Education and Work*, 14(1), pp 133-156.

Hatano, G. and Inagaki, K. (1986) 'Two courses of expertise', in Stevenson, H., Azuma, H. and Hakuta, K. (eds.) *Child Development and Education in Japan*. New York, NY: W.H. Freeman, pp 262-272.

Hmelo-Silver, C.E., Duncan, R.G. and Chinn, C.A. (2007) 'Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006)', *Educational Psychologist*, 42(2), pp 99-107.

Jackson, N.J. (2013) 'The concept of learning ecologies', in Jackson, N.J. and Cooper, G.B. (eds.) *Lifewide Learning, Education and Personal Development*. Guildford: Lifewide Education.

Kasneci, E. et al (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', *Learning and Individual Differences*, 103, article 102274.

Kirschner, P.A., Sweller, J. and Clark, R.E. (2006) 'Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching', *Educational Psychologist*, 41(2), pp 75-86.

Kohen-Vacs, D., Usher, M. and Jansen, M. (2025) 'Integrating Generative AI into Programming Education: Student Perceptions and the Challenge of Correcting AI Errors', *International Journal of Artificial Intelligence in Education*, 35, pp 3166-3184.

Leidig, P. et al (2022) 'The ACM/AIS IS2020 Competency Model for Undergraduate Programs in Information Systems: A Joint ACM/AIS Task Force Report', *Communications of the Association for Information Systems*, 50, article 25.

Markauskaite, L. and Goodyear, P. (2017) *Epistemic Fluency and Professional Education: Innovation, Knowledgeable Action and Actionable Knowledge*. Dordrecht: Springer.

Pelgrim, E.A.M. et al (2022) 'Professionals' adaptive expertise and adaptive performance in educational and workplace settings: An overview of reviews', *Advances in Health Sciences Education, Vol 27, No. 5, pp 1245-1263.*

Rittel, H.W.J. and Webber, M.M. (1973) 'Dilemmas in a general theory of planning', *Policy Sciences*, 4(2), pp 155-169.

Sajadieh, S. et al (2026) *The AI Index 2026 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

Savin-Baden, M. (2014) 'Using problem-based learning: New constellations for the 21st century', *Journal on Excellence in College Teaching*, 25(3&4), pp 197-219.

Schön, D.A. (1983) *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books.

Sridharan, B. and Kinshuk (2024) 'Empowering the assessment of "wicked" competencies: A multi-level knowledge management architecture', in Li, S. (ed.) *Information Management: ICIM 2024*. Communications in Computer and Information Science, vol. 2102. Cham: Springer, pp 212-222.

Steen, O. and Pierce, P. (2020) 'Swedish undergraduate information systems curricula: A comparative study', in Siarheyeva, A., Barry, C., Lang, M., Linger, H. and Schneider, C. (eds.) *Advances in Information Systems Development: Information Systems Beyond 2020*. Cham: Springer Nature, pp 131-145.

Steen, O. and Pierce, P. (2023) 'Do IS curricular guidelines match employer expectations in Swedish IS job ads?', in *Information Systems Development: Organizational Aspects and Societal Trends (ISD2023 Proceedings)*. 31st International Conference on Information Systems Development, Lisbon, Portugal, 30 August–1 September 2023.

Topi, H. et al (2010) 'IS 2010: Curriculum guidelines for undergraduate degree programs in information systems', *Communications of the Association for Information Systems*, 26, article 18.

Wegerif, R. (2007) *Dialogic Education and Technology: Expanding the Space of Learning*. New York, NY: Springer.

Active Learning Map: A Novel Assimilation Guided Adaptive Learning Platform

Suhan Roy[1], Jaswant Jupalli Prabhas[1], Praseeda[1], Tulika Saha[1], Prasad Ram[2], Sushree Behera[1], Srinath Srinivasa[1]

[1]International Institute of Information Technology Bangalore, India
[2]Gooru Learning, California, USA

suhan.roy@iiitb.ac.in
jupalli.jaswant@iiitb.ac.in
praseeda@iiitb.ac.in
tulika.saha@iiitb.ac.in
pram@goorulearning.com
sushree.behera@iiitb.ac.in
sri@iiitb.ac.in

Abstract: Identifying individualized assimilation patterns of students can play an important role in personalizing learning interventions and improving their learning outcomes. While studying from similar course materials, learners frequently exhibit significant variability in how they internalize and prioritize core concepts. Recognizing this variability helps in designing meaningful personalized interventions, which in turn can lead to greater learner engagement and outcomes. An effective way to represent assimilation patterns is in the form of a high dimensional data structure called a Polyline that captures the differentiated way in which a given student has assimilated knowledge across different topics in the course. In this work, we extend this representational concept to create a model free Deep Reinforcement Learning Framework that can provide personalized recommendations to learners based on their Polylines. This is then visualized on a two-dimensional progression space called an Active Learning Map that enables the assimilation patterns to be visually displayed and interpreted. We cluster learners based on assimilation pattern and learning style using a Gaussian Mixture Model and train a Conditional Generative Adversarial Network (CGAN) as an Assimilation Estimator to simulate how their Polylines evolve. Using this CGAN as an environment, we train a Deep Q-Network (DQN) to discover individualized recommendation policies. Simulations demonstrate that this CGAN based environment enables the DQN to find highly effective policies using a very small sample of real student data. Furthermore, we find our DQN recommendation engine consistently outperforms a random policy, a coherent resource-matching heuristic, and a Proximal Policy Optimization (PPO) algorithm. Our proposed system functions as an autonomous intelligent agent within the Learning Map and acts as a guide, helping learners to visualize exactly where they are among a complex web of topics and lessons, as well as providing the best resources to help the learners close their knowledge gap.

Keywords: Assimilation, Polyline, Adaptive Learning Systems, Deep Reinforcement Learning, Active Learning Map

1. Introduction

One of the key elements of learning that drives outcomes is assimilation or how learners internalize their learning (Ryan, Connell and Grolnick, 1992). Students who have taken the same course often display wide differences in what they consider key learnings and takeaways from the course. When adaptive learning environments factor learner assimilation in the way their learning interventions are designed, it results in meaningful learning and retention (Ausubel, 2000). However, in online learning environments, capturing assimilation patterns and representing them in a way that is amenable to being processed by algorithms is a key challenge. Many online learning environments adopt adaptive learning architectures that orchestrate personalized learning instructions to meet the unique needs of individuals. The crux of a good adaptive learning system is to design a recommendation system that sequentially outputs customized recommendations, leading to the best suitable learning path for every single learner. Some of the early approaches to adaptive learning implemented an ‘’averagian’’ model where the class average was used as a benchmark to calibrate individual learners. However, later work recognizes learner variability as a rule, rather than an exception, and emphasizes the need for understanding individual variability while designing learning interventions. Assimilation or internalization of learning is a latent phenomenon that is not directly observable. In this regard, constructivist learning (Bevevino, Dengel and Adams, 1999), where learners are required to periodically construct their own knowledge and theories based on what they have learned, offers a glimpse into underlying assimilation patterns. Using this approach, a representation called “Polyline” was introduced for representing a learner's individualized assimilation pattern (also called a “jagged profile'') across the different topics representing a course (Praseeda et al., 2025). While Polylines provide a computationally sound representation to reason with assimilation, there

is still a need to factor this into recommendation engines that can provide personalized learning interventions. This paper addresses the problem of designing personalized pedagogical interventions using high-dimensional learner Polylines to capture the non-linear, highly individualized nature of knowledge assimilation. While traditional predictive neural network models (Li et al., 2023) tend to regress to the mean and discrete frameworks like Bayesian Knowledge Tracing (Subramanian and Mostow, 2021) cannot map to continuous, high-dimensional spaces, we model these trajectories by training a Conditional Generative Adversarial Network (CGAN) as an Assimilation Estimator Model. To ground our work in real world educational dynamics, we introduce ATLAS (Assimilation Tracking from Lecture Assigned Summaries), our dataset consisting of independent, constructivist student summaries collected post-lecture over a graduate level course "Advances in NLP". We construct Polylines from these summaries to provide a high-resolution, longitudinal trace of continuous knowledge growth. To capture shifting assimilation patterns, a Gaussian Mixture Model (GMM) clusters students based on learning styles and academic inclinations. We then develop a model-free Deep Q-Network (DQN) engine to recommend personalized learning resources that maximize learning gains. Simulated evaluations against random and rule-based heuristic baselines demonstrate the policy's efficiency. Finally, we deploy the trained DQN model onto a two-dimensional progression space called an Active Learning Map, enabling visually interpretable tracking of student assimilation patterns and resource recommendations.

2. Related Work

2.1 Assimilation Patterns and Mental Models

Previously researchers have tried to map the assimilation of students in terms of understanding and representing their internalization of knowledge by constructing mental models. Constructing mental models of students has been explored in key disciplines such as mathematics, chemistry (Coll and Treagust, 2001), and in studies on how learners internalize abstract ideas about non-living entities (Stavy, 1988). However, there exist no such studies to understand the assimilation pattern of a large number of learners over an entire structured course and how assimilation of learners can change over the consumption of new topics and diverse learning materials.

2.2 Representation of Learner Profiles

A learner's profile represents their progress in terms of competencies or proficiencies across all the topics included in the course. Several previous works have chosen to represent learner's proficiencies as a vector of binary latent traits. Each latent trait here represents complete mastery in a particular topic as a discrete binary signal. Such signals are often erroneous and provide incomplete context about the learner's proficiency, as competency over a topic or skill can often be in terms of multiple levels of mastery and there is continuous room for improvement. Representation of these learning traits as continuous values provide more detail and granularity in terms of identifying where learners can improve. Item Response Theory (Lord, 1980) and its variants (Ackerman et al., 2003) use a continuous scale vector representation that has been widely used in designing Intelligent Tutoring Systems (Burns and Capps, 2013) that help learners to improve their skills and reach their target outputs (van der Linden, 2005). The continuous vector representation is a powerful idea because it translates abstract, qualitative human learning into a rigorous, m-dimensional space that we can navigate and optimize with Reinforcement Learning.

2.3 Reinforcement Learning in Education

Reinforcement Learning (RL) is a paradigm of Machine Learning that has been rapidly growing over the past few years and has found key applications in real-world problems. In contrast to supervised and unsupervised learning, RL involves an agent that learns through direct contact with its environment. The agent takes consecutive actions and earns rewards from its environment as feedback after every action and thus learns to optimize its path. Thus, in order to maximize its overall cumulative sum of expected rewards, the agent learns by acquiring knowledge and experience rather than learning from instructions. RL has found a wide variety of applications in Education and particularly in e-learning platforms as a framework for personalised education to enhance learning outcomes. Several RL methods, namely Markov Decision Process (MDP), POMDP, Markov Chain Monte Carlo (MCMC), and the widely popular Deep RL Framework have been successfully applied in learning-based scenarios (Fahad Mon et al., 2023).

2.4 Adaptive Learning Systems

Personalised recommendations of educational activities for learners enrolled in a course to improve their learning goals have been deeply explored in literature related to Intelligent Tutoring Systems (Chen et al., 2017). Several works have adopted the Contextual Multi-Armed Bandits framework (Manickam et al., 2017) for personalised resource recommendation owing to their ability to balance exploration (exploring new activities) and exploitation or recommending effective known content (Nguyen et al., 2023). On considering the state of the learner as a dynamic and evolving entity, researchers have chosen to model the problem as a Markov Decision Process (MDP). Researchers have used popular Deep RL algorithms such as DQNs and Actor-Critic to learn Instructional policies that can optimize for long-term rewards in terms of Higher assessment scores or Improved learning output.

3. Methodology

3.1 Jagged Learner Profiles and Active Learning Map

In our research, we extend the Polyline model to represent the current assimilation state of learners as a Jagged profile, which captures learning outcomes across multiple dimensions (Praseeda et al., 2025). A Learning Entity refers to any entity involved in the course, including learners, learning resources, quizzes or assessments, instructors, and Teaching Assistants (TA). An *LE* also includes a set of basis topics ***T*** as:

$$\boldsymbol{T} = \{T_1, T_2, T_3, \ldots, T_n\}$$

Every learning entity in the course is represented as a polyline across n fundamental syllabus topics, as illustrated in Figure 1.

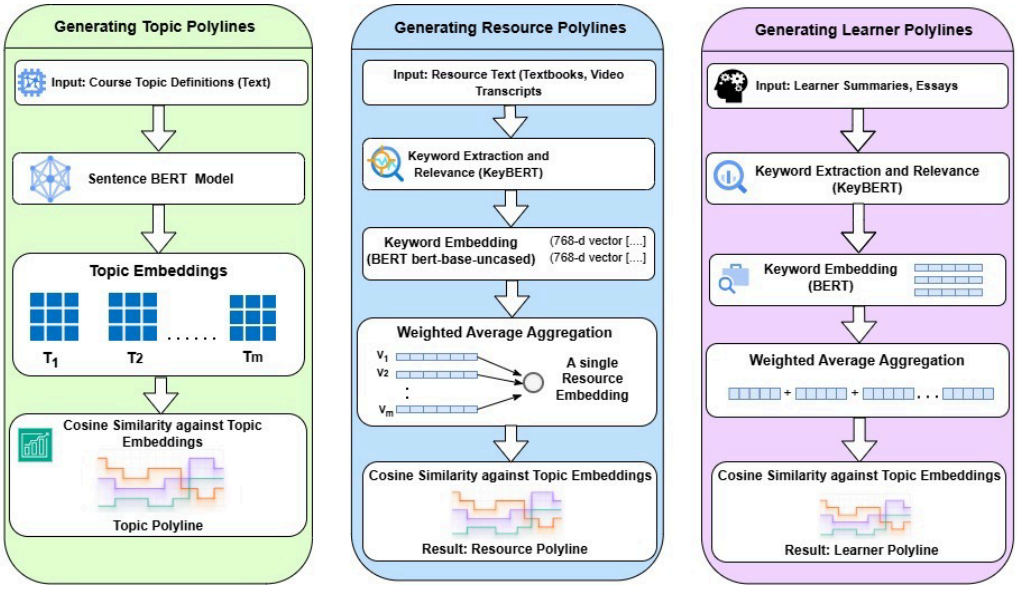


Figure 1: Creating Topic, Resource and Learner Polyline

Polylines in the Progression space follow the covering property. The covering property dictates that a specific Learning Entity *A* covers a distinct Learning Entity *B* provided that each individual dimensional coordinate within the Polyline of *A* maintains a value at least as high as its respective counterpart in the Polyline of *B*. The Highline for a group of students represents the highest achievable upper bound over the set of Basis Topics included in the course. The Highline of a course serves as a target vector that represents the highest proficiencies achievable over the Basis Topics for a new student who is enrolled in the course. The Highline of a course can be traced from the historical polylines of the learners enrolled in the course, or it can be set by the instructor or Teaching Assistant. The Active Learning Map functions as an interpretable two-dimensional space designed to visually track and navigate student progression throughout a course. At its core, the framework relies on high-dimensional vector representations called polylines, which quantify the mastery level of learning resources and individual learners across a fixed set of foundational concepts. To project these multi-dimensional vectors onto the two-dimensional map, we implement the radial mapping algorithm (Praseeda et al., 2025). In this geometric layout, each fundamental course topic is represented as a distinct radial axis emanating from a central origin.

The magnitude of a specific topic's mastery in the polyline vector determines the distance out along its corresponding axis and the algorithm computes a weighted sum of these directional components to compress the high-dimensional knowledge profile into a single, interpretable coordinate point. An intelligent recommendation engine then analyses a student's current position on this map to dynamically prescribe the next most effective learning resource, optimizing the trajectory toward comprehensive topic mastery.

3.2 An Assimilation Learning Dataset

Traditionally, Intelligent tutoring systems have relied on discrete scoring metrics such as lesson wise assessments and binary skill labels to estimate a learner's proficiency profile. While these models successfully use scalar scores to make recommendations, they inherently reduce the continuous, multidimensional aspect of knowledge assimilation into a single quantity. To address this gap, we introduce ATLAS (Assimilation Tracking from Lecture Assigned Summaries), a dataset curated from students enrolled in the Advances in NLP course that provides a granular, longitudinal trace of acquired knowledge. By capturing individual summaries after every lecture and resource interaction, our dataset records the progressive assimilation of technical concepts across the entire student cohort. The course curriculum includes 19 different Basis Topics and there is one learning resource for every topic. From the student summaries, we construct 19-dimensional Learner Polylines which serve as a backbone for understanding how the students internalize the curriculum over time.

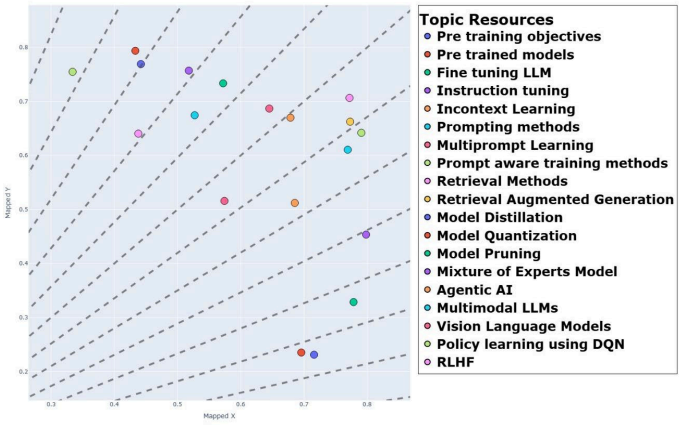


Figure 2 : Active Learning Map: High-dimensional Polyline representation projected into 2D progression space using the Radial Mapping Algorithm.

3.3 Formulating Adaptive Learning as a Markov Decision Process

The typical defining feature of DRL is to use deep learning to approximate reinforcement learning's value function and solve high-dimensional MDPs. The key components of our MDP are as follows:

1. State *s :* We define the current Polyline *PL* describing the learner's assimilation profile as the state.
2. Action *a :* Let the learning materials available in the adaptive learning system be indexed by 1,2,....*L*. The action space is $A = \{1,2,...,L\}$ and the action *a* in the adaptive learning system is the index of recommended learning resource which is a discrete value.
3. *Reward R :* To minimize the pedagogical steps required to reach course mastery, our reward function is strictly defined as the negative L_1 distance between the current Learner Polyline and the target Course Highline.

$$R(PL,a,HL) = - ||HL-PL||_1$$

By penalizing the agent at each step until the polyline aligns with the target highline, the environment frames routing as a shortest-path problem thus forcing the agent to minimize the knowledge gap in the fewest possible steps.

3.4 Clustering Learners based on Assimilation Patterns

To dynamically categorize and update learner profiles based on their Polyline, we use a GMM model. The objective of the clustering algorithm is to group the learners based on their interests and style of learning, based on information from their Polyline.

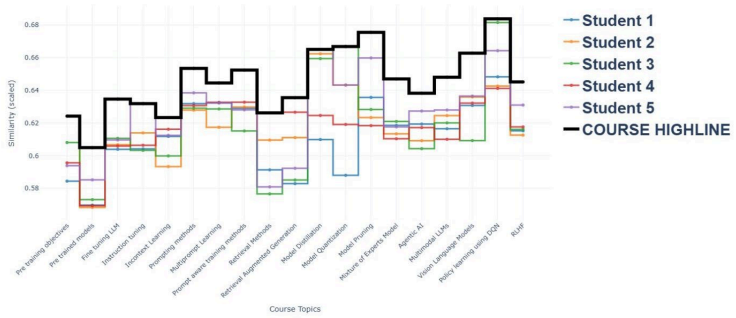


Figure 3: The Polyline representing the Highline for the course "Advances in NLP".

3.5 Conditional GAN as an Assimilation Estimator

Extensive research has demonstrated that neural networks are highly effective transition model estimators, capturing the complex, non-linear mappings required to precisely track and predict state transitions across learning trajectories (Li et al., 2023). To capture the non-deterministic nature of this progression, we employ a Conditional Generative Adversarial Network (CGAN) to estimate a probability distribution over potential future states rather than a single deterministic outcome (Mirza and Osindero, 2014). Trained on sequential logs from the ATLAS dataset, the network utilizes a supervised triplet format: an input polyline (current state), a condition (resource ID), and a label polyline (next state). By using the resource ID as a conditioning variable, the generator (G) learns to recognize underlying progression patterns to synthesize a targeted next polyline directly from the previous polyline based on the specific resource visited. Meanwhile, the discriminator (D) acts as a critic, comparing the generated paths against real student data from the ATLAS dataset to ensure the predicted steps are realistic and accurate.

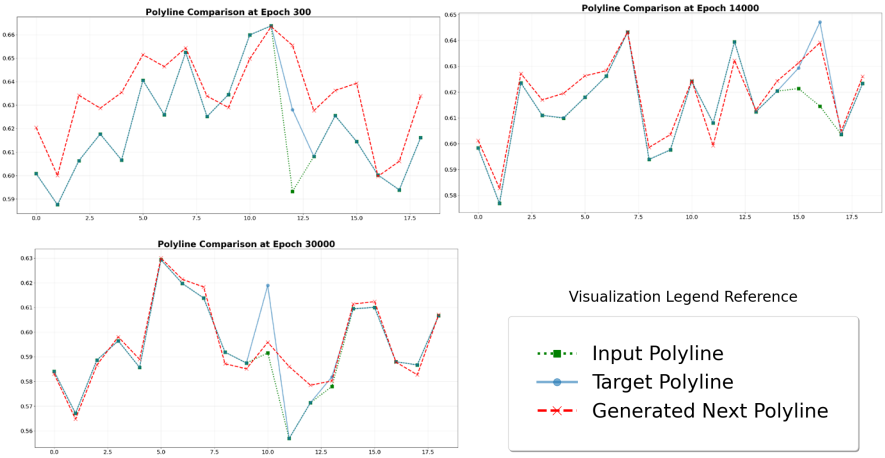


Figure 4: Our CGAN Model effectively learnt to understand and accurately predict assimilated knowledge states from our learner dataset.

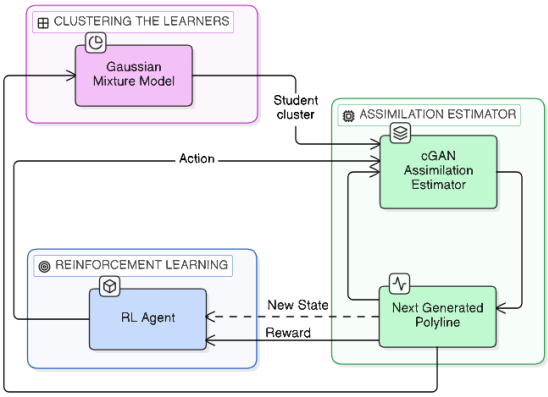


Figure 5: Diagram illustrates the proposed setup and Methodology to train our RL models.

3.6 Proposed Formulation to Train RL Algorithm

To discover novel pedagogical paths rather than relying purely on familiar sequences, the framework leverages distinct exploration mechanics that mirror instructional experimentation. The DQN engine employs an ϵ-greedy strategy, starting with high randomized exploration (ϵ = 0.9) to aggressively test diverse resource combinations before systematically decaying to a minimum floor (ϵ = 0.05) as it converges on a highly targeted, effective recommendations. Conversely, PPO explores inherently through a stochastic policy, sampling actions from a broad probability distribution over available resource. Early in training, this distribution is high-entropy, forcing the model to unroll widely divergent 50-step student journeys. As the engine accumulates compounding rewards, the policy sharpens around the optimal sequences. This robust exploration prevents both models from falling into local optimum traps such as repeating a predictable, basic sequence and forces them to discover hidden structural shortcuts across the continuous polyline space to find accelerated trajectories that minimize the knowledge gap in fewer steps.

4. Simulations Overview

To evaluate the performance of our trained models, we conducted 1000 independent simulation episodes per policy over a fixed 50 step horizon. At the beginning of each episode, the learner's initial knowledge state is represented as an all zero Polyline paired with a randomly selected starting cluster ID. To test the agents' capacity to optimize paths from arbitrary configurations, the first action at step (t=0) is forced to a random selection from the initial resource subset [0, 1, 2]. For all subsequent time steps, the Polyline state evolves dynamically based on policy recommendations. The PPO agent samples actions directly from its optimized stochastic policy, while the DQN agent follows an ϵ-greedy policy with a minimal exploratory baseline (ϵ = 0.1). We retained this nonzero exploration baseline during simulations to understand the performance of the engine to prepare for real world deployment on an e-learning platform where it must handle new learners. In production, real human behaviour and cognitive profiles inevitably deviate from synthetic simulations. A purely deterministic deployment risks faltering when encountering out-of-distribution (OOD) student states, potentially trapping the engine in suboptimal loops. Throughout these test trajectories, the agents interacted continuously with the independent CGAN environment simulator, which updated cluster profiles via the GMM and strictly enforced the unidirectional covering property to prevent knowledge regression. To measure policy efficiency, we tracked Total sum of cumulative negative L_1 distance rewards across each complete trajectory, directly quantifying how effectively each engine minimized the remaining knowledge gap against the target Course Highline.

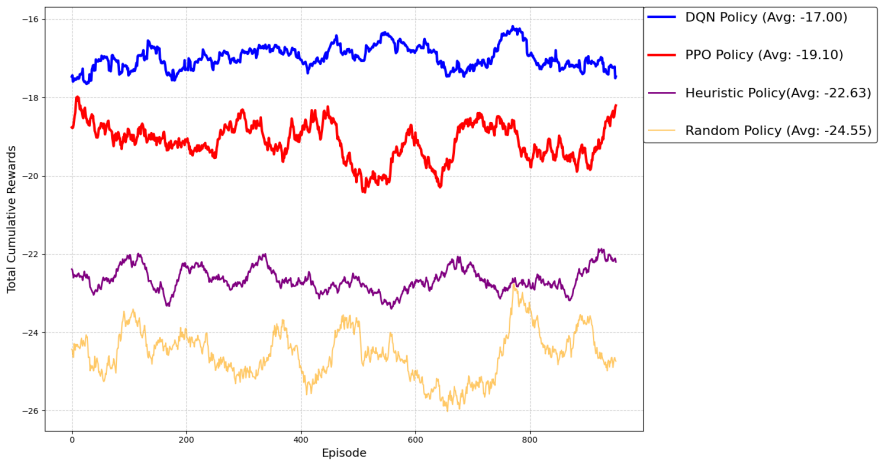


Figure 6: Smoothed rewards as learning gains over DQN, PPO, Heuristic and Random policies.

5. Results and Discussion

We evaluate our trained reinforcement learning policies against two baselines, an L_1 distance based heuristic and a Random recommendation policy. The heuristic policy selects semantically proximate resources to maintain coherence and prevent cognitive jumps, and the random recommendation policy serves as a robust baseline in Intelligent Tutoring Systems (ITS) where any curricular exposure can trigger baseline gains. Performance is

quantified using the cumulative sum of negative rewards, which measures the aggregate knowledge gap between the evolving Polyline and the target Course Highline. As this reward moves closer to zero (becomes less negative), it represents a policy that successfully maximizes learning gains by routing the student toward the Highline within the fixed horizon. A reward closer to zero (less negative) indicates an optimal, accelerated learning path that minimizes instructional friction and reaches the Highline in fewer steps. The DQN emerged as the best-performing policy, navigating these dynamics with the highest efficiency.

6. Active Learning Map

The student interface of the 2D Active Learning Map provides a comprehensive, interactive visualization of the learner's educational trajectory. The environment integrates a diverse array of

learning resources and instructional materials, which are spatially distributed across the map. The platform actively displays the learner's current positional coordinates and historical activity log, dynamically updating their state in real time as they engage with the curriculum through actions such as summarizing their learnings. Throughout their progression, students can continuously monitor their evolving Polyline, a visual representation that captures their continuous knowledge assimilation over time.

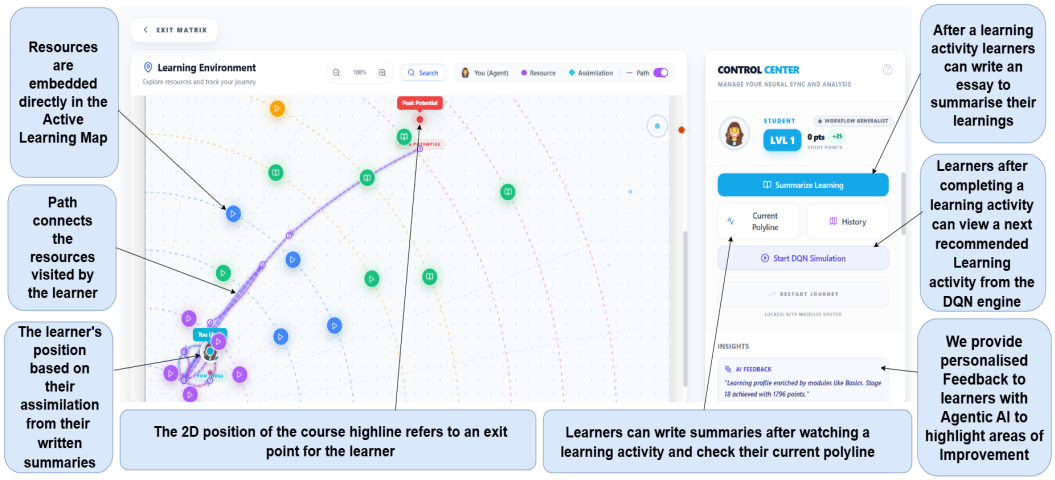


Figure 7: Our DQN Recommendation Engine can visualise and perform recommendations as a learning plan to learners on the Active Learning Map space.

7. Conclusion and Future Work

Understanding and representing assimilation is fundamental to fostering holistic learner transformation. In this work, we introduced a data driven framework for personalized adaptive learning centred on the Active Learning Map. By departing from rigid, traditional structures, our approach prioritizes holistic learning outcomes and captures the complex nature of knowledge assimilation that Adaptive Learning Systems overlook. However, while simulations validate the framework's technical soundness, it has not yet been tested with real learners. Consequently, our immediate future work focuses on integrating the Active Learning Map into a live classroom environment to empirically measure its impact on student engagement and learning outcomes. To capture the full cognitive reality of the student experience, we aim to expand the framework's capabilities by incorporating diverse learning artifacts such as coding assignments and quizzes and scaling the system across multiple courses. Ultimately, accurately modelling real-world assimilation dynamics will rely on collecting and analysing a larger corpus of student data to build increasingly robust machine learning models.

Ethics Declaration

This study was conducted in accordance with the ethical standards of research. All participants were informed about the purpose, and all the data was collected under the consent and supervision of the Course Instructor and Teaching Assistant.

AI Declaration

We used Gemini Pro to improve and refine sentence structure and to suggest alternative phrasings. All AI-generated suggestions were critically reviewed and edited by the authors, and the final content reflects the authors' own analysis and conclusions.

References

Ackerman, T.A., Gierl, M.J. and Walker, C.M., 2003. Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*, 22(3), pp. 37–51. doi: 10.1111/j.1745-3992.2003.tb00136.x.

Ausubel, D.P., 2000. Assimilation theory in meaningful learning and retention processes. In: D.P. Ausubel, ed. *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Springer, pp. 101–145.

Bevevino, M.M., Dengel, J. and Adams, K., 1999. Constructivist theory in the classroom internalizing: Concepts through inquiry learning. *The Clearing House*, 72(5), pp. 275–278.

Burns, H.L. and Capps, C.G., 2013. Foundations of intelligent tutoring systems: An introduction. In: H.L. Burns and J.W. Parlett, eds. *Foundations of intelligent tutoring systems*. New Orleans: Psychology Press, pp. 1–19.

Chen, Y., Li, X., Liu, J. and Ying, Z., 2017. Recommendation System for Adaptive Learning. *Applied Psychological Measurement*, 42(4), pp. 243–257. doi: 10.1177/0146621617697959.

Coll, R.K. and Treagust, D.F., 2001. Learners' Mental Models of Chemical Bonding. *Research in Science Education*, 31(3), pp. 357–382. doi: 10.1023/A:1013159927352.

Fahad Mon, B., Wasfi, A., Hayajneh, M., Slim, A. and Abu Ali, N., 2023. Reinforcement Learning in Education: A Literature Review. *Informatics*, 10(3), p. 74. doi: 10.3390/informatics10030074.

Li, X., Xu, H., Zhang, J. and Chang, H.H., 2023. Deep Reinforcement Learning for Adaptive Learning Systems. *Journal of Educational and Behavioral Statistics*, 48(2), pp. 220–243. doi: 10.3102/10769986221129847.

Lord, F.M., 1980. *Applications of Item Response Theory to Practical Testing Problems*. 1st ed. New York: Routledge. doi: 10.4324/9780203056615.

Manickam, I., Lan, A.S. and Baraniuk, R.G., 2017. Contextual multi-armed bandit algorithms for personalized learning action selection. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA, USA, 5-9 March 2017. IEEE Press, pp. 6344–6348. doi: 10.1109/ICASSP.2017.7953377.

Mirza, M. and Osindero, S., 2014. *Conditional Generative Adversarial Nets*. arXiv preprint arXiv:1411.1784. Available at: https://arxiv.org/abs/1411.1784.

Nguyen, D., Howard, L., Yan, H. and Lin, O., 2023. Adaptive Sequence Learning: Contextual Multi-armed Bandit Approach. In: *2023 IEEE International Conference on Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing, Cloud and Big Data Computing, Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. Abu Dhabi, United Arab Emirates, 14-17 November 2023. IEEE, pp. 0751–0756. doi: 10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361498.

Praseeda, Srinivasa, S. and Ram, P., 2025. Learning map: Representing and reasoning over diverse assimilation patterns of learners. *Next Research*, 2(4), p. 100804. doi: 10.1016/j.nexres.2025.100804.

Ryan, R.M., Connell, J.P. and Grolnick, W.S., 1992. When achievement is not intrinsically motivated: A theory of internalization and self-regulation in school. In: A.K. Boggiano and T.S. Pittman, eds. *Achievement and motivation: A social-developmental perspective*. Cambridge: Cambridge University Press, pp. 167–188.

Subramanian, J. and Mostow, J., 2021. Deep Reinforcement Learning to Simulate, Train, and Evaluate Instructional Sequencing Policies. In: *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*. Paris, France, 29 June - 2 July 2021. International Educational Data Mining Society, pp. 696–700. Available at: https://api.semanticscholar.org/CorpusID:236439662 [Accessed 21 May 2026].

tavy, R., 1988. Children's conception of gas. *International Journal of Science Education*, 10(5), pp. 553–560. doi: 10.1080/0950069880100508.

van der Linden, W.J., 2005. Item Response Theory. In: K. Kempf-Leonard, ed. *Encyclopedia of Social Measurement*. New York: Elsevier, pp. 379–387. doi: 10.1016/B0-12-369398-5/00452-7.

Spatial Reasoning Outcomes in Technology-Enhanced Geometry Teacher Education

Angelika Schmid[1], Tomas Barot[1], Lilla Korenova[2] and Zsolt Lavicza[3]
[1]Faculty of Education University of Ostrava, Ostrava, Czech Republic
[2]Faculty of Education Comenius University Bratislava, Bratislava, Slovakia
[3]School of Education Johannes Kepler University Linz, Linz, Austria

angelika.schmid@osu.cz
tomas.barot@osu.cz
korenova@fedu.uniba.sk
zsolt.lavicza@jku.at

Abstract: Spatial reasoning is widely recognised as an essential competence for mathematics teachers and a significant predictor of achievement across STEM disciplines. However, less attention has been devoted to how these abilities can be systematically supported through technology-enhanced learning environments in higher education, particularly within teacher education. This study examined spatial reasoning outcomes in a technology-enhanced geometry course designed for prospective mathematics teachers. The research focused on comparing current learning outcomes with delayed performance following previous participation in the course. A quasi-experimental design was employed with two university student cohorts enrolled in geometry teacher education courses. One cohort had completed the course in the previous academic year, while the second cohort was involved in the current implementation during the ongoing semester. The learning environment was based on the Geometry Teaching Model, which integrated manual construction activities with digital technologies, particularly GeoGebra. Four spatial reasoning tests were administered during the semester. The tests assessed three components of spatial reasoning: mental rotation, spatial orientation, and spatial visualisation. Data were analysed using descriptive and non-parametric statistical methods. Students participating in the current course implementation achieved significantly stronger results in several complex tasks requiring reconstruction of three-dimensional objects from two-dimensional representations, interpretation of orthographic views, and advanced spatial synthesis. Students from the previous cohort performed comparably in simpler orientation tasks, suggesting partial maintenance of basic spatial skills over time. However, their performance was weaker in more demanding visualisation tasks. The findings suggest that performance in advanced spatial reasoning tasks may be particularly sensitive to the continuity of structured educational support. Sustained technology-enhanced geometry learning may therefore provide valuable opportunities for supporting complex visual-spatial competences in teacher education. The study contributes to current discussions on e-learning design, competence sustainability, and the preparation of future mathematics teachers in higher education.

Keywords: Spatial Reasoning, Teacher Education, Geometry Education, GeoGebra, Technology-Enhanced Learning.

1. Introduction

Spatial reasoning is widely recognised as a fundamental cognitive competence that plays a crucial role in mathematics education and predicts success across science, technology, engineering, and mathematics (STEM) disciplines (Ramful, Lowrie and Logan, 2017; Uttal et al, 2013). It enables learners to mentally manipulate objects, interpret multiple representations, and understand geometric relationships that are essential for solving mathematical problems. As geometry relies heavily on the ability to visualise, transform, and reason about spatial structures, the development of spatial reasoning has become an important objective of mathematics teacher education. Future teachers are expected not only to develop these competences themselves but also to create learning environments that foster spatial reasoning among their pupils.

Recent advances in digital technologies have created new opportunities to support the development of spatial reasoning through interactive geometry learning environments. Dynamic geometry software enables learners to construct, manipulate, and explore geometric objects while seamlessly moving between two-dimensional and three-dimensional representations. These environments encourage active exploration, immediate visual feedback, and inquiry-based learning, making them particularly suitable for developing spatial reasoning within mathematics teacher education.

Although previous research has demonstrated the educational value of technology-enhanced geometry learning, considerably less attention has been devoted to the long-term sustainability of spatial reasoning following technology-supported instruction. Moreover, relatively little is known about the diagnostic potential of different geometry-specific assessment tasks and their ability to distinguish between varying levels of spatial reasoning performance. Understanding which types of tasks are most sensitive to differences in spatial

reasoning is important not only for evaluating educational interventions but also for designing assessment instruments that capture the complexity of spatial thinking.

To address these issues, this study compares the spatial reasoning performance of two cohorts of prospective mathematics teachers enrolled in the same technology-enhanced university geometry course. One cohort was assessed during the course, whereas the second cohort completed the same assessment approximately one year after instruction, enabling the comparison of current and delayed spatial reasoning outcomes. In addition to comparing the two cohorts, the study investigates which characteristics of geometry-specific assessment tasks most effectively differentiate students' spatial reasoning performance. The instructional context was based on the Geometry Teaching Model (GTM), a technology-enhanced instructional framework integrating digital and manual geometry activities, which is described in Section 2.2. The study addresses the research questions presented in Section 3.1.

2. Theoretical Background

This section provides the theoretical framework underpinning the study. It first examines the concept of spatial reasoning and its key components in the context of mathematics education. Subsequently, it reviews the contribution of technology-enhanced geometry learning to the development of spatial reasoning and introduces the GTM, which served as the instructional context for the reported research.

2.1 Spatial Reasoning

Spatial reasoning is generally understood as the ability to generate, manipulate, transform, and interpret mental representations of spatial relationships between objects and within space. It represents a multidimensional cognitive competence that underpins geometric thinking, supports mathematical problem solving, and contributes to achievement across STEM disciplines. Beyond mathematics education, spatial reasoning is recognised as an important predictor of later academic and professional success, highlighting the need for its systematic development throughout education (Ramful, Lowrie and Logan, 2017; Uttal et al, 2013).

Although several theoretical classifications of spatial reasoning have been proposed, this study adopts the framework developed by Ramful, Lowrie and Logan (2017), which distinguishes three complementary constructs of spatial reasoning that are particularly relevant to school mathematics: mental rotation, spatial orientation, and spatial visualisation. This framework was selected because it reflects the types of spatial tasks commonly encountered in geometry and provides a useful basis for educational assessment.

The three components represent distinct but closely interconnected aspects of spatial reasoning and together capture the range of spatial processes commonly required in geometry education.

- **Mental rotation** refers to the ability to mentally rotate two- or three-dimensional objects while maintaining their structural properties.
- **Spatial orientation** involves mentally changing one's viewpoint or determining the relative positions of objects from different perspectives.
- **Spatial visualisation** encompasses more complex multi-step manipulations of spatial information, often requiring the coordination of multiple representations and successive mental transformations.

Because spatial reasoning comprises multiple interrelated cognitive processes, its assessment requires tasks that capture different forms of spatial thinking rather than relying on a single type of spatial operation (Lowrie et al, 2020). Geometry tasks involving object transformations, orthographic representations, cross-sections, and transitions between two-dimensional and three-dimensional representations provide valuable opportunities for diagnosing different aspects of spatial reasoning. Accordingly, the assessment design of the present study incorporates geometry tasks targeting these complementary components of spatial reasoning, enabling a more comprehensive evaluation of students' spatial reasoning abilities.

2.2 Technology-Enhanced Geometry Learning

Technology-enhanced learning has become an integral part of mathematics education, particularly in geometry, where digital technologies offer new opportunities to support conceptual understanding, mathematical exploration, and the development of higher-order cognitive skills. In this context, interactive learning environments enable learners to construct, manipulate, and investigate geometric objects in ways that are difficult to achieve using traditional static representations alone. These environments encourage active

exploration and inquiry while supporting connections between visual and formal mathematical representations (Duval, 2006).

Among the available digital tools, dynamic geometry software has received particular attention because it enables learners to explore geometric relationships through direct manipulation, immediate visual feedback, and seamless transitions between two-dimensional and three-dimensional representations. GeoGebra, one of the most widely used dynamic mathematics environments, facilitates experimentation with geometric constructions while promoting spatial reasoning through interactive visualisation and multiple representations. Previous research suggests that dynamic geometry environments can support students' conceptual understanding, engagement, and geometric reasoning (Hohenwarter and Lavicza, 2007; Sinclair, Holt and Heyd-Metzuyanim, 2016; Weigand, Hollebrands and Maschietto, 2025).

Building on these principles, the GTM was developed as a technology-enhanced instructional framework for university geometry courses in prospective mathematics teacher education (Schmid and Korenova, 2024). The model integrates dynamic geometry software with manual geometric constructions and complementary learning activities to support the systematic development of spatial reasoning while fostering conceptual understanding of geometry. Rather than relying exclusively on digital technologies, the GTM emphasises purposeful transitions between digital and manual representations, enabling students to approach geometric problems from multiple perspectives. The present study was conducted within this instructional framework.

2.3 Research Questions

To investigate prospective mathematics teachers' spatial reasoning performance through geometry-specific assessment tasks implemented within the GTM, the study addressed the following research questions:

- **RQ1:** How did the quasi-control and quasi-experimental cohorts differ in their performance on geometry-specific assessment tasks targeting different components of spatial reasoning?
- **RQ2:** Which characteristics of geometry-specific assessment tasks appear to be most informative for differentiating prospective mathematics teachers' spatial reasoning performance?

3. Methodology

3.1 Research Design

This study employed a quasi-experimental comparative design involving two cohorts of prospective mathematics teachers enrolled in the *Synthetic Geometry II* course at the University of Ostrava. The first cohort consisted of students who had completed the course during the previous academic year and were assessed approximately one year after instruction, allowing the examination of delayed performance. The second cohort comprised students enrolled in the current course, who completed the same assessment tasks during the semester as part of their regular coursework.

Both cohorts completed an identical set of geometry-specific assessment tasks designed to represent the three components of spatial reasoning: mental rotation, spatial orientation, and spatial visualisation. The assessments were administered under identical conditions, using the same instructions and time limits, to ensure the comparability of the collected data. Each assessment task targeted a specific geometry-related cognitive process associated with one of the three spatial reasoning components.

Beyond comparing students' learning outcomes, the present study draws on the quasi-experimental dataset to examine the characteristics of individual geometry-specific assessment tasks and the extent to which different task types differentiated spatial reasoning performance.

3.2 Participants

The study involved prospective mathematics teachers from two successive cohorts enrolled in the *Synthetic Geometry II* course at the University of Ostrava. The quasi-control and quasi-experimental course cohorts comprised 24 and 16 enrolled students, respectively. Participation in the study was voluntary, and the number of participants varied across the four test administrations. In the quasi-control cohort, Tests 1–4 were completed by 20, 20, 20, and 16 students, respectively, while the corresponding numbers in the quasi-experimental cohort were 16, 13, 10, and 15. Both cohorts followed the same curriculum and were taught by the same instructor. All collected data were anonymised prior to analysis.

3.3 Instructional Context

The study was conducted within the Synthetic Geometry II course for prospective mathematics teachers at the University of Ostrava. The course was delivered using the GTM, a technology-enhanced instructional approach integrating GeoGebra, manual geometric constructions, inquiry-based learning, and collaborative problem-solving activities. A detailed description of the model is provided in Section 2.2. The instructional activities were designed to support the development of the three components of spatial reasoning addressed in this study – mental rotation, spatial orientation, and spatial visualisation – through a sequence of progressively more complex geometry tasks. Digital technologies were used alongside traditional geometric constructions to encourage students to connect multiple representations and justify their reasoning. The assessment tasks analysed in this study were administered independently of regular course assessment and were used to examine students' spatial reasoning performance across different geometry-specific task types.

3.4 Assessment Framework

The geometry-specific assessment framework was developed as part of the broader design-based research process accompanying the iterative refinement of the GTM. The assessment tasks were designed not only to evaluate students' spatial reasoning outcomes but also to monitor their development across successive iterations of the instructional model and to provide evidence supporting its continuous refinement.

Table 1: Overview of the Geometry-Specific Assessment Framework

Test	Task ID	Task	Geometry Context	Representation(s) Involved	SR Compo-nent	Primary Cognitive Demand	Max Score
Test 1	T1.1	Cube nets	Cube nets	2D net → 3D cube	SV	Mental reconstruction of a three-dimensional object	6
	T1.2	Rolling die	Rolling cube	Dynamic 3D transformations	MR	Mental simulation of sequential rotations	30
	T1.3	Cube–net matching	Cube and net	2D ↔ 3D	SV	Matching corresponding representations	4
Test 2	T2.1	3D oblique view → Floor plan	Orthographic projections	3D → 2D	SO	Interpretation of spatial relationships from a three-dimensional representation	2
	T2.2	Floor plan → 3D oblique view	Orthographic projections	2D → 3D	SO	Mental reconstruction of a three-dimensional arrangement from a plan view	3
	T2.3	3D oblique view → Three orthographic views	Orthographic projections	3D → multiple 2D views	SO	Coordination of multiple orthographic representations	6
	T2.4	Three orthographic views → 3D oblique view	Orthographic projections	Multiple 2D views → 3D	SO	Mental reconstruction from multiple projections	3
Test 3	T3.1	Perspective viewpoint	Perspective scene	3D → viewpoint selection	SO	Perspective taking	1
	T3.2	Maze orientation	Spatial navigation	Perspective ↔ top view	SO	Mental change of viewpoint	1
	T3.3	House net	Polyhedron net	2D net → 3D object	SV	Mental reconstruction of a folded object	1
	T3.4	Prism cross-section	Cross-sections	3D object → 2D net	SV	Mental mapping of planar intersections onto a net	1
Test 4	T4.1	Odd-one-out	Block structures	3D objects	MR	Mental reflection and rotation	3
	T4.2a	Cube	Cross-sections	3D object → section	SV	Prediction of planar intersections	4

Test	Task ID	Task	Geometry Context	Representation(s) Involved	SR Compo-nent	Primary Cognitive Demand	Max Score
		cross-section					
	T4.2b	Complex cross-section	Cross-sections	3D object → section	SV	Prediction of complex planar intersections	10

In the present study, the framework is examined from a different perspective, focusing on the characteristics of individual task types and their ability to differentiate prospective mathematics teachers' spatial reasoning performance.

The assessment tasks were researcher-developed and informed by the classification principles of the Spatial Reasoning Instrument (SRI) (Ramful, Lowrie and Logan, 2017; SERC, 2022), but did not constitute a standardised administration of the SRI. Each task was classified as targeting one of the three spatial reasoning components: mental rotation (MR), spatial orientation (SO), or spatial visualisation (SV).

The assessment framework comprised four tests administered during the study, each consisting of tasks targeting one or more of these components through a range of geometric contexts, representational formats, and cognitive demands. The tasks were designed to reflect the content and reasoning processes encountered in university geometry education. An overview of the assessment framework is presented in Table 1.

The scoring procedure reflected the structure and response format of each task. Discrete response items or sub-items were scored dichotomously (0 = incorrect, 1 = correct), with scores summed when a task comprised multiple sub-items. For more complex constructed-response tasks, partial credit was awarded for predefined components or steps of a correct solution, allowing partially successful performance to be represented rather than reducing the entire response to a binary outcome (May et al, 2023). Consequently, the maximum score for each task reflected the number of scorable elements required for its complete solution.

3.5 Data Analysis

The collected data were processed using Microsoft Excel and analysed using PAST Statistics 5.0. Descriptive statistics were calculated for each assessment task and cohort. Prior to inferential analysis, data normality was assessed using the Shapiro–Wilk test. As the assumption of normality was not satisfied across the assessment tasks ($p < 0.05$), non-parametric statistical methods were applied.

Differences between the quasi-control and quasi-experimental cohorts were analysed separately for each assessment task using the Mann–Whitney U test, with the significance level set at $\alpha = 0.05$. The results are presented using descriptive statistics, boxplots, and statistical comparisons of individual assessment tasks.

4. Findings

This section presents the comparative results of the quasi-control and quasi-experimental cohorts across the four geometry-specific assessment tests. The analysis focuses on differences in performance at the task level in relation to the assessed components of spatial reasoning. Descriptive statistics are complemented by Mann–Whitney U tests to identify statistically significant differences between the cohorts. Boxplots are presented only for tasks in which statistically significant differences were identified, providing a visual representation of the score distributions.

4.1 Overall Statistical Comparison

The overall statistical results are presented in Tables 2 and 3.

Table 2: Results of the Normality and Group Comparison Tests

Test	Task ID	Normality Test Shapiro–Wilk (p)	Group Comparison *Mann–Whitney U (p)*
Test 1	T1.1	1.19E-03**	0.7319
	T1.2	8.35E-06***	0.7302
	T1.3	1.32E-03**	0.4879

Test	Task ID	Normality Test Shapiro–Wilk (p)	Group Comparison *Mann–Whitney U (p)*
Test 2	T2.1	1.85E-08***	1
	T2.2	3.20E-03**	0.0248*
	T2.3	6.54E-08***	0.0373*
	T2.4	5.61E-05***	0.0249*
Test 3	T3.1	3.54E-06***	0.2091
	T3.2	2.69E-09***	0.5369
	T3.3	1.86E-06***	Not calculated (zero variance)
	T3.4	2.89E-07***	0.0163*
Test 4	T4.1	3.48E-06***	0.1026
	T4.2a	1.13E-02*	0.2315
	T4.2b	4.99E-05***	0.9909

Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Median Scores for the Quasi-Control and Quasi-Experimental Cohorts

Test	Task ID	Quasi-Control Group	Quasi-Experimental Group
Test 1	T1.1	0.833	0.917
	T1.2	5.500	5.500
	T1.3	0.750	0.750
Test 2	T2.1	1.000	1.000
	T2.2	2.000	1.000
	T2.3	1.000	1.000
	T2.4	2.500	0.000
Test 3	T3.1	1.000	1.000
	T3.2	1.000	1.000
	T3.3	0.000	0.000
	T3.4	1.000	0.000
Test 4	T4.1	1.000	1.000
	T4.2a	2.000	3.000
	T4.2b	2.000	2.000

Higher or equal median values are shown in **bold**.

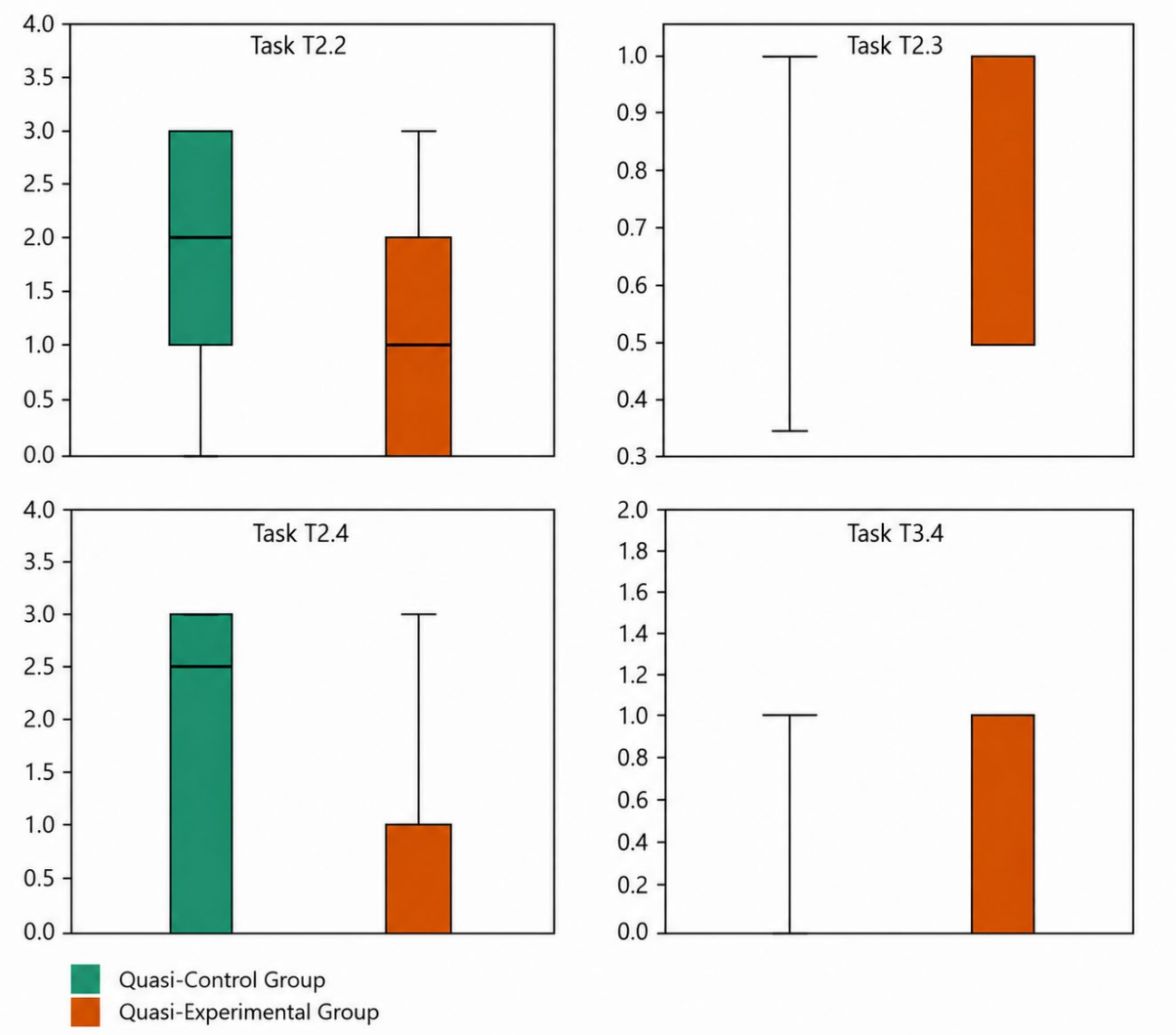


Figure 1: Boxplots of score distributions for Tasks T2.2, T2.3, T2.4, and T3.4

Table 2 summarises the results of the Mann–Whitney U tests for all assessment tasks, while Table 3 presents the corresponding median scores achieved by both cohorts. Statistically significant differences between the quasi-control and quasi-experimental cohorts were identified in four tasks: T2.2, T2.3, T2.4, and T3.4. The corresponding score distributions are presented in Figure 1**.**

4.2 Comparison by Assessment Test

To examine the results in more detail, the performance of the quasi-control and quasi-experimental cohorts was analysed separately for each of the four assessment tests. The following sections summarise the statistical comparisons and median scores for each assessment test, while Figure 1 presents the score distributions for the four tasks in which statistically significant differences were identified.

4.2.1 Test 1

No statistically significant differences were identified between the quasi-control and quasi-experimental cohorts in any of the three tasks included in Test 1 (Table 2). As shown in Table 3, both cohorts achieved identical median scores in Tasks T1.2 and T1.3, while only a small difference was observed in Task T1.1. Overall, the results indicate comparable performance in tasks assessing spatial visualisation and mental rotation.

4.2.2 Test 2

Among the four assessment tests, Test 2 yielded the highest number of statistically significant differences between the cohorts. No statistically significant difference was found in Task T2.1 (Table 2). In contrast,

significant differences were identified in Tasks T2.2, T2.3, and T2.4. Table 3 shows higher median scores for the quasi-control cohort in Tasks T2.2 and T2.4, whereas both cohorts achieved identical median scores in Task T2.3. The score distributions for the statistically significant tasks are presented in Figure 1.

4.2.3 Test 3

No statistically significant differences were observed in Tasks T3.1, T3.2, or T3.3 (Table 2). Table 3 shows identical median scores for both cohorts in these tasks. A statistically significant difference was identified only in Task T3.4, where the quasi-control cohort achieved a higher median score. The corresponding score distribution is presented in Figure 1.

4.2.4 Test 4

No statistically significant differences were identified between the cohorts in any of the tasks included in Test 4 (Table 2). Median scores were identical in Tasks T4.1 and T4.2b, whereas a higher median score was observed for the quasi-experimental cohort in Task T4.2a (Table 3). Despite this difference in median scores, no statistically significant difference was identified between the cohorts.

5. Discussion

The findings indicate that differences between the quasi-control and quasi-experimental cohorts were not observed across all components of spatial reasoning. Comparable performance in simpler tasks suggests that basic spatial abilities, particularly those involving mental rotation and elementary spatial orientation, may be retained over time. In contrast, statistically significant differences emerged in tasks requiring the reconstruction of three-dimensional objects from two-dimensional representations, interpretation of orthographic views, and more advanced spatial synthesis. These findings suggest that higher-order spatial reasoning abilities may be more sensitive to continuous and structured educational support than more fundamental spatial skills. This interpretation is consistent with research demonstrating that spatial abilities can be systematically developed through targeted training addressing specific spatial factors and solution strategies (Maresch, 2015).

The observed pattern of results suggests that not all components of spatial reasoning develop and are retained in the same way. The more demanding tasks required students to coordinate several cognitive processes simultaneously, including the interpretation of multiple representations, mental reconstruction of three-dimensional objects, and spatial synthesis, which is consistent with Duval's (2006) view that successful mathematical understanding depends on coordinating and converting between different semiotic representations. Such tasks place substantially higher demands on cognitive processing than isolated mental rotation or elementary orientation tasks, which may explain why differences between the cohorts became apparent only in these more complex activities.

Beyond the comparison of cohort performance, the findings also provide insights into the characteristics of geometry assessment tasks that are most effective in differentiating levels of spatial reasoning. The tasks that revealed significant differences shared common cognitive demands, including the coordination of multiple representations, interpretation of orthographic projections, and mental reconstruction of three-dimensional structures. By contrast, tasks targeting more isolated spatial processes, such as basic mental rotation or simple spatial orientation, showed limited discriminatory capacity. These findings suggest that geometry-specific assessment tasks requiring the integration of multiple spatial reasoning processes may provide a more sensitive assessment of advanced spatial competence in prospective mathematics teachers.

The findings also have implications for the design of technology-enhanced geometry learning environments in teacher education. Rather than supporting only the acquisition of basic spatial skills, such environments should provide sustained opportunities to engage with cognitively demanding tasks that require the coordination of multiple representations, dynamic visualisation, and three-dimensional reasoning. Recent research similarly indicates that technology-enhanced learning environments incorporating interactive three-dimensional visualisation tools can contribute to the development of students' spatial visualisation abilities (Medina Herrera, Juárez Ordóñez and Ruiz-Loza, 2024). The present findings suggest that technology-enhanced instruction may be particularly valuable for maintaining and strengthening these more advanced aspects of spatial reasoning. Consequently, the educational value of digital technologies depends not only on their interactive features but also on the extent to which they enable repeated engagement with cognitively demanding geometric tasks.

This study has several limitations that should be considered when interpreting the findings. The research involved students from a single teacher education programme, and the sample sizes were relatively small, which limits the generalisability of the results. The findings should therefore be interpreted as context-specific evidence rather than as representative of prospective mathematics teachers more broadly. In addition, the quasi-experimental design does not allow causal conclusions regarding the effectiveness of the instructional approach. Future research should include larger and more diverse cohorts, longitudinal designs, and further validation of geometry-specific assessment tasks to better understand the development and long-term sustainability of different components of spatial reasoning. It should also investigate whether similar patterns can be observed across different educational contexts and with alternative forms of technology-enhanced geometry instruction.

6. Conclusion

This study compared spatial reasoning outcomes between a current cohort of prospective mathematics teachers and a cohort assessed approximately one year after completing the same technology-enhanced geometry course. While both cohorts demonstrated comparable performance in simpler spatial reasoning tasks, significant differences emerged in more demanding tasks requiring the coordination of multiple representations, interpretation of orthographic views, and advanced spatial visualisation. These findings suggest that performance in higher-order spatial reasoning tasks may be particularly sensitive to the continuity of structured educational support.

Beyond the comparison of cohort performance, the study highlights the importance of carefully designed geometry assessment tasks for evaluating advanced spatial competence. Tasks requiring the integration of multiple spatial reasoning processes appear to provide more informative evidence of students' spatial thinking than tasks focusing on isolated spatial skills. The findings contribute to current research on technology-enhanced mathematics teacher education by informing the design of learning environments and assessment approaches that support the long-term development of spatial reasoning.

Ethics Declaration

Participation in the study was voluntary. All participants provided informed consent for the use of their anonymised data for research purposes, in accordance with the ethical guidelines and institutional requirements of the authors' universities for research involving human participants.

AI Declaration

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to improve the clarity, coherence, and linguistic quality of the English text. All AI-assisted output was critically reviewed, substantially revised, and verified by the authors, who assume full responsibility for the final content of the manuscript.

Acknowledgement

This paper is supported by grant KEGA 077UK-4/2025 “Edu-Steam laboratory in primary education teaching”

References

Duval, R. (2006) “A cognitive analysis of problems of comprehension in a learning of mathematics”, *Educational Studies in Mathematics*, Vol 61, No. 1–2, pp 103–131.

Hohenwarter, M. and Lavicza, Z. (2007) “Mathematics teacher development with ICT: towards an International GeoGebra Institute”, *Proceedings of the British Society for Research into Learning Mathematics*, Vol 27, No. 3, pp 49–54.

Lowrie, T., Resnick, I., Harris, D. and Logan, T. (2020) “In search of the mechanisms that enable transfer from spatial reasoning to mathematics understanding”, *Mathematics Education Research Journal*, Vol 32, No. 2, pp 175–188.

Maresch, G. (2015) “How to Develop Spatial Ability? Factors, Strategies, and Gender Specific Findings”, *Journal for Geometry and Graphics*, Vol 19, No. 1, pp 133–157.

May, T., Koskey, K., Bostic, J., Stone, G., Kruse, L. and Matney, G. (2023) “Examining how using dichotomous and partial credit scoring models influence sixth-grade mathematical problem-solving assessment outcomes”, *School Science and Mathematics*, Vol 123, No. 2, pp 54–67.

Medina Herrera, L.M., Juárez Ordóñez, S. and Ruiz-Loza, S. (2024) “Enhancing mathematical education with spatial visualization tools”, *Frontiers in Education*, Vol 9, Article 1229126.

Ramful, A., Lowrie, T. and Logan, T. (2017) **"**Measurement of Spatial Ability: Construction and Validation of the Spatial Reasoning Instrument for Middle School Students", *Journal of Psychoeducational Assessment*, Vol 35, No. 7, pp 709–727.

SERC (2022) **"**Spatial Reasoning Instrument (SRI)", University of Canberra, [online], available at: https://serc.edu.au/spatial-reasoning-instrument/

Schmid, A. and Korenova, L. (2024) **"**Enhancing Geometry Learning with GeoGebra: A Study", in *Proceedings of the 23rd European Conference on e-Learning*, pp 487–496, Academic Conferences International Limited.

Sinclair, N., Holt, D. and Heyd-Metzuyanim, E. (Eds.) (2016) *The Mathematics Teacher in the Digital Era: An International Perspective on Technology Focused Professional Development*, Springer, Dordrecht.

Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., Alden, A.R., Warren, C. and Newcombe, N.S. (2013) **"**The Malleability of Spatial Skills: A Meta-Analysis of Training Studies", *Psychological Bulletin*, Vol 139, No. 2, pp 352–402.

Weigand, H.-G., Hollebrands, K. and Maschietto, M. (2025) **"**Geometry education at secondary level – A systematic literature review", *ZDM – Mathematics Education*, Vol 57, No. 4, pp 829–843.

Integrating Generative AI for Assessment and Feedback in Higher Education in Albania

Xhensila Sinanaj, Ervis Sanço, Erion Selimi, Doris Madhi and Elvira Fetahu
Aleksandër Xhuvani University, Elbasan, Albania

xhensila.abazi@uniel.edu.al
ervis.sanco@uniel.edu.al
erion.selimi@uniel.edu.al
doris.madhi@uniel.edu.al
elvira.fetahu@uniel.edu.al

Abstract: The rapid development of Generative Artificial Intelligence (GenAI) is transforming assessment and feedback practices in higher education, creating both opportunities and challenges for teaching and learning. This empirical study, situated within Aleksandër Xhuvani University investigates how GenAI is integrated into assessment and feedback practices in higher education. The study investigates 57 educators' and 128 students' perceptions, experiences and expectations regarding the use and feedback with the aim of informing evidence-based pedagogical innovation. It seeks to answer five research questions. (1) What GenAI tools are integrated in assessment and feedback activities? (2) What do educators and students need for GenAI integration? (3) How to integrate GenAI tools in pedagogical practice? (4) What are the experienced and desired impacts of GenAI tools? and (5) What are the foreseeable pitfalls of using GenAI tools? Methodologically, the study applies a mixed-method, qualitative exploration through open-ended questions and a quantitative survey based research design administered to higher education educators and students. The findings indicate that currently the most commonly used GenAI tools are large language models and plagiarism detection tools by both groups of respondents. The main reason for using these tools appears to be for research purposes such as essays or scientific papers. The respondents provided almost identical answers regarding the perceived risks of using AI tools and gave valuable information about their needs and the integration of GenAI into higher education institutions. Concerns persist regarding plagiarism, over-dependence on AI-generated content and the absence of clear regulatory frameworks in Albanian universities. The results contribute to Albanian educational research by identifying priority areas for institutional policy development, capacity building, and the design of ethical and inclusive GenAI-supported assessment frameworks. They offer transferable insights for higher education institutions and policymakers seeking to integrate GenAI responsibly and effectively into assessment and feedback practices across Albanian context and beyond.

Keywords: Generative AI, Assessment, Feedback, Higher Education, Albania

1. Introduction

Assessment and feedback represent critical components of the learning process, as they help evaluate student learning outcomes while guiding students toward academic improvement. However, traditional assessment practices often face challenges such as large class sizes, limited time for detailed feedback, and increasing administrative workload for educators. In this context, GenAI presents a promising opportunity for both educators and students. GenAI tools offer opportunities for educators to enhance efficiency by assisting educators in grading, generating feedback, and designing assessments.

Despite growing global interest in artificial intelligence in education, empirical evidence regarding the adoption and perceptions of GenAI among academics remains limited, particularly in developing higher education contexts. Understanding how educators and students perceive and use these technologies is crucial for designing institutional policies and pedagogical strategies that support responsible AI integration.

This study therefore aims to explore the use of generative AI tools in assessment and feedback practices in higher education. The study is situated within Aleksandër Xhuvani University, a public university in Albania founded in 1991 which is considered one of the most prominent universities in the country for training teachers, specialists, and young researchers in various fields of study. Aleksandër Xhuvani University comprises over 6,480 students across 3 PhD, 24 MSc, 21 Professional Masters, 8 Professional Programs and 33 BSc programs, with 217 researchers and full time educators and 326 part-time educators. Specifically, the study investigates the level of familiarity with GenAI among educators and students, the frequency of its use in assessment-related tasks, and the perceived risks associated with its adoption. It seeks to answer five research questions. (1) What GenAI tools are integrated in assessment and feedback activities? (2) What do educators and students need for GenAI integration? (3) How to integrate GenAI tools in pedagogical practice? (4) What are the experienced and desired impacts of GenAI tools? and (5) What are the foreseeable pitfalls of using GenAI tools?

2. Literature Review

The integration of Artificial Intelligence (AI) into higher education has accelerated significantly in recent years, particularly following the emergence of Generative Artificial Intelligence (GenAI) tools such as ChatGPT, Claude, and Google Gemini. These technologies are increasingly influencing assessment and feedback practices by enabling automated content generation, personalized responses, and adaptive learning support. The growing accessibility of large language models (LLMs) has generated extensive scholarly debate regarding their pedagogical potential, ethical implications, and impact on academic integrity in higher education.

Early studies on AI in education focused primarily on intelligent tutoring systems and automated grading technologies (Luckin et al., 2016). However, the emergence of generative AI has shifted attention toward more sophisticated systems capable of producing human-like text, supporting feedback generation, and assisting both educators and students in academic tasks (Kasneci et al., 2023). According to Holmes et al. (2022), AI-supported education offers opportunities for improving efficiency, scalability, and personalization in teaching and assessment processes. These developments are particularly relevant in higher education institutions facing increasing workloads, larger student populations, and demands for faster and more individualized feedback.

One of the most widely discussed applications of GenAI in higher education concerns assessment and feedback practices. Feedback is recognized as a central component of effective learning because it helps students identify strengths, weaknesses, and opportunities for improvement (Hattie & Timperley, 2007). Nevertheless, traditional feedback processes are often constrained by limited educator time and institutional resources. Generative AI tools can support educators by automating aspects of grading, generating formative feedback, and assisting in rubric-based evaluations (Bearman et al., 2023). Studies suggest that AI-generated feedback may improve response time and consistency while reducing administrative burden for educators (Chan & Hu, 2023).

Research also indicates that students increasingly use GenAI tools for academic writing, brainstorming, summarization, coding, and problem solving. Cotton et al. (2023) argue that tools such as ChatGPT are transforming student learning practices by functioning as “academic assistants” capable of generating explanations, essays, and draft responses. Similarly, Dwivedi et al. (2023) emphasize that generative AI has the potential to fundamentally reshape educational practices by supporting creativity, collaboration, and personalized learning experiences. However, these opportunities are accompanied by concerns regarding the accuracy and reliability of AI-generated outputs.

A substantial body of literature highlights the ethical and pedagogical risks associated with GenAI adoption in higher education. One of the primary concerns relates to academic integrity and plagiarism. Several scholars argue that generative AI systems may facilitate unauthorized assistance during assessments and make it increasingly difficult to distinguish between original student work and AI-generated content (Tlili et al., 2023). Perkins (2023) notes that traditional plagiarism detection systems are often ineffective in identifying AI-generated text, creating new challenges for universities attempting to preserve academic honesty. These concerns have prompted calls for revised assessment methods emphasizing critical thinking, reflection, and authentic learning activities.

Another frequently discussed issue concerns overreliance on AI technologies. Excessive dependence on generative AI may reduce students’ independent problem-solving abilities and critical thinking skills (Baidoo-Anu & Ansah, 2023). Researchers warn that students may become passive consumers of AI-generated knowledge rather than active participants in the learning process. Furthermore, educators express concerns that increased automation could weaken educator–student interaction and reduce opportunities for meaningful academic engagement (Selwyn, 2019).

Bias, transparency, and data privacy also represent important themes in the literature. Generative AI systems are trained on large datasets that may contain cultural, linguistic, or social biases, potentially reproducing discriminatory patterns in educational contexts (Bender et al., 2021). Zawacki-Richter et al. (2019) argue that institutions must carefully evaluate the ethical implications of AI deployment, particularly regarding fairness and accountability in automated assessment systems. In addition, concerns regarding data protection and student privacy have become increasingly important, especially when external AI platforms process sensitive academic information.

Despite these challenges, numerous scholars argue that responsible integration of GenAI can enhance educational quality when accompanied by appropriate pedagogical strategies and institutional policies. UNESCO (2023) emphasizes the importance of developing ethical guidelines and AI literacy among educators and students to ensure responsible use of generative technologies in education. Similarly, the European Commission

has highlighted the importance of trustworthy AI frameworks that prioritize transparency, human oversight, and inclusion in educational applications.

Research also demonstrates that successful AI integration depends heavily on educator readiness and institutional support. According to Rudolph et al. (2023), many educators recognize the potential benefits of generative AI but lack adequate training and guidance regarding its pedagogical use. Faculty development programs, digital infrastructure, and interdisciplinary collaboration are therefore considered essential for sustainable implementation. These findings are particularly relevant in developing and transitional higher education systems where institutional policies regarding AI remain underdeveloped.

Within the Balkan and Albanian context, empirical studies examining GenAI integration remain limited. Existing research on digital transformation in Albanian higher education has primarily focused on e-learning adoption and digital competencies rather than generative AI applications specifically. Consequently, there is a significant research gap regarding how educators and students in Albanian universities perceive, adopt, and evaluate generative AI tools in assessment and feedback activities.

Overall, the literature suggests that generative AI technologies have the potential to transform assessment and feedback practices by improving efficiency, accessibility, and personalization. However, their implementation also raises important concerns related to academic integrity, ethical governance, transparency, and pedagogical quality. The balance between innovation and responsible use therefore remains a central issue for higher education institutions worldwide.

3. Methodology

In order to achieve the stated research objective, the authors applied a mixed-methods: a qualitative exploration through open-ended questions and a quantitative survey based research design administered to higher education educators and students. Data was collected using an online questionnaire administered through Google Forms for the period February-March 2026. The questionnaire aimed to capture the use and perceptions of generative AI tools in higher education assessment and feedback practices. Specifically, aimed to gain information on prior experience with generative AI tools, frequency of AI use in assessment and feedback, and perceptions of potential risks associated with AI integration. Participation in the survey was voluntary and anonymous, allowing respondents to provide honest and unbiased responses. The final dataset consisted of 57 valid responses from academic staff members and 128 valid responses from students representing different academic disciplines. Risk perceptions section were measured using a five-point Likert scale ranging from 1 (low risk) to 5 (high risk). The collected data were analyzed using descriptive statistical methods, including frequency distributions, percentages, and mean values for Likert-scale variables. These descriptive statistics were used to identify patterns in the adoption of generative AI tools and to examine educators' and students' perceptions of associated risks.

4. Results and discussions

The survey collected responses from 57 academic staff members and 128 students across Aleksandër Xhuvani University from different disciplines. The aim was to examine the use of Generative AI (GenAI) tools in assessment and feedback processes and to identify perceived risks and adoption patterns. Table 1 and Table 2 present some descriptive statistics in order to identify the familiarity with GenAI tools and frequency of use.

Table 1: Prior Experience with GenAI

Experience with GenAI	Percentage (Staff)	Percentage (Students)
Yes	68.4%	85%
No	31.6%	15%
Total	100%	100%

The results confirm that a majority of both respondents reported prior experience using GenAI tools such as ChatGPT or similar systems in teaching-related activities.

Table 2: Frequency of GenAI Use

Frequency of Use	Percentage (Staff)	Percentage (Students)
Occasionally (3–5 times/semester)	33.3%	25%
Rarely (1–2 times/semester)	24.6%	10%
Frequently (monthly)	17.5%	20%
Regularly (weekly)	12.3%	40%
Never	12.3%	5%

The results indicate that GenAI use is primarily occasional rather than routine for the academic staff but almost two-thirds of students use GenAI at least monthly, with a significant portion using it weekly. More than half of staff respondents reported using AI rarely or occasionally. This suggests GenAI is still in an exploratory stage in higher education practice for the academic staff and it is actually more explored by the students. Such frequent use suggests that these tools are and not merely supplementary but are increasingly perceived as essential for academic productivity.

4.1 Research Question 1: What GenAI Tools are Integrated in Assessment and Feedback Activities?

Table 3: Use of GenAI Tools

Genai Tools	Percentage (Staff)	Percentage (Students)
LLM (GPT based tools, Claude, Bard)	59.6%	85%
Plagiarism detection tools	33.3%	8%
Automated assessment and grading platforms	26.3%	6%
Feedback generation tools	14%	10%
Institution provided GenAI tool	8.8%	-
Other GenAI tools	15.8%	8%
None	14%	15%

Results show that LLMs dominate assessment and feedback activities, used by the majority of both groups of respondents. There is a noticeable increase in the use of plagiarism detection tools and automated grading platforms among academic staff, while feedback tools remain a secondary, yet significant, resource for students. The results on table 3 show some differences according the use of GenAI tools among academic staff and students with students across IT and Informatics-Economics programs being the most active users.

4.2 Research Question 2: What do Educators and Students Need for GenAI Integration?

Survey results indicate that educators identify several institutional and professional requirements for the successful integration of generative AI tools in higher education. The most frequently cited needs include research and technological infrastructure, staff qualification and training. This suggests that both technical capacity and educator readiness are perceived as equally critical for effective implementation.

Additionally, international cooperation and academic mobility opportunities were also highlighted as important elements supporting AI integration. These responses indicate that staff respondents view generative AI adoption not only as a technological issue but also as part of a broader academic ecosystem that involves collaboration, knowledge exchange, and exposure to global best practices. Overall, the findings suggest that educators perceive GenAI integration as requiring both institutional investment in infrastructure and systematic professional development for academic staff. On the other hand, the most common uses of GenAI tools among students are writing essays and research papers (70%), solving problems and performing calculations (60%), and preparing project reports (50%). Additionally, 30% of students use GenAI to generate written feedback. A smaller but still notable proportion (25%) use these tools for exam preparation and practicing multiple-choice questions.

4.3 Research Question 3: How to Integrate GenAI Tools in Pedagogical Practice

Respondents were also asked how generative AI tools should be integrated into pedagogical practice. The results indicate a strong preference for gradual and controlled implementation strategies.

The most frequently selected approach was pilot programs in specific departments or courses. This finding suggests that educators prefer testing AI tools in limited contexts before wider institutional adoption.

The second most common approach was a hybrid implementation model combining standardized institutional tools with department-specific applications. This reflects a desire to balance institutional oversight with disciplinary flexibility.

Other proposed approaches include faculty-led adoption with institutional oversight and full integration across programs using standardized tools. A smaller number of respondents supported centralized implementation managed by academic technology teams. These results suggest that educators prefer incremental integration strategies that allow experimentation and adaptation across disciplines rather than immediate large-scale institutional deployment. There is also strong support for integration by students point of view with almost 75% of them confirming it and 10% disagreeing. Integration is seen as most useful in theoretical subjects and project based learning.

4.4 Research Question 4: Experiences and Desired Impacts of GenAI Tools

Educators reported generally moderate to positive perceptions of the potential benefits of generative AI in assessment and feedback processes. Respondents indicated that GenAI tools may help reduce time spent on grading and feedback preparation, support alternative assessment approaches, and enable faster feedback cycles for students. Many educators also noted that AI tools could help reduce lecturer workload and support clear and understandable feedback for students. However, slightly lower levels of agreement were observed regarding the ability of GenAI tools to provide highly personalized feedback or fully consistent grading outcomes. Overall, the findings suggest that educators recognize the efficiency benefits of generative AI technologies, while still expressing moderate reservations about their pedagogical reliability and long-term impact on learning processes. It is also observed a positive impact (75%) of GenAI tools among students. Students report improved productivity, easier project preparation and support in learning practices.

4.5 Research Question 5: Foreseeable Pitfalls of Using GenAI Tools

Respondents also evaluated potential risks associated with the use of generative AI in assessment and feedback practices. The analysis reveals several important concerns among educators and students as described in table 4.

The most significant concern relates to the potential misuse of AI tools by students for academic dishonesty. Many educators expressed concerns that generative AI systems could produce essays, assignments, or exam responses that may bypass traditional plagiarism detection systems. Additional risks identified by educators include privacy and data protection concerns, lack of transparency in AI-generated feedback, possible bias in automated responses, and the risk that excessive reliance on AI tools could reduce meaningful interaction between educators and students. Plagiarism is the most frequently mentioned concern by students (40%), dependance on AI (25%) and accuracy of information (20%). About 15% of students reported no major concerns.

These findings highlight the importance of developing institutional policies, ethical guidelines, and pedagogical strategies that ensure responsible and transparent use of generative AI technologies in higher education.

Table 4: Educators Perceived Risks of GenAI in Assessment

Risk Factor	Mean Score (Staff)
Student misuse for academic dishonesty	4.07
Over-reliance reducing educator–student interaction	3.65
Privacy and data security concerns	3.58
Lack of transparency for students	3.47
Bias in AI-generated feedback	3.26
Technical reliability issues	3.12

From a pedagogical perspective, the results suggest that integration of GenAI into the curriculum is both necessary and inevitable. Educators and students are already using these tools for essays, problem solving, project writing and so on as stated earlier in the paper. Formalizing their use within teaching practices could help ensure that students develop critical skills in evaluating, contextualizing and ethically applying AI-generated content. Ignoring this trend risks widening the gap between institutional expectations and students' lived academic practices. The diversity of responses across faculties and study levels highlights the importance of tailored policies. Integration strategies should be discipline-specific, balancing opportunities for innovation with safeguards against misuse taking in consideration the differences in the adoption and attitudes among different programs.

Acknowledgement

This research has been co-funded by the European Union's *Erasmus+ Programme* under the Cooperation Partnerships in Higher Education (KA220-HED), **project code: 2025-1-DK01-KA220-HED-000364628**. The support of the European Commission does not constitute an endorsement of the contents, which reflect only the authors' views.

Ethics Declaration

This study was conducted in accordance with the ethical principles for research involving human participants and followed standards of voluntary participation, confidentially, anonymity, and informed consent. Participation in the study was entirely voluntary and all respondents were informed about the purpose of the research prior completing the questionnaire. The data were collected anonymously and treated confidentially as stated in the methodology section. The study did not involve any physical, psychological or social risks beyond those encountered in everyday academic and research activities. Furthermore, the authors declare that there are no conflicts of interest related to this study and that all findings are reported honestly and transparently.

AI Declaration

The authors acknowledge the use of Generative Artificial Intelligence (GenAI) tools, including ChatGPT, Dola and Copilot, during the preparation of this manuscript. These tools were used solely to support language refinement, grammar improvement, structuring of ideas, and drafting assistance during the writing process. All conceptual development, research design, data collection, data analysis, interpretation of findings, and final academic judgments were conducted exclusively by the authors. The authors critically reviewed, edited, and validated all AI-assisted content to ensure accuracy, originality, academic integrity, and compliance with ethical research standards. No AI tool was used to generate, manipulate, or fabricate research data, results, citations, or empirical findings. The authors take full responsibility for the content of this manuscript.

References

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence: Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52–62.

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in assessment and feedback. Assessment & Evaluation in Higher Education, 48(7), 1121–1135.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. Education and Information Technologies, 28(9), 1–18.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 60(6), 1–12.

Dwivedi, Y. K., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.

Gillani, N., Eynon, R., Chiappe, A., & Hjorth, I. (2023). Responsible AI in education: Ethical considerations and implications for policy. Learning, Media and Technology, 48(2), 1–15.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston: Center for Curriculum Redesign.

Kasneci, E., Sessler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. London: Pearson Education.

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era. Journal of University Teaching & Learning Practice, 20(2), 1–12.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342–363.

Selwyn, N. (2019). Should Robots Replace Teachers? AI and the Future of Education. Cambridge: Polity Press.

Tlili, A., Shehata, B., Adarkwah, M. A., et al. (2023). What if the devil is my guardian angel? ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10(15), 1–24.

UNESCO. (2023). Guidance for Generative AI in Education and Research. Paris: UNESCO.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(39), 1–27.

Zhai, X. (2022). ChatGPT user experience: Implications for education. Education and Information Technologies, 28(5), 1–18.

Zhou, L., Wang, Y., & Sun, J. (2023). Generative artificial intelligence in higher education: Opportunities and challenges for assessment and feedback. Computers and Education: Artificial Intelligence, 4, 100134.3

The Prisoner's Dilemma of AI: Cognitive Offloading Among Future Mathematics Teachers

Otto Suchanek, Tomas Javorcik, Tatiana Havlaskova and Magdalena Zavodna
Faculty of Education, University of Ostrava, Czech Republic

otto.suchanek@osu.cz
tomas.javorcik@osu.cz
tatiana.havlaskova@osu.cz
magdalena.zavodna@osu.cz

Abstract: The integration of Large Language Models (LLMs) into education has introduced substantial opportunities as well as challenges related to students' cognitive processes and professional development. For pre-service teachers, generative Artificial Intelligence (AI) creates a professional paradox. While AI systems can increase efficiency, support learning processes, and provide rapid access to information, excessive reliance on such systems may lead to cognitive offloading, automation bias, and gradual weakening of critical thinking and professional autonomy. Within educational contexts, this tension may be interpreted as a conflict between short-term individual gains and long-term consequences for human expertise and independent reasoning. This study compares pretest and posttest perceptions of generative AI among first-year pre-service mathematics teachers following a targeted educational intervention focused on principles of AI functioning and related cognitive risks. The Prisoner's Dilemma is used as a conceptual and pedagogical lens for framing the tension between the immediate benefits of AI use and its possible long-term educational consequences; it is not empirically operationalised or tested as a game-theoretic model. Within this framework, delegating cognitive activities to AI may represent a rational strategy at the individual level while simultaneously producing undesirable outcomes for broader educational development. A quasi-experimental pretest–posttest design without a control group was employed. Participants consisted of undergraduate students enrolled in a Mathematics Teacher Education programme at the University of Ostrava (pretest n = 18; posttest n = 14). Data collection was conducted using a semantic differential instrument consisting of sixteen bipolar dimensions measuring students' perceptions of AI characteristics. The intervention included instructional activities explaining principles of large language models, cognitive offloading, automation bias, and ethical implications of excessive dependence on AI systems. Data were analysed using descriptive statistics and non-parametric methods, specifically the Mann–Whitney U test. The results did not reveal statistically significant differences between pretest and posttest measurements across any of the analysed dimensions. Several mean scores differed descriptively between the measurements, particularly for transparency, predictability, and the role of AI in deeper learning, but these differences were not statistically significant. Given the small, unmatched samples and the absence of a control group, they cannot be interpreted as evidence of an intervention effect. These results highlight the importance of integrating critical AI literacy into teacher education programmes.

Keywords: Artificial Intelligence, Cognitive Offloading, Game Theory, Mathematics Education, Pre-service Teachers

1. Introduction

The integration of large language models (LLMs) into education represents one of the most significant transformations in the contemporary educational environment. Unlike previous technological innovations, tools based on generative artificial intelligence, such as ChatGPT, Gemini, and Claude, have spread among students and teachers more rapidly than stable institutional, methodological, and ethical frameworks for their use have been established. As a result, students and educators now commonly work with systems capable of generating texts, solving tasks, and creating educational materials without necessarily possessing sufficient understanding of the underlying principles and limitations of these models (Kasneci et al, 2023).

From a pedagogical perspective, LLMs do not represent merely another digital tool but rather a substantial transformation of the structure of the learning process itself. Generative models alter the ways in which information is searched, knowledge is organised, and problems are solved. At the same time, part of the cognitive work is transferred from humans to algorithms, raising the issue of *cognitive offloading*, defined as the delegation of mental operations to external systems. Current research suggests that this process may produce a dual effect. On one hand, AI may reduce cognitive load and allow students to focus on higher-order thinking processes; on the other hand, it may weaken deeper understanding, metacognitive control, and long-term intellectual autonomy (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024).

The issue of metacognitive calibration represents a particularly important challenge in the context of generative AI. Metacognitive calibration refers to an individual's ability to realistically evaluate their own knowledge and competencies. Studies indicate that AI-supported environments may create an illusion of understanding, in which students overestimate their actual abilities because they can rapidly generate plausible-looking answers

without deeper comprehension of the problem. If students repeatedly accept AI outputs without critical verification, their ability to independently monitor their own thinking may gradually decline, increasing reliance on algorithmic support systems (Lee et al, 2025; Lin and Sohail, 2026).

This issue is particularly significant in mathematics education and other exact disciplines. Mathematical reasoning relies on logical precision, structured argumentation, and the ability to identify even minor errors in reasoning processes. Although LLMs can effectively explain procedures or generate examples, and some studies suggest a positive influence on students' conceptual understanding and mathematical confidence (Canonigo, 2024), their operation is fundamentally based on probabilistic prediction of subsequent tokens rather than genuine logical reasoning (Soygazi and Oğuz, 2023). Generative models therefore do not "think" in the human sense but instead produce outputs based on statistical patterns in data. The authoritative nature of such responses may consequently lead students toward uncritical acceptance of incorrect mathematical procedures or arguments.

Closely related to this phenomenon are the concepts of *automation bias* and *automation complacency*. Automation bias refers to the tendency to trust automated systems even in situations where their outputs contradict human judgement or contain errors. Repeated experiences with apparently successful AI systems may reduce the level of critical monitoring and support the development of over-reliance on algorithmic systems. Research also suggests that frequent delegation of analytical tasks to AI may be associated with weakening of certain cognitive processes, particularly in areas requiring deeper logical analysis and independent argumentation (Parasuraman and Manzey, 2010; Nguyen, 2024; Zhai, Wibowo and Li, 2024).

Within the context of preparing future mathematics teachers, an important professional paradox emerges. Generative AI offers substantial improvements in efficiency, speed, and accessibility of supportive materials; however, it may simultaneously weaken the development of competencies that constitute the foundation of teaching professionalism. Current AI literacy frameworks therefore emphasise that teacher education should focus not only on technical mastery of tools but also on the development of critical thinking, epistemological vigilance, and the ability to reflect on the limitations of algorithmic systems (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026).

One possible conceptual lens for interpreting this paradox is the Prisoner's Dilemma, used here as a pedagogical analogy rather than as an empirically tested game-theoretic model. From a short-term perspective, delegating part of cognitive work to AI may appear rational because it provides immediate advantages in efficiency and performance. However, if such behaviour becomes the dominant strategy within a group, it may result in a collective loss of deeper understanding, professional autonomy, and critical thinking abilities. Evolutionary models based on the Prisoner's Dilemma further suggest that changes in the "payoff matrix" may substantially influence strategic behaviour and support the emergence of cooperative actions (Perc et al, 2024).

Although current literature increasingly discusses AI literacy among future teachers (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026), automation bias and over-reliance on AI systems (Parasuraman and Manzey, 2010; Nguyen, 2024), as well as cognitive offloading in generative AI environments (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024), relatively limited attention has been devoted to a game-theoretical interpretation of these processes in the context of mathematics teacher education. Existing studies primarily focus on technical AI integration, pedagogical effectiveness, or AI competency development, whereas the strategic conflict between short-term efficiency and long-term professional autonomy remains insufficiently explored (Perc et al, 2024).

Therefore, this study compares pretest and posttest perceptions of generative AI among pre-service mathematics teachers in the context of an educational intervention focused on the principles of LLM functioning, automation bias, cognitive offloading, and game theory. The research employs a pretest–posttest design based on semantic differential scales to analyse differences between the two measurements.

2. Methodology

2.1 Participants and Research Design

The study employed a quasi-experimental pretest–posttest design without a control group. The aim of the research was to compare pretest and posttest perceptions of generative artificial intelligence in the context of a targeted educational intervention focused on the principles of AI functioning, cognitive offloading, automation bias, and the application of game theory within an educational context.

The study used a convenience sample of first-year undergraduate students enrolled in a Mathematics Teacher Education programme at the Faculty of Education, University of Ostrava. A total of 18 students participated in the pretest, while 14 students completed the posttest after the intervention. The difference in the number of respondents was caused by regular absenteeism during face-to-face classes. Participation in the study was voluntary, and all participants were informed about the purpose of the research.

The study was conducted in three consecutive phases:

1. Pretest
2. Educational intervention
3. Posttest

The following research questions were formulated:

RQ1: What differences in perceptions of generative AI are observed between the pretest and posttest measurements among pre-service mathematics teachers?

RQ2: Do the pretest and posttest measurements differ in dimensions related to a critical and realistic perception of generative AI?

The study followed institutional ethical guidelines, and all participants provided informed consent before participation.

2.2 Research Procedure

The educational intervention was implemented between the pretest and posttest and consisted of two interconnected instructional blocks focusing on the principles of generative AI, cognitive offloading, automation bias, and the ethical implications of over-reliance on AI systems.

2.2.1 Block 1: How Large Language Models Work

The first part of the intervention focused on explaining the principles underlying neural networks and large language models. Students were introduced to the idea that neural networks do not represent a “digital mind” but rather mathematical models based on recognising statistical patterns in data.

The instructional content included explanations of:

- next-token prediction,
- attention mechanisms,
- embeddings,
- vector representations of words,
- high-dimensional vector spaces.

To explain vectorisation principles, a food-taste analogy was used. Students were shown that each food item can be described using several characteristics (e.g., sweetness, saltiness, sourness, and bitterness), creating a vector representation.

The analogy was subsequently transferred to language models, where words are represented using hundreds or thousands of dimensions describing semantic relationships between concepts. Students were also introduced to a simplified example:

$$King - Man + Woman = Queen$$

as an illustration of mathematical operations performed on vector representations of words.

The purpose of this block was to reduce anthropomorphic perceptions of AI and strengthen understanding of the statistical nature of language models.

2.2.2 Block 2: AI as a Tool and a Trap

The second part focused on ethical and pedagogical implications of AI use. Students initially completed a SWOT analysis focused on the use of AI in educational contexts in order to identify potential benefits and risks associated with AI systems.

Algorithmic atrophy was subsequently discussed through the following question:

> *"Would you be able to identify an error in an AI-generated mathematical proof if you could not construct the proof yourself?"*

A key part of the intervention involved applying the Prisoner's Dilemma framework to AI use in educational settings. The Prisoner's Dilemma describes situations in which individually rational decisions may lead to collectively suboptimal outcomes. In this context, excessive reliance on AI was presented as a potentially beneficial short-term strategy that may produce undesirable long-term consequences for learning and professional development.

Students discussed a simulated scenario involving an algebra test with two possible strategies:

- Option A – study independently
- Option B – use hidden AI support

Using concepts of Nash equilibrium and Pareto optimality, students discussed the conflict between short-term individual benefits and long-term consequences for professional development. The payoff matrix represented possible outcomes associated with different strategies, such as independent learning versus excessive use of AI. Students discussed how changes in educational conditions may alter incentives and shift behaviour towards more desirable learning outcomes. In this study, however, the Prisoner's Dilemma served solely as a conceptual and pedagogical framework within the intervention; no payoff parameters, strategy choices, equilibria, or other game-theoretic variables were operationalised or tested in the empirical analysis.

The intervention also included activities designed to modify these incentives. For example:

> *"AI generated the following mathematical procedure. Identify three logical errors and explain why they occurred."*

2.3 Research Instrument and Statistical Analysis

Data collection was conducted using a questionnaire based on a semantic differential approach. The instrument consisted of sixteen bipolar scales measured using a seven-point semantic differential. Higher values represented more positive evaluations of the given characteristic.

Students evaluated AI using dimensions such as:

- Non-transparent – Transparent
- Biased – Impartial
- Unpredictable – Predictable
- Competitive – Cooperative
- Manipulative – Honest
- Reducing understanding – Deepening understanding
- Leading to surface learning – Leading to deep learning
- Passive – Activating
- Dangerous – Safe
- Unethical – Moral

among others.

Due to the ordinal nature of data obtained through semantic differential scales, non-parametric statistical procedures were employed. Data analysis included descriptive statistics (mean, median, and standard deviation) followed by comparisons between pretest and posttest results.

Because pretest and posttest data were not paired at the individual respondent level, paired statistical procedures could not be applied. Therefore, pretest (n = 18) and posttest (n = 14) responses were analysed as two independent samples using the Mann–Whitney U test.

Arithmetic means were used only for descriptive and graphical purposes. Effect sizes were additionally examined to support interpretation of practical significance. Given the small group sizes, statistical power was limited, particularly for detecting small effects. Accordingly, non-significant results were not treated as evidence of equivalence or of the absence of any difference. All statistical analyses were performed using NCSS software.

3. Results

3.1 Descriptive Statistics

Descriptive statistics including mean (M), standard deviation (SD), median (Md), and interquartile range (IQR) were calculated for both pretest and posttest measurements. Given the ordinal nature of the seven-point scale, medians and interquartile ranges were additionally considered because they provide a more robust description of central tendency and variability.

In the pretest measurement, the highest mean scores were observed for the dimensions *Confrontational–Supportive* (M = 5.17), *Self-oriented–Prosocial* (M = 5.11), and *Reducing Understanding–Deepening Understanding* (M = 4.83), suggesting a relatively positive perception of generative AI prior to the intervention. The lowest scores were found for *Passive–Activating* (M = 3.61), *Unpredictable–Predictable* (M = 3.72), and *Leading to Surface Learning–Leading to Deep Learning* (M = 3.72). Following the intervention, *Confrontational–Supportive* remained the highest-rated dimension (M = 5.43), followed by *Biased–Impartial*, *Competitive–Cooperative*, and *Self-oriented–Prosocial* (all M = 4.71). In contrast, *Leading to Surface Learning–Leading to Deep Learning* (M = 3.21) and *Passive–Activating* (M = 3.79) remained among the lowest-rated dimensions.

Overall, the descriptive results suggest relatively stable perceptions of AI across most dimensions, while several dimensions associated with deeper learning and cognitive engagement showed less favourable evaluations following the educational intervention.

Table 1: Descriptive Statistics for Pretest and Posttest Measurements

Item	Pretest M	Pretest SD	Pretest Md	Pretest IQR	Posttest M	Posttest SD	Posttest Md	Posttest IQR
Non-transparent - Transparent	3.83	1.27	4	1.25	4.21	1.14	4	2
Biased - Impartial	4.61	1.26	4.5	2.25	4.71	1.35	5	2
Unpredictable - Predictable	3.72	1.14	4	1.25	4.21	1.4	4	2.25
Difficult - Easy to Use	4.78	1.02	4.5	2	4.57	1.48	5	3
Competitive - Cooperative	4.44	1.53	4	3	4.71	1.47	5	3
Self-oriented - Prosocial	5.11	1.34	5	2.25	4.71	1.16	5	1.25
Manipulative - Honest	4.5	1.27	5	1.25	4.43	1.31	5	2.25
Confrontational - Supportive	5.17	1.31	5	2	5.43	0.87	5.5	1
Reduces Understanding - Deepens Understanding	4.83	1	5	0.5	4.64	1.42	5	2.25
Surface Learning - Deep Learning	3.72	1.14	3	1.25	3.21	1.28	3	1.25
Discouraging Thinking - Supporting Thinking	3.83	1.45	4	2.25	4.07	1.52	4	2.25
Passive - Activating	3.61	1.26	3	1.25	3.79	1.34	4	2
Unregulated - Regulated	4.11	1.39	4	2	4	1.39	4	2.25
Dangerous - Safe	4.17	1.42	4	2	4.5	0.85	4	2
Harmful - Helping Humans	5.11	1.08	5	2	4.86	1.06	5	2

Item	Pretest M	Pretest SD	Pretest Md	Pretest IQR	Posttest M	Posttest SD	Posttest Md	Posttest IQR
Unethical-Moral	4.56	1.15	5	1	5	1.12	5	2

Note: M = Mean; SD = Standard Deviation; Md = Median; IQR = Interquartile Range.

3.2 Semantic Differential Profile

To visualise changes in the perception of generative artificial intelligence between pretest and posttest measurements, a semantic differential profile was created. The figure presents mean scores for individual dimensions measured on a seven-point scale, where higher values indicate more positive evaluations of a given characteristic.

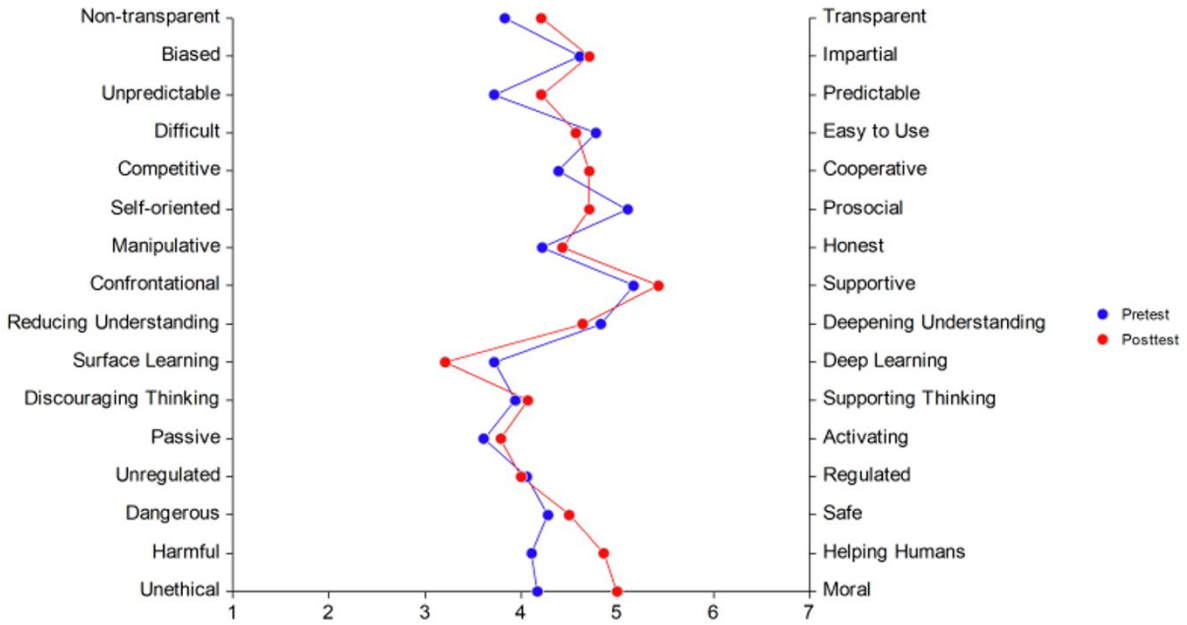


Figure 1: Semantic differential profile – comparison of pretest and posttest measurements

Visual comparison of the profiles suggests that the overall shape of the curves remained relatively similar across both measurements, indicating a certain degree of stability in students' perceptions of generative AI. Nevertheless, several partial changes can be observed.

Slight increases in mean values were identified for dimensions such as:

- Non-transparent – Transparent (3.83 → 4.21)
- Unpredictable – Predictable (3.72 → 4.21)
- Confrontational – Supportive (5.17 → 5.43)
- Dangerous – Safe (4.17 → 4.50)

In contrast, decreases were observed for dimensions such as:

- Self-oriented – Prosocial (5.11 → 4.71)
- Surface Learning – Deep Learning (3.72 → 3.21)

These differences represent descriptive changes in mean values only and do not indicate statistical significance by themselves. Therefore, statistical significance was subsequently examined using non-parametric methods.

3.3 Comparison of Pretest and Posttest Results Using the Mann–Whitney U Test

To examine differences between pretest and posttest measurements, the Mann–Whitney U test was employed. This procedure was selected because pretest and posttest responses could not be reliably matched at the individual participant level. Consequently, both measurements were analysed as two independent samples.

Furthermore, the data consisted of ordinal responses obtained through seven-point semantic differential scales, supporting the use of non-parametric statistical procedures.

The results of the Mann–Whitney U test did not reveal statistically significant differences between pretest and posttest measurements in any of the observed dimensions. In all cases, p-values exceeded the significance threshold of 0.05. The lowest p-value was observed for the *Unpredictable–Predictable* dimension (p = 0.273); however, even this result did not reach statistical significance. Effect sizes ranged from r = 0.010 to r = 0.194, corresponding to negligible to small effects.

Table 2: Mann–Whitney U test comparing Pretest and Posttest Measurements

Dimension	U	Z	p	r	Effect Interpretation
Non-transparent - Transparent	144	0.709	0.478	0.125	Small
Biased - Impartial	134.5	0.332	0.74	0.059	Negligible
Unpredictable - Predictable	154	1.096	0.273	0.194	Small
Difficult - Easy to Use	116.5	-0.373	0.709	0.066	Negligible
Competitive - Cooperative	139	0.509	0.611	0.09	Negligible
Self-oriented - Prosocial	105	-0.824	0.41	0.146	Small
Manipulative - Honest	122.5	-0.137	0.891	0.024	Negligible
Confrontational - Supportive	139.5	0.532	0.595	0.094	Negligible
Reducing Understanding - Deepening Understanding	115.5	-0.42	0.674	0.074	Negligible
Surface Learning - Deep Learning	106	-0.801	0.423	0.142	Small
Discouraging Thinking - Supporting Thinking	137	0.43	0.667	0.076	Negligible
Passive - Activating	143	0.666	0.506	0.118	Small
Unregulated - Regulated	124.5	-0.059	0.953	0.01	Negligible
Dangerous - Safe	146	0.801	0.423	0.142	Small
Harmful - Helping Humans	111.5	-0.564	0.573	0.1	Small
Unethical - Moral	143	0.666	0.505	0.118	Small

Note: $r = |Z|/\sqrt{N}$, where N = 32. Effect sizes were interpreted as negligible for $r < 0.10$, small for approximately $r \geq 0.10$, medium for $r \geq 0.30$, and large for $r \geq 0.50$.

No statistically significant differences were detected between the pretest and posttest measurements in any semantic differential dimension. Because the samples were small and unmatched, the minor differences in mean scores are exploratory descriptive patterns only and should not be interpreted as evidence of individual attitude change or an intervention effect. The results therefore provide no statistical basis for concluding that the intervention altered students' general attitudes towards generative AI.

4. Discussion

The aim of this study was to compare pretest and posttest perceptions of generative AI among pre-service mathematics teachers in the context of an educational intervention focused on AI principles, cognitive offloading, automation bias, and game theory. No statistically significant differences were detected in any semantic differential dimension. The observed mean differences were descriptive only and cannot be attributed to the intervention. Accordingly, this study provides no evidence that the short intervention altered students' general attitudes towards AI; at most, the descriptive patterns identify dimensions for further investigation. This interpretation corresponds with previous studies on AI literacy among future teachers, which suggest that educational interventions often influence awareness, critical reflection, and understanding of AI principles rather than immediately changing overall attitudes or trust in technology (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026).

One exploratory descriptive pattern was the lower posttest mean for the Surface Learning–Deep Learning dimension; this difference was not statistically significant. Although this pattern is compatible with greater awareness of risks related to cognitive offloading and excessive reliance on AI systems, the present design cannot determine its cause. The intervention also included explanations of the probabilistic nature of large language models and discussions of automation bias, but the data do not establish that these components produced more critical reflection. Related concerns have been reported in research on cognitive offloading, metacognitive calibration, and automation bias in AI-supported environments (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024; Lee et al, 2025; Nguyen, 2024; Parasuraman and Manzey, 2010).

From the perspective of mathematics teacher education, the findings suggest that AI literacy should not focus solely on technical use of tools but should also support the development of critical thinking and understanding of the limitations of generative systems. This is particularly important in mathematics and other exact disciplines, where correctness of results and logical consistency represent fundamental aspects of the educational process (Canonigo, 2024; Soygazi and Oğuz, 2023).

The findings should be interpreted in light of important design limitations. The study used a small convenience sample from a single programme, included no control group, and could not match individual pretest and posttest responses. The resulting independent-sample analysis had limited statistical power, particularly for detecting small effects, and the design does not support causal attribution. The small, non-random sample also limits generalisability beyond this cohort. The relatively short intervention may have further reduced the possibility of observing differences. In addition, the Prisoner's Dilemma functioned as a conceptual and pedagogical lens rather than an empirically operationalised model; the study therefore does not test game-theoretic predictions.

5. Conclusion

This study compared pretest and posttest perceptions of generative artificial intelligence among pre-service mathematics teachers in the context of an educational intervention focused on AI functioning and related cognitive risks.

No statistically significant differences were identified between the two measurements. The small differences in mean scores were descriptive only and do not demonstrate individual change or an effect of the intervention. Accordingly, the study cannot determine whether the intervention altered students' attitudes or understanding.

The study nevertheless illustrates how the Prisoner's Dilemma can be used as a conceptual pedagogical lens for discussing cognitive offloading and excessive reliance on AI in teacher education. It also highlights the relevance of connecting technical AI literacy with critical reflection, without claiming that the present intervention produced this outcome.

Future research should examine these findings using larger samples and longitudinal research designs.

Ethics Declaration

Participation in the study was voluntary and all participants were informed about the purpose of the study before providing informed consent.

AI Declaration

Generative AI tools (ChatGPT, OpenAI) were used for language refinement, text structuring, and stylistic editing. All scientific interpretation, analysis, and final decisions remained the responsibility of the authors.

References

Bilbao Eraña, A. and Arroyo-Sagasta, A. (2025) “Fostering AI literacy in pre-service teachers: Impact of a training intervention on awareness, attitude and trust in AI”, Frontiers in Education, Vol. 10, Article 1668078. DOI: 10.3389/feduc.2025.1668078.

Canonigo, A.M. (2024) “Leveraging AI to Enhance Students' Conceptual Understanding and Confidence in Mathematics”, Journal of Computer Assisted Learning, Vol. 40, No. 6, pp. 3215–3229. DOI: 10.1111/jcal.13065.

Grinschgl, S. et al. (2025) “Cognitive offloading or cognitive overload? How AI alters the mental ecology of learning”, Frontiers in Psychology, Vol. 13, Article 1699320. DOI: 10.3389/fpsyg.2025.1699320.

Hsu, H.P. (2026) “Artificial Intelligence Literacy and Competency in Pre-Service Teacher Education”, Encyclopedia, Vol. 6, No. 4, Article 76. DOI: 10.3390/encyclopedia6040076.

Kasneci, E. et al. (2023) “ChatGPT for good? On opportunities and challenges of large language models for education”, Learning and Individual Differences, Vol. 103, Article 102274. DOI: 10.1016/j.lindif.2023.102274.

Lee, H. et al. (2025) “Learning Behaviors Mediate the Effect of AI-powered Support for Metacognitive Calibration on Learning Outcomes”, in Proceedings of the ACM Learning at Scale Conference. DOI: 10.1145/3706598.3713960.

Lin, Z. and Sohail, A. (2026) “Recalibrating academic expertise in the age of generative AI”, Patterns, Vol. 7, No. 1, Article 101473. DOI: 10.1016/j.patter.2025.101473.

Nguyen, T. (2024) “ChatGPT in Medical Education: A Precursor for Automation Bias?”, JMIR Medical Education, Vol. 10, e50174. DOI: 10.2196/50174.

Parasuraman, R. and Manzey, D.H. (2010) “Complacency and Bias in Human Use of Automation: An Attentional Integration”, Human Factors, Vol. 52, No. 3, pp. 381–410. DOI: 10.1177/0018720810376055.

Perc, M. et al. (2024) “Emergence of cooperation in the one-shot Prisoner’s Dilemma when AI mediates interactions”, Journal of the Royal Society Interface, Vol. 21, No. 218, Article 20240212. DOI: 10.1098/rsif.2024.0212.

Soygazi, H. and Oğuz, A. (2023) “An Analysis of Large Language Models and LangChain in Mathematics Education”, in Proceedings of the 2023 International Conference on Educational Technology. DOI: 10.1145/3633598.3633614.

Zhai, C., Wibowo, S. and Li, L.D. (2024) “The effects of over-reliance on AI dialogue systems on students’ cognitive abilities: a systematic review”, Smart Learning Environments, Vol. 11, Article 28. DOI: 10.1186/s40561-024-00316-7.

The Use of Generative AI in Higher Education Student Assessments: A Synthesis of Empirical Studies

Chun Meng Tang
James Cook University Singapore, Singapore

chunmeng.tang@jcu.edu.au

Abstract: Generative artificial intelligence (GenAI) has become increasingly common among higher education students as an aid for completing assessments. Although previous studies have examined the implications of GenAI use in assessments, there is a need to synthesise existing findings to better understand how students use GenAI and how it affects their experiences. This study aimed to synthesise empirical research on higher education students' use of GenAI in assessments. Guided by the Population, Exposure, and Outcome (PEO) framework, the study adopted a hybrid deductive and inductive codebook thematic analysis of 14 empirical studies identified through a search of the Education Resources Information Center (ERIC) database. The analysis identified four main themes: uses of GenAI in assessments, perceived benefits of GenAI, challenges and limitations of GenAI, and academic integrity and ethical implications of GenAI. The findings suggest that students use GenAI for writing and content development, research and information support, idea generation and planning, and feedback and assessment support. Participants also perceived several benefits associated with GenAI use, including improved assessment performance, increased efficiency, and enhanced learning support. However, the use of GenAI also presents challenges and limitations, including ineffective use, concerns about accuracy and reliability, and effects of over-reliance on learning skills. In addition, participants expressed concerns about plagiarism and academic misconduct, fairness and academic equity, and the need for transparency and ethical GenAI use. The findings suggest that effective and responsible GenAI use requires institutional guidelines and adequate proficiency in GenAI competencies. This study contributes to the growing literature on GenAI in higher education assessments and provides insights for students, educators, and policymakers on the integration of GenAI into assessment designs.

Keywords: Assessment, Generative AI, Higher Education, PEO Framework, Thematic Analysis

1. Introduction

The launch of ChatGPT by OpenAI in late 2022 provided many students with their first exposure to a generative artificial intelligence (GenAI) tool (OpenAI, 2022). Since then, students have gained access to many other GenAI tools, such as Google Gemini, Anthropic Claude, and Microsoft Copilot, and are increasingly using these tools as an aid to assist with assessment tasks (Kirsanov et al., 2026; Oc et al., 2025), leveraging on their ease of use and ability to quickly generate text, images, videos, and other content (Comscore, 2026; Statista, 2026a). However, the use of GenAI in student assessments, if misused, raises concerns about academic integrity (Akiba and Garte, 2024; Cong-Lem et al., 2024) and over-reliance (Aydeniz, 2025; Chkarka and Fatmi, 2026), which may hinder the development of the knowledge, skills, and attitudes necessary for learning (Xia et al., 2024; Zirar, 2023).

Although some previous studies have examined students' use of GenAI in assessments, they have generally focused on the broader use of GenAI rather than on specific GenAI interventions in specific assessment tasks. Examples of such studies include Biju et al.'s (2024) study on how AI-assisted language assessment influenced second language learners' anxiety, motivation, attitudes, and writing skills; Nikolic et al.'s (2023) study on how ChatGPT affected assessment tasks in engineering education; and Oc et al.'s (2025) study on higher education students' adoption of GenAI tools in coursework. However, examining the general use of GenAI provides limited insight into how students use it to complete specific assessment tasks. Furthermore, findings from previous studies that examined specific applications of GenAI in assessments remain fragmented.

Some previous studies have conducted qualitative reviews and analyses of GenAI use in assessments. Examples include Wangdi et al.'s (2025) qualitative content analysis of 17 articles, which examined the practices and limitations of using ChatGPT for student assessment; Xia et al.'s (2024) scoping review of 32 articles, which examined how GenAI affected higher education assessment at the levels of students, teachers, and institutions; and Zirar's (2023) thematic analysis of 25 articles, which examined how large language models influenced learning and assessment in higher education. However, the existing literature synthesised in these studies primarily focuses on the general use of GenAI and provides limited insight into students' actual use of GenAI in assessments.

This study, therefore, aimed to synthesise findings on how GenAI has affected the ways in which higher education students complete assessments. To achieve this aim, the study searched the digital library of the Education Resources Information Center (ERIC) for empirical studies conducted after the launch of ChatGPT and synthesised their findings using the Population, Exposure, and Outcome (PEO) framework. The synthesis

attempted to identify patterns in aspects such as the types of assessments completed using GenAI tools, how these tools were used, and how their use influenced students' experiences.

This study contributes to the existing body of knowledge by synthesising findings from previous studies to inform students', educators', and policymakers' decisions regarding the use of GenAI in student assessments and the design of student assessments for e-learning environments.

The remaining sections provide an overview of GenAI, its use in education and student assessments, and the PEO framework; describe the research design and method; present the study findings; and conclude with a discussion of the findings.

2. Literature Review

2.1 Generative Artificial Intelligence (GenAI)

Built on models trained on large datasets, GenAI can generate content, including text, images, and videos, based on user prompts and is expected to disrupt a range of industries, from technology to banking (Statista, 2026a; Statista, 2026b). Since the launch of OpenAI's ChatGPT in late 2022, many other GenAI tools have become widely accessible, including Google Gemini, Anthropic Claude, and Microsoft Copilot. The widespread availability of these tools has further accelerated the adoption of GenAI across industries. According to Microsoft's AI diffusion report, GenAI adoption reached 16.3% of the world's population in the second half of 2025 (Microsoft, 2026). The global GenAI market is projected to reach US$86.70 billion in 2026, with an annual growth rate of 24.83% from 2026 to 2032 (Statista, 2026a).

2.2 GenAI in Education and Student Assessments

The launch of OpenAI's ChatGPT attracted the attention not only of students but also of educators and educational institutions because of its potential implications for teaching and learning. Singh and Ngai (2024) analysed the initial responses of leading US and UK universities to ChatGPT and found that its uses go beyond content generation to include supporting student learning, assisting educators in teaching and assessment design, and fostering skill development such as critical thinking and research. They also highlighted that ChatGPT has the potential to transform or disrupt work practices, support the development of educational materials and assessments, and threaten academic integrity if misused.

Concerning the educational impact of GenAI, Chiu (2024) examined the perspectives of primary and secondary school teachers and leaders. The study identified four main themes: (1) student learning, including AI education, critical reasoning and thinking, digital, media, and information literacies, and generic skills such as problem-solving, communication, and self-directed learning; (2) teacher learning, including AI-integrated curriculum development, teacher AI literacy, new facilitation skills, and interdisciplinary teaching; (3) assessment, including assessment approaches and dimensions of assessment; and (4) administration, including administrative effectiveness and attitudes.

Regarding the use of GenAI in assessment, Xia et al.'s (2024) scoping review summarised that GenAI can affect higher education assessment at the levels of students, teachers, and institutions. At the student level, the study identified three opportunities: perceived unbiased feedback, immediate and diverse feedback, and self-assessment, as well as one key challenge related to academic integrity. At the teacher level, the study identified six impacts: teacher assessment literacy, more diverse and innovative assessment practices, assessment-driven teaching, poor student skill development, teacher beliefs about assessment, and balancing GenAI and human assessment. At the institutional level, the study identified five implications: assessment policies, new literacies and professional development for teachers, shifting educational focus and learning objectives, interdisciplinary assessment, and the disruption of traditional assessment methods.

2.3 The Population, Exposure, and Outcome (PEO) Framework

The PEO framework helps ensure that research questions remain focused and relevant. Population refers to the participants involved in a study, such as individuals, organisations, or groups. Exposure refers to the phenomenon or experience to which the population is exposed, such as an event, policy, or intervention. Outcome refers to the impacts or results of the exposure on the population, such as changes in behaviour or perceptions (van der Waldt, 2025). The PEO framework has been widely used in social science, health, and technology research to examine a range of phenomena, for example, antecedents of involuntary retirement

(MacLean, 2022), air pollution and noise (Mayntz et al., 2024), and digital transformation and Industry 4.0 (Quynh and Buics, 2024).

3. Research Design and Method

3.1 Database and Article Screening

As this study focused on the use of GenAI in higher education student assessments, the ERIC database was selected for its specific coverage of educational literature, rather than broader multidisciplinary databases such as Scopus and Web of Science. Because OpenAI launched ChatGPT on 30 November 2022 (OpenAI, 2022), this study searched the ERIC database for primary empirical studies published in English-language peer-reviewed journals from that date until 12 March 2026, when the search was conducted. The search used a combination of key terms and searched the full text of the articles. Figure 1 presents the search string and article screening criteria.

The search yielded 769 articles. Following an initial screening, this study removed 689 articles that did not match the inclusion criteria and selected 80 articles for full-text screening. The full-text screening process further removed 66 articles. Thus, 14 articles were selected for the thematic analysis.

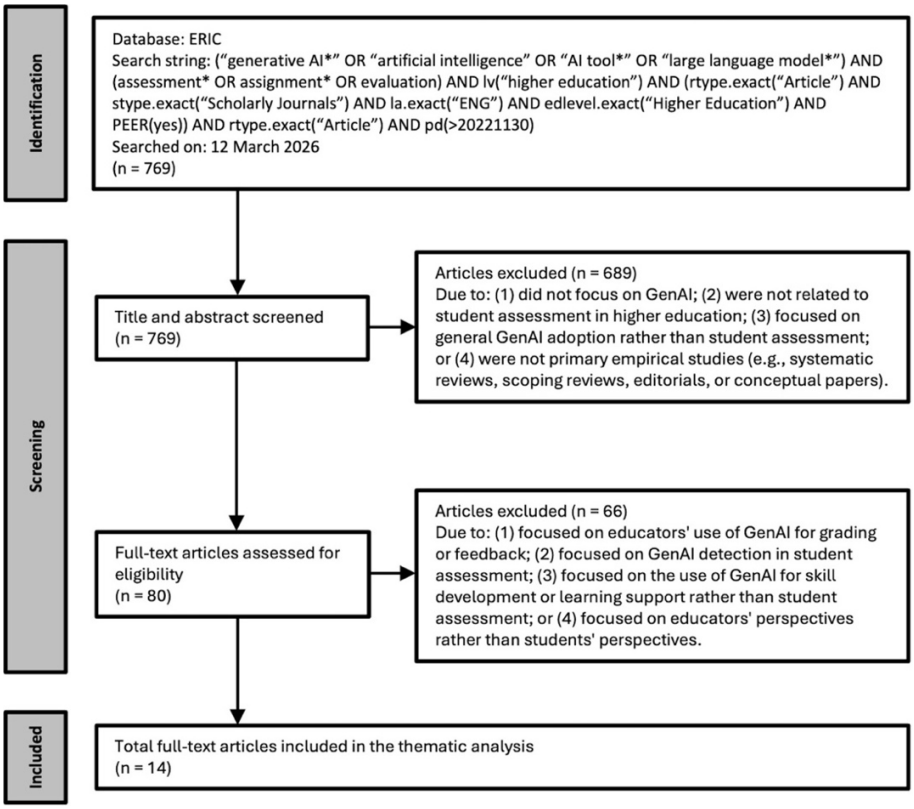


Figure 1: Flow diagram of the screening process (adopted from the PRISMA 2020 statement, Page et al., 2021)

3.2 Data Analysis

This study adopted a hybrid deductive and inductive codebook approach to thematic analysis (see Braun and Clarke, 2022). The 14 articles selected through the full-text screening process formed the dataset for the analysis. To maintain consistency, the same coder performed all coding throughout the analysis using QualCoder 3.8.2 (Curtain et al., 2025).

The study used the a priori PEO framework in a deductive approach to guide the analysis, i.e., population (higher education students), exposure (GenAI tools), and outcomes (students' experiences associated with GenAI use). Because the analysis focused on how the use of GenAI tools influenced students' experiences, the coder adopted an inductive approach to code the relevant data extracts related to students' experiences.

Since a single data extract could be assigned to multiple codes, the coder carefully reviewed each data extract to ensure that all relevant meanings were captured. This process generated a preliminary list of 34 codes. Because this study involved only one coder, the coder conducted two separate coding rounds two weeks apart to ensure intra-coder reliability (Schreier, 2012).

The coder subsequently reviewed the preliminary codes to identify conceptual similarities. Following an iterative approach, similar codes were merged, reducing the final number of codes to 29. These 29 codes were organised into four main themes and 13 subthemes, which capture the distinct dimensions of students' experiences associated with the use of GenAI tools.

4. Study Findings

50% of the studies employed a mixed-methods research design, whereas the remaining studies adopted qualitative approaches such as case study, phenomenological study, and exploratory study. These studies used various data collection methods, such as interviews, surveys, experiments, and text analysis of reflections, Padlet discussions, and written work. The studies primarily involved undergraduate and postgraduate students from universities in the US, the UK, Switzerland, and Australia.

The studies exposed participants to a range of GenAI-supported assessment activities, such as essay writing, formative e-assessments, creative writing, research tasks, exam preparation, and language learning. About 60% of the studies focused specifically on ChatGPT, whereas the remaining studies were not limited to a specific GenAI tool.

The analysis resulted in 29 codes organised into four main themes and 13 subthemes. The four themes capture the uses, perceived benefits, challenges and limitations, and academic integrity and ethical implications associated with GenAI use in assessments. Figure 2 presents a thematic map that summarises the study's findings.

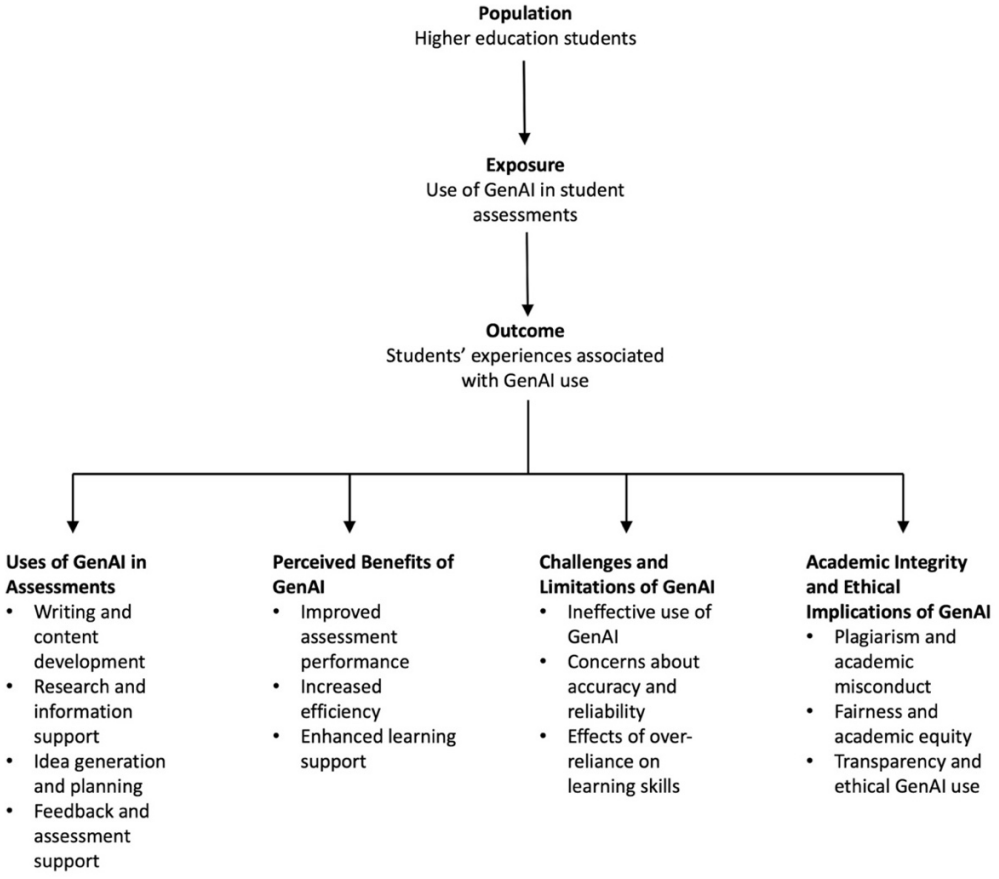


Figure 2: Thematic map

4.1 Uses of GenAI in Assessments

This main theme concerns how participants used GenAI tools to assist with assessment completion. The analysis identified four subthemes: (1) writing and content development, including content generation and drafting (e.g., drafting written work) and writing improvement (e.g., grammar checking and paraphrasing) (Johnston et al., 2025; Reinke et al., 2025; Wang, 2025; Williams, 2025); (2) research and information support, including information search (e.g., gathering background information) and research support (e.g., summarising texts and verifying information) (Behrens et al., 2025; Huang et al., 2024; Johnston et al., 2025; Vinyard and Roosa, 2025); (3) idea generation and planning, including brainstorming and idea development (e.g., generating ideas) and assignment planning and structuring (e.g., creating outlines and assignment structures) (Cotelli Kureth et al., 2025; Johnston et al., 2025; Wang, 2025; Williams, 2025); and (4) feedback and assessment support, including receiving feedback (e.g., obtaining timely responses and suggestions) and revising work (e.g., using feedback to refine work) (Behrens et al., 2025; Johnston et al., 2025; Song and Tang, 2025; Tossell et al., 2024).

4.2 Perceived Benefits of GenAI

This main theme concerns the perceived benefits of GenAI tools from the participants' perspective. The analysis identified three subthemes: (1) improved assessment performance, including higher-quality written work (e.g., improved essay quality and structure) and better assessment outcomes (e.g., higher grades and writing scores) (Baines et al., 2026; Cotelli Kureth et al., 2025; Song and Tang, 2025; Vincent and Lear, 2025); (2) increased efficiency, including time savings (e.g., completing tasks more quickly and receiving quick responses) and reduced cognitive effort (e.g., reducing cognitive load) (Baines et al., 2026; Huang et al., 2024; Wang, 2025; Williams, 2025); and (3) enhanced learning support, including improved conceptual understanding (e.g., obtaining definitions and explanations), increased confidence (e.g., greater confidence in completing assessments), and self-directed learning (e.g., reduced dependence on instructional staff) (Cotelli Kureth et al., 2025; Johnston et al., 2025; Reinke et al., 2025; Vincent and Lear, 2025).

4.3 Challenges and limitations of GenAI

This main theme concerns the issues, concerns, and limitations participants encountered or perceived when using GenAI tools. The analysis identified three subthemes: (1) ineffective use of GenAI, including misuse of GenAI (e.g., using unverified GenAI-generated outputs) and variable GenAI literacy (e.g., ineffective prompt creation and varying levels of GenAI proficiency) (Finkel-Gates, 2025; Johnston et al., 2025; Tossell et al., 2024; Williams, 2025); (2) concerns about accuracy and reliability, including quality and credibility of GenAI-generated outputs (e.g., inaccurate and misleading information) and citation and source verification (e.g., fake citations and lack of citations) (Behrens et al., 2025; Johnston et al., 2025; Tossell et al., 2024; Vinyard and Roosa, 2025); and (3) effects of over-reliance on learning skills, including reduced critical thinking skills (e.g., taking shortcuts and undermining critical thinking), diminished creativity and originality (e.g., producing less creative and original work), and over-reliance on GenAI (e.g., excessive dependence on GenAI) (Reinke et al., 2025; Song and Tang, 2025; Wang, 2025; Williams, 2025).

4.4 Academic Integrity and Ethical Implications of GenAI

This main theme concerns academic integrity and ethical issues associated with the use of GenAI tools. The analysis identified three subthemes: (1) plagiarism and academic misconduct, including plagiarism (e.g., writing substantial portions of assignments) and academic dishonesty (e.g., cheating and unfair academic practices) (Cotelli Kureth et al., 2025; Finkel-Gates, 2025; Vincent and Lear, 2025; Vinyard and Roosa, 2025); (2) fairness and academic equity, including unfair academic advantages (e.g., providing some students with an advantage over others) and inequitable access and use (e.g., disparities in access to or effective use of GenAI) (Behrens et al., 2025; Finkel-Gates, 2025; Vinyard and Roosa, 2025; Williams, 2025); and (3) transparency and ethical GenAI use, including the need for clear guidelines (e.g., guidance on appropriate GenAI use), transparency in GenAI use (e.g., disclosure regarding GenAI use), and ethical GenAI engagement (e.g., using GenAI responsibly) (Cotelli Kureth et al., 2025; Song and Tang, 2025; Vinyard and Roosa, 2025; Wang, 2025).

5. Discussion and Conclusions

The study's findings suggest that participants used GenAI for various assessment-related purposes, including writing and content development, research and information support, idea generation and planning, and feedback and assessment support. Participants used GenAI for activities such as generating ideas, organising

content, obtaining explanations, and receiving feedback (Cotelli Kureth et al., 2025; Huang et al., 2024; Johnston et al., 2025). These findings support the notion that students primarily use GenAI as a learning support tool rather than as a substitute for their own academic work (Hao et al., 2025; Tseng and Lin, 2024). This underscores the role of GenAI in assisting students in completing assessments (Furze et al., 2024; Zirar, 2023).

The study's findings also identified several benefits that participants associated with their use of GenAI. Participants reported improved assessment performance, increased efficiency, and enhanced learning support. GenAI was perceived to benefit participants in areas such as improving the quality of written work, enhancing conceptual understanding, reducing cognitive effort, and increasing confidence (Baines et al., 2026; Vincent and Lear, 2025; Wang, 2025). These findings are consistent with the notion that GenAI can provide personalised academic support with immediate responses and guidance tailored to students' needs (Keyhan, 2025; Zheldibayeva, 2025). This suggests that GenAI has the potential to contribute positively to student learning, particularly when used to support conceptual understanding and independent learning (Slobodsky and Durcheva, 2025; Wang, 2026).

Despite these benefits, the study also revealed several challenges and limitations that concerned participants, including ineffective GenAI use, inaccurate and unreliable GenAI-generated outputs, and reduced learning skills due to over-reliance on GenAI. Participants highlighted issues such as inaccurate information, fake citations, difficulty verifying GenAI-generated content, and undermining critical thinking and creativity (Behrens et al., 2025; Cotelli Kureth et al., 2025; Reinke et al., 2025). These findings are consistent with the challenges and limitations of GenAI reported in previous studies (Alzubi et al., 2025; Sarwanti et al., 2024). This suggests that although GenAI can enhance learning, its effectiveness depends on students' adequate proficiency in GenAI competencies (Asio, 2024; Naamati-Schneider and Alt, 2024).

The study's findings revealed that participants were concerned about academic integrity and ethical issues related to plagiarism and academic misconduct, fairness and academic equity, and transparency and ethical GenAI use. Participants highlighted concerns such as plagiarism, unfair academic advantages, and the need for clear institutional guidelines (Behrens et al., 2025; Cotelli Kureth et al., 2025; Johnston et al., 2025). These findings are consistent with previous studies that suggest concerns about fairness, accountability, and academic conduct in relation to the use of GenAI (Alrayes et al., 2024; Tsai et al., 2024). This reinforces the importance of developing institutional guidelines and assessment policies that support responsible GenAI use (Gonsalves, 2025; Mpolomoka, 2025).

In conclusion, the study's findings demonstrate that GenAI presents both opportunities and challenges in assessment contexts. While students perceive GenAI as a valuable tool for supporting learning and assisting in assessments, they also express concerns about reliability, over-reliance, and academic integrity. Therefore, the successful integration of GenAI into higher education is likely to depend on balancing its educational benefits with appropriate institutional guidelines and supporting students in developing the GenAI competencies required to use it effectively and responsibly (Asio, 2024; Gonsalves, 2025).

5.1 Research Limitations

This study has two main limitations. First, the thematic analysis included only 14 studies, which may limit the breadth and generalisability of the findings. Second, these studies were primarily conducted at universities in the US, the UK, Switzerland, and Australia. Consequently, the findings may not fully reflect students' experiences with GenAI use in different educational, cultural, or institutional contexts.

Ethics and AI Declaration

This study did not require ethics approval as it involved only a review of published articles, and the use of AI was mainly for preliminary searches and grammatical checks.

References

Akiba, D. and Garte, R. (2024) "Leveraging AI Tools in University Writing Instruction: Enhancing Student Success While Upholding Academic Integrity", *Journal of Interactive Learning Research*, Vol 35 No 4, pp 467-480.

Alrayes, A., Henari, T.F. and Ahmed, D.A. (2024) "ChatGPT in Education: Understanding the Bahraini Academics Perspective", *Electronic Journal of E-Learning*, Vol 22 No 2, pp 112-134.

Alzubi, A.A.F., Nazim, M. and Alyami, N. (2025) "Do AI-generative Tools Kill or Nurture Creativity in EFL Teaching and Learning?", *Education and Information Technologies*, Vol 30 No 11, pp 15147-15184.

Asio, J.M.R. (2024) "AI literacy, Self-Efficacy, and Self-Competence Among College Students: Variances and Interrelationships Among Variables", *Malaysian Online Journal of Educational Sciences,* Vol 12 No 3, pp 44-60.
Aydeniz, M. (2025) "Reflections on The Merit and Perils of AI in Higher Education: Five Early Adopter's Perspectives", *International Journal of Technology in Education and Science,* Vol 9 No 4, pp 522-544.
Baines, S., Patrao, A., Zhupa, X., Gramcheva, L., Paskalev, V. and Pauldy, C.J.O. (2026) "Evaluating the Impact of ChatGPT on Student Performance in Academic Writing", *International Journal of Technology in Education,* Vol 9 No 1, pp 109-126.
Behrens, K.A., Marbach-Ad, G. and Kocher, T.D. (2025) "AI in the Genetics Classroom: A Useful Tool but not a Replacement for Creative Writing", *Journal of Science Education and Technology*, Vol 34 No 4, pp 621-635.
Biju, N., Nasser Said, G.A., Bakiyeva, K., Prasad, K.D.V. and Jember, B. (2024) "Which One? AI-assisted Language Assessment or Paper Format: An Exploration of the Impacts on Foreign Language Anxiety, Learning Attitudes, Motivation, and Writing Performance", *Language Testing in Asia*, Vol 14, Article 45.
Braun, V. and Clarke, V. (2022) *Thematic Analysis: A Practical Guide*, SAGE, Thousand Oaks.
Chiu, T.K.F. (2024) "The Impact of Generative AI (GenAI) on Practices, Policies and Research Direction In Education: A Case of ChatGPT and Midjourney", *Interactive Learning Environments*, Vol 32 No 10, pp 6187-6203.
Chkarka, F. and Fatmi, H. (2026) "Comparative Examination of Master's Students' and Faculty Members' Maintenance of Academic Integrity in the Age of AI", *Journal of Academic Ethics,* Vol 24, Article 10.
Comscore (2026) "Consumer AI Chatbot Usage Rankings Show Claude Gaining Share*",* [online], Available at: https://www.comscore.com/Insights/Press-Releases/2026/5/Comscore-March-2026-Consumer-AI-Chatbot-Usage-Rankings-Show-Claude-Gaining-Share (Accessed 30 May 2026).
Cong-Lem, N., Tran, T.N. and Nguyen, T.T. (2024) "Academic Integrity in the Age of Generative AI: Perceptions and Responses of Vietnamese EFL Teachers" *Teaching English with Technology,* Vol 24 No 1, pp 28-48.
Cotelli Kureth, S., Paliot, E. and Zink, S. (2025). "Fostering Transparency: A Critical Introduction of Generative AI in Students' Assignments", *Language Learning in Higher Education*, Vol 15 No 1, pp 63-85.
Curtain, C., Dröge, K., Missaghieh-Poncet, J. and Salomón, L. (2025) *QualCoder 3.8.2* [Computer software]. Retrieved from https://github.com/ccbogel/QualCoder/releases/tag/3.8.2
Finkel-Gates, A. (2025) "ChatGPT in Academic Assessments: Upholding Integrity", *Journal of Learning Development in Higher Education*, Vol 36, pp 1-20.
Furze, L., Perkins, M., Roe, J. and MacVaugh, J. (2024) "The AI Assessment Scale (AIAS) in Action: A Pilot Implementation of GenAI-supported Assessment", *Australasian Journal of Educational Technology*, Vol 40 No 4, pp 38-55.
Gonsalves, C. (2025) "Addressing Student Non-Compliance in AI Use Declarations: Implications for Academic Integrity and Assessment in Higher Education", *Assessment and Evaluation in Higher Education*, Vol 50 No 4, pp 592-606.
Hao, T., Zhao, X. and Lu, Z. (2025) "Spontaneous Use of Generative Artificial Intelligence in Preparing for College English Test: GenAI Adoption Factors, Learners' Motivation, and Self-Efficacy", *Technology in Language Teaching & Learning*, Vol 7 No 3, Article 102640.
Huang, D., Huang, Y. and Cummings, J.J. (2024) "Exploring the Integration and Utilisation of Generative AI in Formative E-Assessments: A Case Study in Higher Education", *Australasian Journal of Educational Technology*, Vol 40 No 4, pp 1-19.
Johnston, H., Eaton, M.L., Henry, I., Deeley, E.M. and Parsons, B.N. (2025) "Discovering How Students Use Generative Artificial Intelligence Tools for Academic Writing Purposes", *Journal of Learning Development in Higher Education*, Vol 34, pp 1-28.
Keyhan, J. (2025) "Designing a Model for an AI-Based Intelligent Assistant for Personalized Learning in Higher Education", *International Journal of Technology in Education and Science,* Vol 9 No 2, pp 255-269.
Kirsanov, O., Kushwah, L. and Selvaretnam, G. (2026) "Beyond Detection: How Students Use--and Hide--AI in Online Assessments and What Authentic Tasks Can Do About It", *Journal of Academic Ethics,* Vol 24, Article 14.
MacLean, M.B. (2022) "Factors Explaining Involuntary Retirement: A Systematic Review", *The Journal of Aging and Social Change*, Vol 12 No 2, pp 19-35.
Mayntz, S.P., Rosenbech, K.E., Mohamed, R.A., Lindholt, J.S., Diederichsen, A.C.P., Frohn, L.M. and Lambrechtsen, J. (2024) "Impact of Air Pollution and Noise Exposure on Cardiovascular Disease Incidence and Mortality: A Systematic Review", *Heliyon*, Vol 10 No 21, Article e39844.
Microsoft (2026) "Global AI Adoption in 2025 A Widening Digital Divide", [online], Available at: https://www.microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025/ (Accessed 30 May 2026).
Mpolomoka, D.L. (2025) "Utilizing Artificial Intelligence for Assessment in Higher Education", *Pedagogical Research*, Vol 10 No 3, Article em0243.
Naamati-Schneider, L. and Alt, D. (2024) "Beyond Digital Literacy: The Era of AI-Powered Assistants and Evolving User Skills", *Education and Information Technologies*, Vol 29 No 16, pp 21263-21293.
Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., . . . Sandison, C. (2023) "ChatGPT Versus Engineering Education Assessment: A Multidisciplinary and Multi-Institutional Benchmarking and Analysis of this Generative Artificial Intelligence Tool to Investigate Assessment Integrity", *European Journal of Engineering Education*, Vol 48 No 4, pp 559-614.
Oc, Y., Gonsalves, C. and La, T.Q. (2025) "Generative AI in Higher Education Assessments: Examining Risk and Tech-Savviness on Student's Adoption", *Journal of Marketing Education*, Vol 47 No 2, pp 138-155.
OpenAI (2022) "Introducing ChatGPT", [online], Available at: https://openai.com/index/chatgpt/ (Accessed 30 May 2026).

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., . . . Moher, D. (2021) "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews", *BMJ*, Vol 372, No 71, pp 1-9.

Quynh, T.N.N. and Buics, L. (2024) "The Evolution of Digitalization Transformation and Industry 4.0 in Supply Chain Management: A Systematic Literature Review", *Engineering Proceedings*, Vol 79 No 1, Article 65.

Reinke, N.B., Parkinson, A.L. and Kafer, G.R. (2025) "A Tutorial Activity for Students to Experience Generative Artificial Intelligence: Students' Perceptions and Actions", *Advances in Physiology Education*, Vol 49 No 2, pp 461-470.

Sarwanti, S., Sariasih, Y., Rahmatika, L., Islam, M.M. and Riantina, E.M. (2024) "Are They Literate on ChatGPT? University Language Students' Perceptions, Benefits and Challenges in Higher Education Learning", *Online Learning,* Vol 28 No 3, pp 105-130.

Schreier, M. (2012) *Qualitative Content Analysis in Practice*, SAGE, Thousand Oaks.

Singh, R.G. and Ngai, C.S.B. (2024) "Top-ranked U.S. and U.K.'s Universities' First Responses to GenAI: Key Themes, Emotions, and Pedagogical Implications for Teaching and Learning", *Discover Education*, Vol 3 No 1, Article 115.

Slobodsky, P. and Durcheva, M. (2025) "Using ChatGPT Errors to Encourage Student Exploration and Self-Learning in Mathematics", *International Journal of Mathematical Education in Science and Technology*, Vol 56 No 8, pp 1401-1425.

Song, D. and Tang, A.F. (2025) "Exploring AI-assistance in L2 Chinese Writing with Standardized Assessment Tasks", *Language Learning & Technology*, Vol 29 No 2, pp 161-189.

Statista (2026a) "Generative AI - Worldwide", [online], Available at: https://www.statista.com/outlook/tmo/artificial-intelligence/generative-ai/worldwide?srsltid=AfmBOorh-hxJz1GNttUBm_BAaEmZYPAMBci1xca4hl1m_X8zn_ap7zG5#market-size (Accessed 30 May 2026).

Statista (2026b) "Generative Artificial Intelligence (AI) - Statistics & Facts", [online], Available at: https://www.statista.com/topics/10408/generative-artificial-intelligence/?srsltid=AfmBOoosQyYgUWk33hwAxZfrzxdqGnBnJ0w1ZYFHSO7Ca4yw_0yiHY0f#topicOverview (Accessed 30 May 2026).

Tossell, C.C., Tenhundfeld, N.L., Momen, A., Cooley, K. and de Visser, E.J. (2024) "Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence", *IEEE Transactions on Learning Technologies*, Vol 17, pp 1069-1081.

Tsai, C.Y., Lin, Y.T. and Brown, I.K. (2024) "Impacts of ChatGPT-assisted Writing for EFL English Majors: Feasibility and Challenges", *Education and Information Technologies*, Vol 29 No 17, pp 22427-22445.

Tseng, Y.C. and Lin, Y.H. (2024) "Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design", *Electronic Journal of E-Learning*, Vol 22 No 2, pp 78-97.

van der Waldt, G. (2025) "Formulating Research Questions in Social Science Research", *Journal for Transdisciplinary Research in Southern Africa*, Vol 21 No 1, pp 1-10.

Vincent, M.A. and Lear, B.J. (2025) "Using Exam Preparation and Reflection to Introduce Artificial Intelligence Tools in Honors General Chemistry", *Journal of Chemical Education*, Vol 102 No 10, pp 4470-4478.

Vinyard, M. and Roosa, M. (2025) "Student Perspectives on Using Generative Artificial Intelligence for Research: A Qualitative Approach", *Portal: Libraries and the Academy,* Vol 25 No 4, pp 729-752.

Wang, C. (2025) "Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers", *Technology, Knowledge and Learning*, Vol 30 No 3, pp 1825-1846.

Wang, Q. (2026) "Construction and Empirical Analysis of an Artificial Intelligence-Based Educational Assessment Model", *International Journal of Information and Communication Technology Education*, Vol 22 No 1, pp 1-16.

Wangdi, T., Rigdel, K.S., Dawa, T. and Tshering, K. (2025) "Using ChatGPT As an Assessment Tool in Education: A Systematic Literature Review of Practices and Limitations", *Issues in Educational Research*, Vol 35 No 2, pp 818-837.

Williams, A. (2025) "Integrating Artificial Intelligence Into Higher Education Assessment", *Intersection: A Journal at the Intersection of Assessment and Learning*, Vol 6 No 1, pp 128-154.

Xia, Q., Weng, X., Ouyang, F., Lin, T.J. and Chiu, T.K.F. (2024) "A Scoping Review on How Generative Artificial Intelligence Transforms Assessment in Higher Education", *International Journal of Educational Technology in Higher Education*, Vol 21 No 1, Article 40.

Zheldibayeva, R. (2025) "GenAI as a Learning Buddy for Non-English Majors: Effects on Listening and Writing Performance", *Education Process: International Journal*, Vol 14 No 1, Article e2025051.

Zirar, A. (2023) "Exploring The Impact of Language Models, such As ChatGPT, on Student Learning and Assessment", *Review of Education*, Vol 11 No 3, Article e3433.

Modality Matters for Extraversion: A Big Five Meta-Regression in Online Learning Environments

Eisuke Tokiwa
Sunblaze Co., Ltd., Tokyo, Japan

eisuke.tokiwa@sunblaze.jp

Abstract: The author's parent meta-analysis (Research Square preprint, DOI 10.21203/rs.3.rs-9513298/v1) pooled k = 9–10 primary studies per Big Five trait across online-learning settings; only Conscientiousness retained a robust positive correlation with achievement (r = 0.167 [0.089, 0.243]). Modality was pre-registered as a moderator but not executed because no cell met the k ≥ 10 rule. The present paper completes that registered-but-unrun moderator. Modality codes were re-checked against original PDFs for four primary-pool studies whose preprint extraction was blank (A-15 Elvers; A-23 Rodrigues; A-26 Wang; A-30 Kaspar). REML with HKSJ-adjusted CIs were used; a long-format weighted-OLS interaction model contrasted asynchronous (A) and mixed-online (M) instruction. Modality moderates the trait–achievement correlation in a trait-specific way: Q_between is highly significant for Extraversion (Q = 17.60, df = 1, p < .001), trend-level for Agreeableness (Q = 3.28, p = .070), and not significant for Conscientiousness, Neuroticism, or Openness. Asynchronous Extraversion shows r = −0.121 [−0.246, 0.007], reversing in mixed-online to r = +0.059 [−0.027, 0.145]; the Extraversion × Mixed interaction is +0.385 Fisher z, p = .009. The joint Wald test on all four interaction terms is $\chi^2(4) = 13.64$, p = .0085. Modality does not change the bottom line for Conscientiousness (r = 0.180–0.190 across A and M) but materially changes Extraversion, a trait the trait-only pool reported as null. For online course designers the message is asymmetric: support self-regulation regardless of synchrony, but expect Extraversion to be a silent disadvantage in async-only programmes.

Keywords: Meta-Analysis, Big Five, Online Learning, Learning Modality, Extraversion.

1. Introduction

The Big Five personality traits — Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N) — have been examined as predictors of academic achievement for over four decades (Poropat, 2009; Vedel, 2014; Mammadov, 2022). Meta-analytic syntheses of face-to-face samples converge on Conscientiousness as the dominant personality predictor (rho ~ 0.20-0.27 corrected; Poropat 2009), with smaller positive contributions from Openness and weaker or null contributions from the other three traits. Whether this established pattern transfers cleanly to online learning environments is, however, a more recent and less settled question.

The author's own systematic review and meta-analysis, deposited as a preprint on Research Square in April 2026 (DOI 10.21203/rs.3.rs-9513298/v1), is the first quantitative synthesis dedicated specifically to online learning environments. Across k = 9 - 10 primary studies per trait (pooled N = 3,384), only Conscientiousness retained a robust positive pooled correlation with achievement (r = 0.167, 95 % CI [0.089, 0.243]). Openness, Agreeableness, Neuroticism and especially Extraversion (r = 0.002) were null or directionally inconsistent. Two pre-registered moderators reached significance at p < .001: Extraversion x Region (Asian samples r = -0.131 vs non-Asian r = +0.050) and Extraversion x Outcome Type (objective r = -0.038 vs self-rated r = +0.117).

The preprint is, however, *silent* on a moderator that is theoretically central to its scope: learning modality. The Methods Deviations subsection explicitly notes that modality, education level, instrument family, publication year, log-sample-size, and risk-of-bias score were pre-registered as moderators but did not meet the k >= 10 per-level rule and were therefore reported only narratively. Modality is the most theoretically pressing of these omissions because it directly conditions the four dimensions of online learning that distinguish it from face-to-face teaching: social presence, temporal flexibility, self-regulation demand, and platform-mediated affordance.

Three theoretical predictions follow if modality matters. First, synchronous instruction reintroduces social cues and instructor-led pacing; if Extraversion's null result in the trait-only pool is driven by social-cue deprivation, synchronous samples should re-acquire a positive Extraversion-achievement link. Second, asynchronous instruction shifts the load to learner self-regulation (Pintrich, 2004; Zimmerman, 2000); Conscientiousness — already the dominant predictor in the pool — should remain so or strengthen further. Third, mixed-online and blended modalities sit between these poles and are likely to produce intermediate, more heterogeneous outcomes empirically.

This paper completes the modality moderator analysis that the parent preprint pre-registered but did not quantitatively execute, and adds a long-format modality x trait interaction model. Its contribution is therefore both methodological (closing a registered analytic gap) and substantive (reporting a trait-specific modality effect

that the trait-only pool obscures). It does not duplicate the preprint: every modality-stratified estimate, every interaction coefficient, and the joint Wald test reported below are absent from the preprint Results section (audit log: *preprint_audit.md*).

2. Method

2.1 Inheritance from the Parent Preprint

The corpus, eligibility criteria, screening flow, and risk-of-bias scoring are inherited unchanged from the parent preprint. PRISMA 2020 standards were followed (Page et al., 2021); the corresponding flow diagram is reproduced in Figure 1. Of the 31 primary studies catalogued at full-text assessment, 25 were retained for qualitative synthesis and 12 contributed at least one extractable Pearson correlation to the primary achievement pool. Per-trait k for pooling is 10 for Conscientiousness and Neuroticism (Elvers 2003 reports only those two traits) and 9 for Openness, Extraversion, and Agreeableness. Detailed extraction sheets are in *metaanalysis/analysis/data_extraction_populated.csv*; the canonical replication of the parent pool is in *metaanalysis/analysis/CANONICAL_RESULTS.md*.

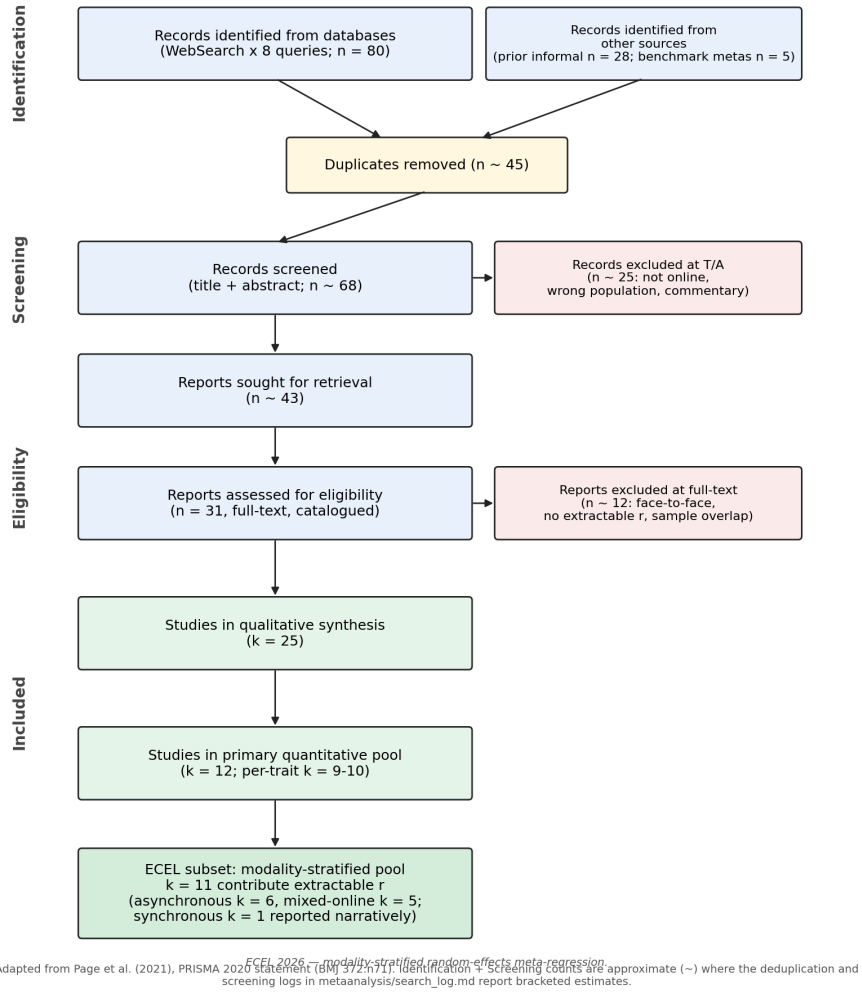


Figure 1: PRISMA 2020 flow diagram. Identification → Screening → Eligibility → Included counts trace from *metaanalysis/search_log.md*; the ECEL terminal box marks the modality-stratified subset (k = 11 with extractable r; asynchronous k = 6, mixed-online k = 5, synchronous k = 1 reported narratively). Adapted from Page et al. (2021), *BMJ*, 372, n71.

2.2 Modality Coding

A key limitation of modality-stratified meta-analysis is that primary studies often do not report their delivery format explicitly, and even when they do, the boundary between "asynchronous" and "mixed-online" can be

ambiguous. The re-coding decisions documented below should be interpreted with this caveat in mind; alternative plausible codings could shift per-cell k and the resulting Q_between contrasts.

Modality categories were operationalised as five mutually exclusive levels:

- A (asynchronous): self-paced LMS or MOOC; no scheduled live sessions.
- M (mixed-online): both synchronous and asynchronous components within a single online course.
- S (synchronous): live videoconferencing-driven instruction.
- B (blended): mix of online and face-to-face.
- U (unspecified): the source paper does not state the format.

The master extraction coded modality where the source PDF stated it explicitly. Four primary-pool studies had a blank modality field; their modality was re-checked against the original PDFs: A-15 Elvers (2003) was coded A (self-paced LMS, no live sessions), while A-23 Rodrigues (2024), A-26 Wang (2023), and A-30 Kaspar (2023) were coded M (evidence of both synchronous and asynchronous components). After override, no unspecified-modality studies remain in the primary pool.

2.3 Statistical Model

For each (trait, modality) cell with k >= 2, a random-effects pool was computed using REML estimation of tau-squared and a Hartung-Knapp-Sidik-Jonkman (HKSJ) adjustment for the confidence interval (Hartung & Knapp, 2001). Effect sizes were Fisher's z-transformed Pearson correlations with sampling variance v = 1 / (N - 3); the back-transformed pooled r and its CI are reported. Synchronous studies (k = 1) are reported narratively only. Between-modality heterogeneity was tested via Q_between with the sub-pool point estimates (df = 1 per trait). To formalise the interaction, a weighted-OLS regression was fit on the full long-format design matrix (n = 42 observations across 5 traits and 2 modality levels), with weights = 1 / (v + tau-squared), tau-squared set to the median of per-trait REML estimates (= 0.0105). The design includes trait dummies, a modality dummy, and four trait x modality interaction terms; a joint Wald chi-squared test on the interaction coefficients is reported as the primary inferential statistic. Three sensitivity scenarios re-run the modality pools: (i) drop beta-converted studies, (ii) drop the COI study (A-25 Tokiwa; no-op since A-25 contributes no extractable r), and (iii) drop the unspecified-modality bucket (no-op after override).

2.4 Reproducibility

The analysis pipeline and all intermediate results are available in the project's open repository. All computations are deterministic; no pseudorandom-number-generator-dependent operations are used.

3. Results

3.1 Replication of the Parent Preprint

For sanity-check purposes, the trait-only pooled correlations match the parent preprint exactly (within rounding): Conscientiousness r = 0.167 [0.089, 0.243]; Openness r = 0.086 [-0.044, 0.214]; Extraversion r = 0.002 [-0.076, 0.080]; Agreeableness r = 0.112 [-0.032, 0.250]; Neuroticism r = 0.018 [-0.079, 0.114]. Heterogeneity ranges from I-squared = 65 % (Conscientiousness) to 96 % (Agreeableness).

3.2 Modality-Stratified Pooled Correlations

Table 1 reports the per-modality x per-trait pooled correlations.

Table 1: Modality-stratified random-effects pooled correlations (REML + HKSJ).

Trait	Modality	k	N	r [95 % CI]	I-squared	tau-squared
O	A	3	1393	0.223 [-0.374, 0.689]	92.0 %	0.036
O	M	5	1445	0.019 [-0.081, 0.119]	47.8 %	0.002
O	S	1	525	0.070 (k=1, narrative)	—	—

Trait	Modality	k	N	r [95 % CI]	I-squared	tau-squared
C	A	4	1414	0.190 [-0.034, 0.395]	63.5 %	0.007
C	M	5	1445	0.180 [0.051, 0.302]	67.5 %	0.005
C	S	1	525	0.068 (k=1, narrative)	—	—
E	A	3	1393	-0.121 [-0.246, 0.007]	19.6 %	0.000
E	M	5	1445	0.059 [-0.027, 0.145]	27.2 %	0.000
E	S	1	525	0.027 (k=1, narrative)	—	—
A	A	3	1393	0.283 [-0.293, 0.708]	93.7 %	0.035
A	M	5	1445	0.032 [-0.076, 0.138]	55.2 %	0.003
A	S	1	525	-0.013 (k=1, narrative)	—	—
N	A	4	1414	0.076 [-0.001, 0.165]	0.0 %	0.000
N	M	5	1445	-0.001 [-0.059, 0.058]	0.0 %	0.000
N	S	1	525	-0.072 (k=1, narrative)	—	—

Two patterns are immediately visible (see also Figure 2). First, Conscientiousness pooled r is essentially identical across the asynchronous and mixed-online buckets (0.190 vs 0.180), with overlapping CIs. Second, Extraversion pooled r reverses sign across modality (-0.121 in A, +0.059 in M); the asynchronous CI brushes zero from below (upper bound 0.007).

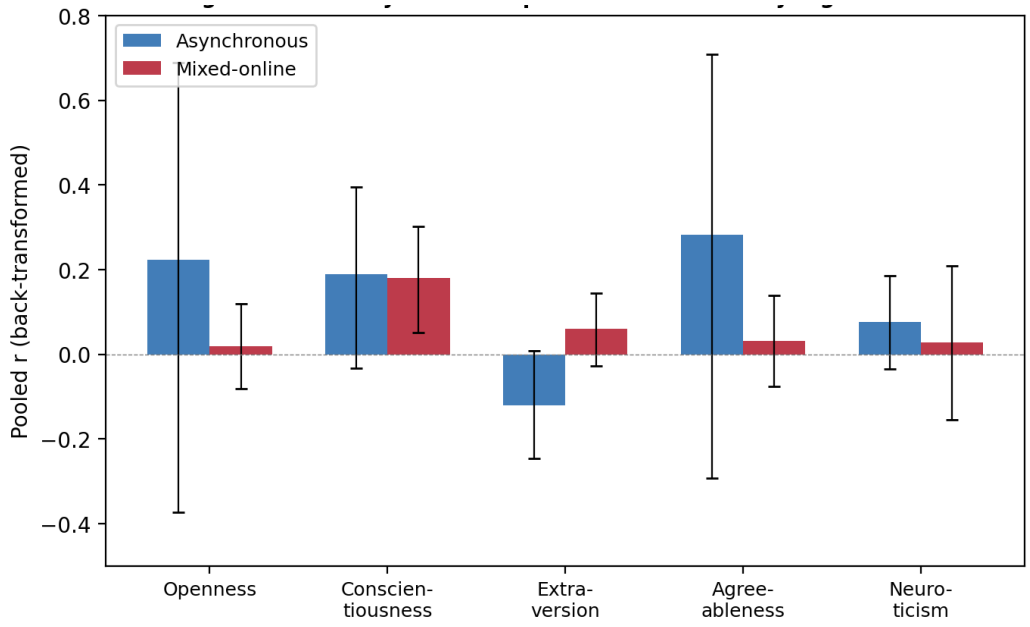


Figure 2: Modality-stratified pooled correlations by Big Five trait (asynchronous = blue; mixed-online = red). Error bars show HKSJ-adjusted 95 % CIs. The Extraversion sign-reversal across modality is the headline finding; Conscientiousness is stable across modality buckets.

3.3 Q_between Across Modality Levels

Table 2 reports the per-trait Q_between contrast.

Table 2: Q_between random-effects contrast across modality levels (df = 1; A vs M).

Trait	Q_between	p_between
O	1.96	.161
C	0.015	.903
E	17.60	< .001
A	3.28	.070
N	0.42	.519

Only Extraversion shows a clear modality dependence; Agreeableness shows a trend (p = .070); the other three traits are flat across modality. Conscientiousness is exceptionally close to zero on Q (0.015), which is itself diagnostic — a trait so tightly tied to self-regulation might have been expected to vary by modality, and the fact that it does *not* is informative.

3.4 Modality x Trait Interaction Model

Table 3 reports the full long-format interaction model (n = 42 observations; reference cell trait = O, modality = A; tau-squared used for weighting = 0.0105).

Table 3: Long-format weighted-OLS interaction model (Fisher z scale).

Term	Estimate	SE	t	p
(Intercept)	0.238	0.079	3.03	.005
trait[C]	-0.034	0.109	-0.31	.758
trait[E]	-0.342	0.111	-3.07	.004
trait[A]	0.084	0.111	0.75	.459
trait[N]	-0.205	0.109	-1.87	.070
modality[M]	-0.226	0.098	-2.31	.027
trait[C]:modality[M]	0.203	0.137	1.49	.147
trait[E]:modality[M]	0.385	0.138	2.78	.009
trait[A]:modality[M]	-0.061	0.138	-0.44	.662
trait[N]:modality[M]	0.222	0.137	1.62	.114

The joint Wald test on the four interaction coefficients is chi-squared(4) = 13.64, p = .0085, indicating that modality moderates trait effects in the aggregate. Inspection of the individual coefficients localises the effect to Extraversion: the trait[E]:modality[M] interaction is +0.385 Fisher z (p = .009), large enough to flip the asynchronous negative sign into a positive effect when modality is mixed.

3.5 Sensitivity

Under the beta-converted-drop scenario (excludes A-28 Yu and A-30 Kaspar), the asynchronous Extraversion cell loses one study (k = 3 -> k = 2) and the CI widens substantially, but the negative point estimate is preserved (-0.121 -> -0.095). The mixed-online cell loses A-30 (k = 5 -> k = 4) with negligible change (+0.059 -> +0.044). The COI-drop and unspecified-modality-drop scenarios are no-ops by construction. Full sensitivity tables are in *results/sensitivity_analysis_summary.md*. The headline interaction effect is qualitatively robust under all three scenarios.

4. Discussion

4.1 The Conscientiousness Invariance

The most striking pattern in Table 1 is what *does not* vary: Conscientiousness pooled r is 0.190 in the asynchronous bucket and 0.180 in the mixed-online bucket — an absolute difference of 0.010 on the back-transformed scale, with a Q_between p of .903 (about as null as a heterogeneity test can be). The parent preprint's headline that Conscientiousness is the dominant Big Five predictor of online-learning achievement holds *at finer modality resolution*. From a practical standpoint this means course designers cannot escape the

Conscientiousness effect by adjusting synchrony — students who score high on self-regulation traits will outperform those who do not in both async and mixed-online programmes, by roughly the same margin.

4.2 The Extraversion Modality Dependence

The headline empirical contribution is the modality dependence of Extraversion. In asynchronous courses the pooled r is -0.121 [-0.246, 0.007] — small, with the upper CI limit just brushing zero — while in mixed-online courses it flips to +0.059. The interaction term trait[E]:modality[M] is +0.385 Fisher z (p = .009), and the joint Wald test on all four interaction terms is chi-squared(4) = 13.64, p = .0085. These results align with the social-presence-of-extraverts hypothesis: extraverts thrive when there is a social channel and lose ground when there is not. A purely asynchronous course removes the live, real-time interaction in which extraversion's sociability functions as an asset; in fact, the literature is converging on the view that extraverts in asynchronous environments may even procrastinate more on solitary activities (Cheng 2023; Quigley 2022), translating into mildly negative achievement effects.

The k = 1 synchronous evidence (A-29 Bahçekapılı 2020 GPA, r = +0.027) is reported only narratively but is consistent with this directional pattern: introducing live elements should restore Extraversion's neutral-to-positive role. A single study cannot carry inferential weight, but the prediction it sets (extraversion is positive in synchronous-heavy courses) is testable in any future synthesis with k > 1 in the synchronous cell.

4.3 Other Traits

- Openness: The asynchronous bucket point estimate is +0.223 with an enormous CI [-0.374, 0.689] driven by k = 3 and high I-squared. The mixed bucket is essentially zero (+0.019). Q_between p = .161 — not significant. We cannot rule out a real Openness x modality interaction but the data do not support one.
- Agreeableness: A trend pattern parallel to Extraversion (A r = +0.283, M r = +0.032; Q_between p = .070), but again k = 3 in async and a wide CI. Worth flagging for future replication; not actionable now.
- Neuroticism: The async cell shows a small positive effect (r = +0.076, CI just touching zero), the mixed cell is essentially null. No modality dependence (p = .519).

4.4 Self-Plagiarism Firewall

This paper completes a moderator analysis the parent preprint *registered* but elected not to *execute* on k-grounds. Both the modality-stratified table and the long-format interaction model are absent from the preprint Methods and Results sections. The disclosure block in the cover letter (template: *metaanalysis/conference_submissions/templates/preprint_disclosure_template.md*) makes this provenance explicit so reviewers can see the boundary between the preprint and the present contribution. The parent preprint will, for now, remain on Research Square and will not be withdrawn; the future journal version of the parent will pick up the modality moderator from this paper as one of its acknowledged extensions.

5. Limitations

The cell-level k is small (k = 1 for synchronous; k = 3 - 5 for the analyzed cells). With k of this size, individual studies materially affect single-cell precision; the beta-converted-drop sensitivity scenario halves the precision of the asynchronous Extraversion cell while preserving the negative sign. Inferences about individual cell estimates should therefore be treated as exploratory; the joint interaction test (which aggregates across cells) carries the inferential weight of this paper.

Modality re-coding for the four U-bucket studies is necessarily an inferential step. The original PDFs contain language consistent with the assigned modality but do not always state synchrony / asynchrony explicitly. The preprint footnote to Table 1 reports that "the asynchronous/synchronous distinction was explicitly reported for only 9 studies"; the present paper extends this to 12 by adding the 4 PDF-grounded inferences, all of which are documented with citations in *preprint_audit.md* section 4. A second-coder audit would strengthen the coding but is not feasible for a single-author submission.

The corpus is dominated by undergraduate samples (8 of 12 primary-pool studies); the modality-stratified findings are therefore best interpreted as undergraduate-online effects. K-12 (A-25, A-26) and graduate (A-37) cells are too sparse to break out separately. The blended (B) bucket is empty in the primary pool because the only blended-eligible study (A-02 Alkis) reports its online subsample separately, which is captured here as M.

The interaction model uses a simple weighted-OLS approximation in lieu of a full random-intercept mixed model; for this corpus the difference is small in practice (the per-trait REML tau-squared estimates are in the range 0.000 - 0.036, and the median used in weighting is 0.0105), but a full mixed-model implementation in *metafor* or *lme4* would be a useful follow-up replication.

6. Conclusion

A modality lens partially changes the Big Five — online-learning-achievement picture. The Conscientiousness story is unchanged: it remains the dominant positive predictor at modality-stratified resolution. But Extraversion, which the trait-only pool reports as null (r = .002), is in fact modality-dependent: a small negative effect in asynchronous courses, reversing to a small positive effect in mixed-online courses, with a joint interaction Wald test at p = .0085.

Two implications for course designers and educational researchers follow. First, in asynchronous-dominant programmes, Big Five-based personalisation should target self-regulation supports rather than personality-fit recommendations: there is no "right" trait profile, only a robust Conscientiousness gradient. Second, course-designer claims that extraverts succeed in online environments cannot be applied uniformly — they hold (weakly) in mixed-online and synchronous formats and *invert* in asynchronous-only formats. This trait-by-modality contingency is exactly the pattern self-regulated-learning theory (Pintrich, 2004; Zimmerman, 2000) predicts.

The dominant message for the field, however, is structural: most online-learning meta-analyses to date have treated modality as noise, in part because original-paper reporting of synchrony is sparse. Of 27 fully-online studies in this corpus, only 9 report sync/async explicitly. Future syntheses should refuse to extract studies that fail to code synchrony, and editors of online-learning journals should require it as a reporting standard.

Acknowledgments

This paper was prepared by the sole author. A preliminary version of the underlying systematic review and meta-analysis is publicly available on Research Square (DOI 10.21203/rs.3.rs-9513298/v1, posted 27 April 2026); the present submission completes the modality moderator that was pre-registered but not quantitatively executed in that preprint, and adds a long-format trait-by-modality interaction model. No co-authors contributed and no external funding was received. The author's ORCID is 0009-0009-7124-6669.

Ethics Declaration

This paper reports a systematic review and meta-analysis based exclusively on previously published research and on the author's own previously published primary study (Tokiwa, 2025, *Frontiers in Psychology* 16:1420996, CC BY 4.0). No new human-subjects data were collected by the author for this paper. Ethical clearance was therefore not required.

AI Declaration

The author used a large language model (Anthropic's Claude) as a research assistant throughout this project. Specific uses included: drafting and editing of the manuscript text; generating Python scripts for the meta-analysis pipeline (REML pooling, Hartung–Knapp–Sidik–Jonkman CI adjustment, Q_between contrasts, weighted-OLS interaction model); formatting the manuscript to ACPI submission specifications; and assistance with literature search support and reference verification. All statistical analyses were re-executed by the author and traced to source CSVs; all bibliographic entries were verified against the corresponding primary-source PDFs; all interpretive and substantive content was reviewed and approved by the author. No AI-generated text, figure, or numerical claim was included in the manuscript without explicit author review and validation.

References

Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *The Internet and Higher Education, 45*, 100724. https://doi.org/10.1016/j.iheduc.2019.100724

Alkış, N., & Taşkaya Temizel, T. (2018). The impact of motivation and personality on academic performance in online and blended learning environments. *Educational Technology & Society, 21*(3), 35–47.

Bahçekapılı, E., & Karaman, S. (2020). A path analysis of five-factor personality traits, self-efficacy, academic locus of control and academic achievement among online students. *Knowledge Management & E-Learning, 12*(2), 191–208. https://doi.org/10.34105/j.kmel.2020.12.010

Cheng, S.-L., Chang, J.-C., Quilantan-Garza, K., & Gutierrez, M. L. (2023). Conscientiousness, prior experience, achievement emotions and academic procrastination in online learning environments. *British Journal of Educational Technology, 54*(4), 898–923. https://doi.org/10.1111/bjet.13302

Elvers, G. C., Polzella, D. J., & Graetz, K. (2003). Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology, 30*(2), 159–162. https://doi.org/10.1207/S15328023TOP3002_13

Hartung, J., & Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine, 20*(24), 3875–3889. https://doi.org/10.1002/sim.1009

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A, 172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Kaspar, K., Burtniak, K., & Rüth, M. (2023). Online learning during the Covid-19 pandemic: How university students' perceptions, engagement, and performance are related to their personal characteristics. *Current Psychology, 42*(30), 26571–26586. https://doi.org/10.1007/s12144-023-04403-9

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality, 90*(2), 222–255. https://doi.org/10.1111/jopy.12663

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996

Quigley, M., Bradley, A., Playfoot, D., & Harrad, R. (2022). Personality traits and stress perception as predictors of students' online engagement during the COVID-19 pandemic. *Personality and Individual Differences, 194*, 111645. https://doi.org/10.1016/j.paid.2022.111645

Rivers, D. J. (2021). The role of personality traits and online academic self-efficacy in acceptance, actual use and achievement in Moodle. *Education and Information Technologies, 26*(4), 4353–4378. https://doi.org/10.1007/s10639-021-10478-3

Rodrigues, J., Rose, R., & Hewig, J. (2024). The relation of Big Five personality traits on academic performance, well-being and home study satisfaction in Corona times. *European Journal of Investigation in Health, Psychology and Education, 14*(2), 368–384. https://doi.org/10.3390/ejihpe14020025

Tokiwa, E. (2025). Who excels in online learning in Japan? *Frontiers in Psychology, 16*, Article 1420996. https://doi.org/10.3389/fpsyg.2025.1420996

Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences, 71*, 66–76. https://doi.org/10.1016/j.paid.2014.07.011

Wang, P., Wang, F., & Li, Z. (2023). Exploring the ecosystem of K-12 online learning: An empirical study of impact mechanisms in the post-pandemic era. *Frontiers in Psychology, 14*, 1241477. https://doi.org/10.3389/fpsyg.2023.1241477

Yu, Z. (2021). The effects of gender, educational level, and personality on online learning outcomes during the COVID-19 pandemic. *International Journal of Educational Technology in Higher Education, 18*(1), Article 14. https://doi.org/10.1186/s41239-021-00252-3

Zheng, Y., & Zheng, S. (2023). Exploring educational impacts among pre, during and post COVID-19 lockdowns from students with different personality traits. *International Journal of Educational Technology in Higher Education, 20*(1), Article 21. https://doi.org/10.1186/s41239-023-00388-4

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

Design of an Educational Platform for Developing Librarians' Digital Competencies

Hana Ulrichová, Michaela Mrazova, Magdalena Zavodna and Tomas Javorcik
University of Ostrava, Faculty of Education, Department of Information and Comunication Technologies, Ostrava, Czech Republic

Hana.Ulrichova@osu.cz
Magdalena.Zavodna@osu.cz
Michaela.Mrazova@osu.cz
Tomas.Javorcik@osu.cz

Abstract: This paper summarizes the results of preliminary research aimed at defining the design requirements for a microlearning application with adaptive features that tailor content to users' individual preferences and prior knowledge levels. The survey involved 395 respondents from among library staff, and the data collected provides crucial insight into the preferences of this specific user group. The results indicate that visual simplicity and a high degree of clarity are key parameters for a successful educational platform. Respondents in the survey clearly preferred a simple, uncluttered design with a minimum of distracting elements. In terms of visual style, a preference for a light mode and an emphasis on readability dominated, even at the cost of displaying less information on the screen at once. The research also identified a need for clear navigation and comprehensible controls. Although standardized icons were found to be highly understandable, most users require accompanying text labels for controls to eliminate uncertainty. The main menu structure with five basic sections (Home, Lessons, Profile, Help, Settings) was evaluated as optimal and user-friendly. In terms of content and pedagogy, the most effective approach appears to be the implementation of microlearning principles, that is, breaking down the curriculum into short, time-efficient blocks lasting 5 to 10 minutes, which allow learners to study at their own pace. Respondents also emphasized the need for personalization, particularly the ability to choose a difficulty level based on their existing knowledge and the automatic adaptation of content format (text vs. video) to their learning styles. The research results show that for the effective development of digital competencies in the library sector, it is essential to design an interface that is intuitive, functional, and respects users' individual time and cognitive capabilities. The findings serve as a direct basis for the creation of a prototype microlearning application with adaptive features that will meet the real needs of library practice in the digital age.

Keywords: Educational Platform, Design, Librarians, Microlearning, Digital Competencies, Adaptive Learning

1. Introduction

The current development of digital technologies and the transformation of information institutions into multifunctional community and informal learning centers place enormous demands on the digital competencies of their staff. In both the European and national contexts, this need is systematically defined by the updated European Digital Competence Framework for Citizens (Cosgrove & Cachia, 2025). In addition to data literacy, cybersecurity, and content creation, it places special emphasis on the ethical and safe use of artificial intelligence and digital well-being (Cosgrove & Cachia, 2025). This phenomenon is particularly pronounced in the Czech library sector. Professionals in the domestic library sector have long since ceased to serve merely as intermediaries for print collections, they are becoming key guides through complex data ecosystems, administrators of local information networks, and instructors in media, digital, and information literacy for the general public.

However, the accelerating pace of technological innovation in the Czech Republic is encountering significant demographic and age-related challenges. Within the Czech socioeconomic context, the library community is characterized by a markedly skewed age structure that skews toward older age groups (Pillerová, 2025).

This is not an exclusively Czech phenomenon. A large-scale DigComp-based study of Hungarian public library staff (N = 1,868) found that many professionals remain at basic or intermediate levels in digital content creation, safety and problem solving, while formal library education is a stronger predictor of advanced proficiency than work experience (Borbély, 2022). A systematic review further showed that although e-learning is the main approach to addressing these gaps, its effectiveness is often limited by workload, insufficient institutional support and poorly designed learning environments (Shahzad et al., 2023).

Continuing professional education thus faces a barrier where generic, technologically complex educational platforms fail to account for the physiological, cognitive, and user-specific needs of older adults. In practice, this

mismatch between interface design and user capabilities leads to premature cognitive exhaustion, demotivation, and an increase in so-called “technostress” (Tarafdar et al., 2019).

The severity of this problem from a demographic perspective is explicitly demonstrated by comprehensive statistical data from the National Library of the Czech Republic, derived from nationwide personnel surveys (Pillerová 2025). According to these empirical findings, the population of domestic librarians is undergoing a continuous aging process: a full 72% of all professional staff are over 41 years of age. The largest age groups are people aged 41 to 50 (30%) and the immediately following category of 51 to 60 (29%). Senior staff over the age of 61 make up a significant 13% of the total workforce, while the share of the youngest generation under 30 has been declining over the long term and stands at a mere 9% (the study also notes a high degree of feminization in the sector, where women account for 85% of employees).

These data clearly demonstrate that the primary users of any newly developed educational system in Czech librarianship are professionals of mature or advanced age. This demographic profile naturally brings with it specific physiological changes, such as a natural decline in visual acuity, reduced sensitivity to color contrast, or specific demands on the motor precision required to operate software controls (Whitenton & Nielsen, 2019). The visual design and interaction architecture of educational systems must therefore strictly adhere to human-centered design methodology and accessibility standards, such as the WCAG 2.2 guidelines, and must not be based solely on the aesthetic trends of young developers (W3C, 2023).

1.1 Problem Domain and User-Specific Considerations

The main research and practical issue addressed in this paper is the lack of design standards for educational interfaces that would reflect the actual needs of older age groups in the Czech library sector. Existing robust educational platforms, including widely used learning management systems (LMS Moodle), tend to be text-heavy, visually cluttered, and have an unintuitive control structure, which contradicts the basic principles of multimedia learning (Mayer, 2020). For librarians who face high workloads and often work in smaller regional institutions with limited opportunities for substitution, studying in such systems represents an additional burden. Instead of absorbing educational content, they must allocate a significant portion of their mental capacity to deciphering a complicated and confusing interface. According to cognitive load theory (Sweller, 2011), there is a rapid increase in external load, which has been shown to reduce computer self-efficacy (Compeau & Higgins, 1995) and triggers a paralyzing fear of making a mistake.

In response to these shortcomings, a project was launched to develop a specialized microlearning web application with adaptive features, designed to effectively build digital skills. However, in order to design a functional visual prototype of the interface, it was essential to conduct robust preliminary research focused on the interface’s ergonomic and aesthetic parameters. The research study thus deliberately distanced itself from the educational content itself and thoroughly analyzed the user preferences of Czech librarians in the area of visual ergonomics. The resulting application design is thus methodologically grounded in empirical requirements gathered from 395 respondents with practical experience.

1.2 Theoretical Foundations: Microlearning, Adaptability, and Cognitive Ergonomics

When seeking the optimal design solution for a demographically specific group such as Czech librarians, it is essential to integrate insights from cognitive ergonomics with modern instructional approaches. Traditional forms of e-learning, characterized by long text-based monologues and robust modules, often encounter working memory limitations in older adults (Sweller, 2011). The concept of microlearning therefore appears to be an effective alternative, defining education as the process of absorbing small, thematically coherent, and time-fragmented units (Hug, 2007). These micro-content blocks (typically lasting 5 to 10 minutes) minimize the risk of cognitive overload and allow users to study flexibly throughout the workday, which directly addresses the needs of library staff with high workloads.

Recent empirical evidence supports this reasoning. A systematic review found that mobile-based microlearning improves knowledge retention and motivation, with interface simplicity and short, self-paced learning as key success factors (Moore, Hwang and Moses, 2024). Similarly, an app-based intervention among older adult care personnel, a workforce comparable to Czech librarians, identified ease of use, accessibility and timely knowledge delivery as critical, with the least digitally experienced employees being most affected by interface friction (Richardson et al., 2023). Zhang and West (2020) further argue that microlearning is most effective when aligned with a competency framework, supporting the proposed platform’s alignment with DigComp 3.0.

The second key pillar of the proposed solution is the adaptability of the interface and content. Adaptive learning systems dynamically modify their behavior, structure, or presentation of information based on the user's individual characteristics, preferences, and current knowledge (Brusilovsky, 2001). A systematic mapping of AI-enabled adaptive learning systems found that adult and workplace populations remain underrepresented despite adaptation based on learners' prior knowledge and preferences (Kabudi, Pappas and Olsen, 2021). The present project addresses this gap.

In the context of developing librarians' digital competencies, adaptability plays a key role in removing barriers and strengthening computer self-efficacy. Older adults exhibit considerable heterogeneity in their initial digital skills. While fixed educational platforms generate fear of failure in less experienced users and boredom in advanced users, an adaptive system can, based on an initial assessment, personalize the difficulty level and tailor the format of educational content (e.g., automatically prioritizing text-based instructions with visual support over dynamic video, or vice versa) to the user's individual preferences. Although the empirical validity of the learning styles model in the sense of fixed cognitive types is questioned in the literature (Pashler et al., 2008), the preference-based selection of content format as a tool for reducing cognitive load and enhancing technological self-efficacy has solid support in the theory of multimedia learning (Mayer, 2020).

From the perspective of cognitive ergonomics and human-computer interaction (HCI), it is crucial that the interface of a microlearning application account for cognitive aging, which is associated with a decline in information processing speed and a reduction in selective attention capacity (Whitenton & Nielsen, 2019). These design implications have since been supported by a systematic review of 132 studies, which identified simplified navigation, larger text and touch targets, high contrast, error-tolerant interaction and clear labelling as key factors for usability among users aged 60 and over, while cognitive overload and low digital literacy remained the main barriers (Amouzadeh et al., 2025).

A minimalist visual style, restraint in color use, and the elimination of redundant decorative elements are not merely aesthetic choices but functional requirements for reducing so-called visual clutter, which has been shown to prolong the time older users spend searching for information on the screen. The comprehensibility of control elements, supported by explicit text labels for icons, then functions as cognitive scaffolding that transforms abstract software operations into understandable actions, thereby eliminating technostress and maximizing user comfort (Tarafdar et al., 2019).

1.3 Research Objectives and Research Questions

This study represents the initial empirical phase of the broader development and testing of a planned educational platform. Given the demographic characteristics and risks of cognitive overload among Czech librarians described above, a preliminary quantitative study was conducted. Its primary goal was to identify key user requirements for the visual, typographic, and didactic-interactive design of the interface. A total of 395 practitioners actively participated in this survey, providing a sufficiently robust data set to clearly define the research problem.

In order to transform the complex set of collected empirical indicators into comprehensive design recommendations, the research project was broken down into three key research questions:

- RQ1 (Visual-Aesthetic Domain): What parameters of visual ergonomics, typography, and color schemes do Czech librarians prefer in order to minimize eye strain and external cognitive load?
- RQ2 (Interaction-Orientation Domain): What architectural structure of navigation and what approach to the design of control elements (including iconography and textual contextualization) ensure optimal orientation and a sense of user confidence?
- RQ3 (Didactic-Content Area): What structural and time allocations for educational modules and what forms of content personalization correspond to the needs and capabilities of adult learners in the library sector?

2. Methodology

Given the nature of the research problem and the need to obtain a robust data set from a broad range of respondents, a quantitative research strategy was chosen, utilizing a questionnaire survey method, which was supplemented by a qualitative content analysis of open-ended responses.

2.1 Research Design and Data Collection Tool

Empirical data were collected using a specially designed, non-standardized questionnaire that was distributed electronically. The instrument was intentionally designed without demographic or professional questions (such as age or years of experience), as the overall demographic profile of the target group has been comprehensively and explicitly mapped over the long term by the National Library of the Czech Republic's comprehensive personnel surveys (Pillerová, 2025).

The questionnaire directly operationalized the research questions (RQ1–RQ3) through 19 items (including 12 icon-comprehensibility test items) structured into three thematic blocks:

- Visual-Aesthetic Domain and Cognitive Ergonomics (RQ1): Evaluated preferences regarding interface simplicity, visual clutter, typography, and color schemes.
- Interaction-Oriented Architecture and Navigation (RQ2): Assessed navigation clarity, confidence in digital environments, preference for text labels over standalone icons, icon comprehension, and the five-section menu layout.
- Didactic Structure and Personalization (RQ3): Examined interest in microlearning modules, preferred content formats, self-paced learning, and personalization options (e.g., difficulty levels, offline PDF downloads).

The questionnaire combined closed-ended questions with single- or multiple-choice options, supplemented by five-point Likert scales (expressing the degree of agreement or preference from "strongly disagree" to "strongly agree") and a total of six open-ended text questions allowing for detailed responses from the respondents.

2.2 Research Sample and Data Distribution

The target population of the research survey consisted of professionals in the library sector in the Czech Republic. The questionnaire was distributed using a method of complete purposive sampling. The director of the Library Institute of the National Library of the Czech Republic actively collaborated on the development of the application and the conduct of the research; through her and using the Library Institute's official communication infrastructure, the electronic questionnaire was distributed directly to all library institutions throughout the Czech Republic. This centralized approach ensured the maximum possible reach across the entire network of Czech libraries (from regional and specialized institutions to small regional branches). In total, N = 395 respondents actively participated in the survey by fully completing and submitting the questionnaire. Given the widespread distribution through the Library Institute of the National Library of the Czech Republic and the established demographic structure of the sector (see Chapter 1), the resulting dataset can be considered a highly representative sample reflecting the actual composition of the workforce in Czech librarianship.

2.3 Data Collection and Processing

Data collection took place between February 1 and February 28, 2026. After collection was completed, the data were exported to XLSX format, and then checked, cleaned, and coded. Given the descriptive and exploratory nature of this phase of the research, descriptive statistics were primarily applied (absolute and relative frequencies, means, medians). The results were then visualized using graphs and tables, which form the basis for the following chapter, Results.

3. Results

The presentation of the empirical study's results is systematically divided into three sections that directly correspond to the research questions. A total of N = 395 fully completed questionnaires were analyzed.

3.1 Visual-Aesthetic Aspects and Cognitive Ergonomics (RQ1)

Research has confirmed the critical need for visual minimalism. On a five-point scale (1 = lowest, 5 = highest), 57.4% of respondents rated their preference for a simple design at the highest level, as illustrated in Figure 1.

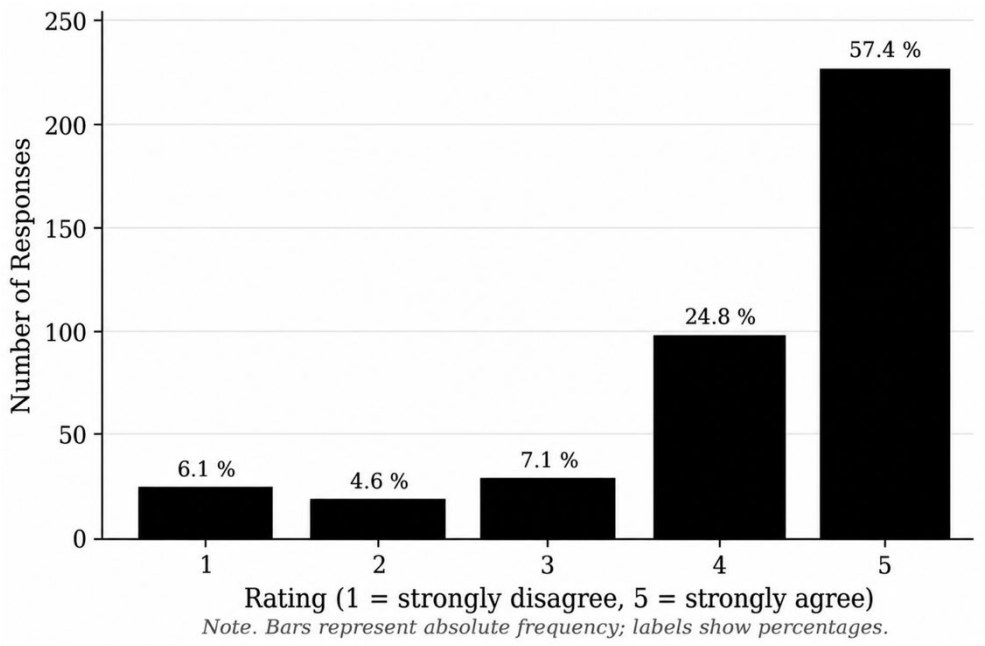


Figure 1: Preference for a simple, uncluttered interface

When it comes to typography, the largest group of respondents (38%) took a neutral stance on font size, but a total of 43.5% (the sum of scores 4 and 5) prefer a larger font even if it means less information (Figure 2). Qualitative data (Questions 9 and 19) supplement this requirement with the need to use a single consistent sans-serif font and ensure high text contrast against the background.

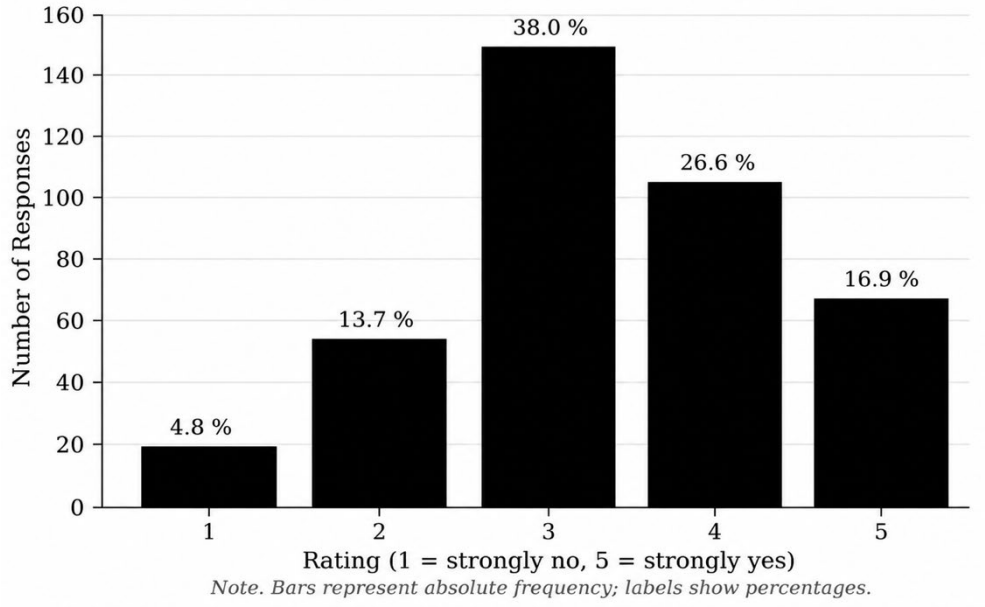


Figure 2: Preference for larger, readable font

Light mode dominates in terms of color scheme (68.1%) as presented in Figure 3. However, the open-ended responses (Question 19) indicated that users would welcome a dark mode as an optional alternative for those with low vision or for working in the evening. In their additional comments, respondents explicitly warned against “aggressive” and “intense” colors (e.g., neon shades) and animated elements, which they consider highly distracting.

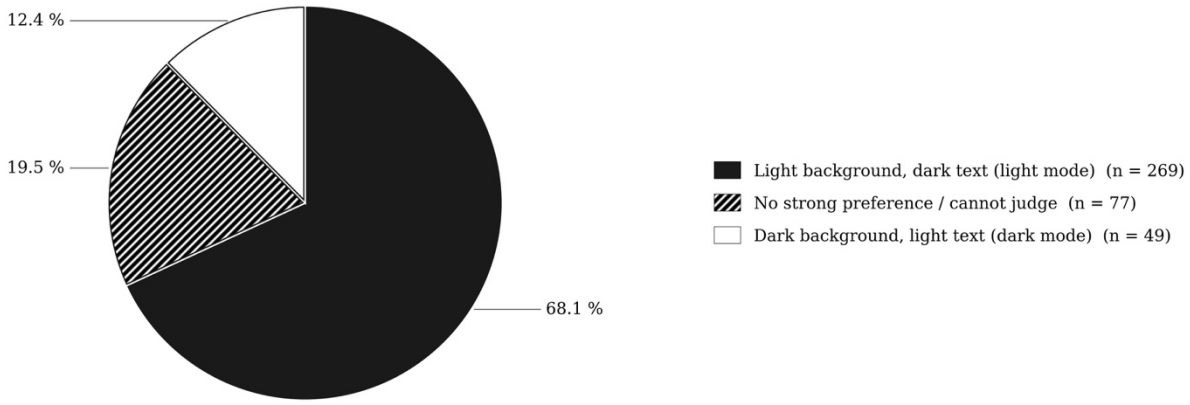


Figure 3: Preferred colour scheme

3.2 Interaction-Oriented Architecture and Navigation (RQ2)

Most librarians rate their ability to navigate new applications as good, with 48.6% experiencing no difficulty and 47.1% reporting some initial difficulty (Figure 4). Nevertheless, the proposed menu structure with five sections (Home, Lessons, Profile, Help, Settings) was accepted with a high degree of agreement (87.1%), as depicted in Figure 5.

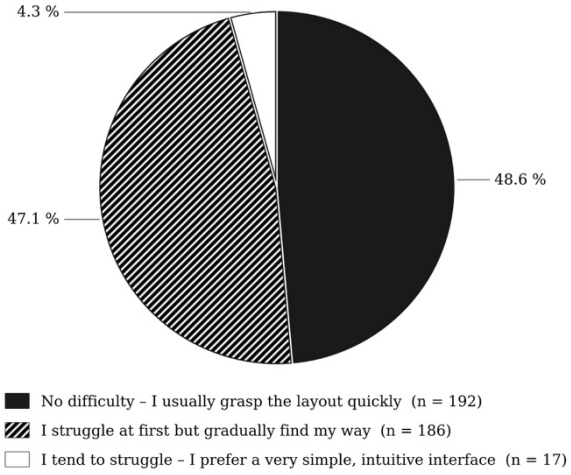


Figure 4: Self-rated ability to orient in a new application

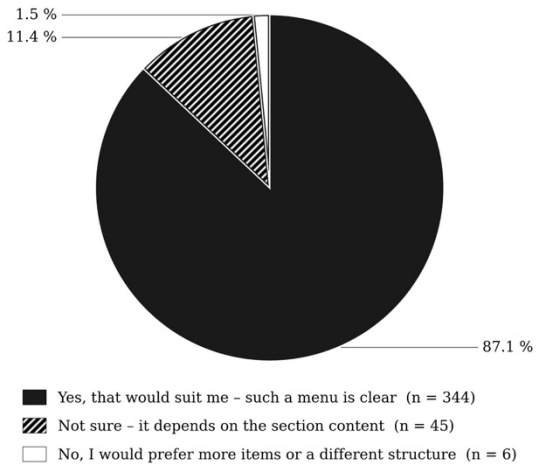


Figure 5: Clarity of 5-section main menu

An analysis of 73 relevant suggestions for improving navigation (Question 9) revealed that the most common request is for the implementation of a search bar (full-text search with autocomplete). Other priorities include:

- Breadcrumb navigation and a fixed ("pinned") page header for quick navigation.
- Chapter titles instead of mere numbering (e.g., "Digital Security" instead of "Lesson 1").
- Horizontal layout (inspired by websites like Seznam.cz, which 82 respondents cited as a model for an easy-to-use website).

A key finding concerns the use of icons. Although standardized icons (magnifying glass, house, gear) show a recognition rate of over 90%, recognition of the "task" icon (pencil) dropped to 57.2% (Figure 6). Furthermore, when asked about overall interface preferences (Question 5), 60.3% of respondents strongly preferred that graphical icons always be accompanied by explicit text labels.

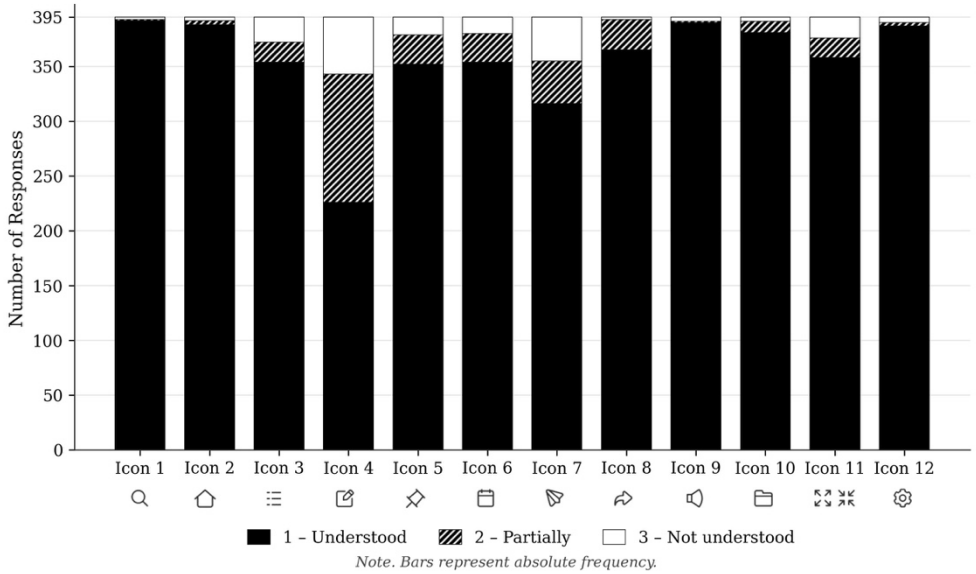


Figure 6: Icon comprehension rating across 12 tested icons

3.3 Didactic and Content Structure and Personalization (RQ3)

Respondents expressed exceptional interest in microlearning, a total of 88.1% (the sum of "Definitely yes" and "Somewhat yes" responses) would welcome lessons in 5–10-minute segments (Figure 7).

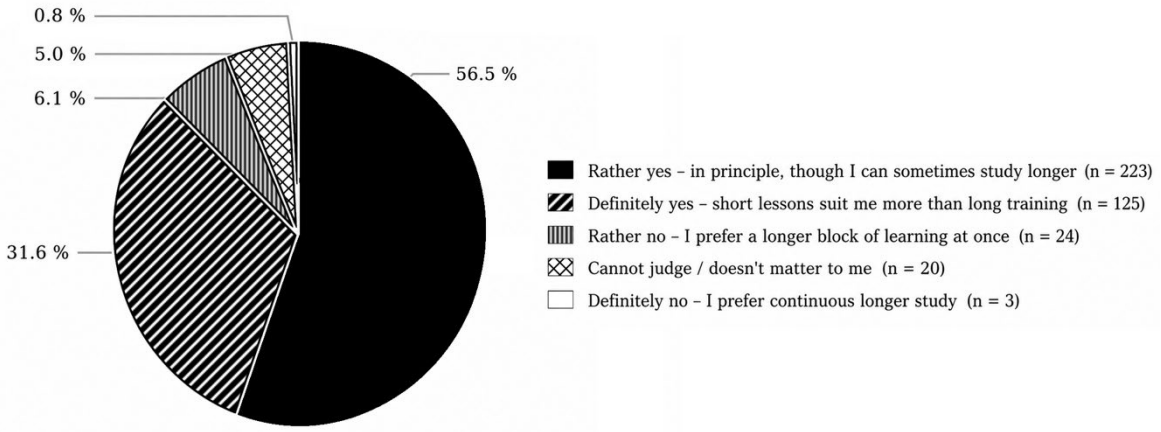


Figure 7: Interest in microlearning (5-10 min lessons)

As depicted in Figure 8, when it comes to preferred content, text-based materials (78%) and interactive elements (73.7%) dominate, followed by video (65.6%). App personalization is considered particularly important when it comes to selecting the difficulty level (62.3%).

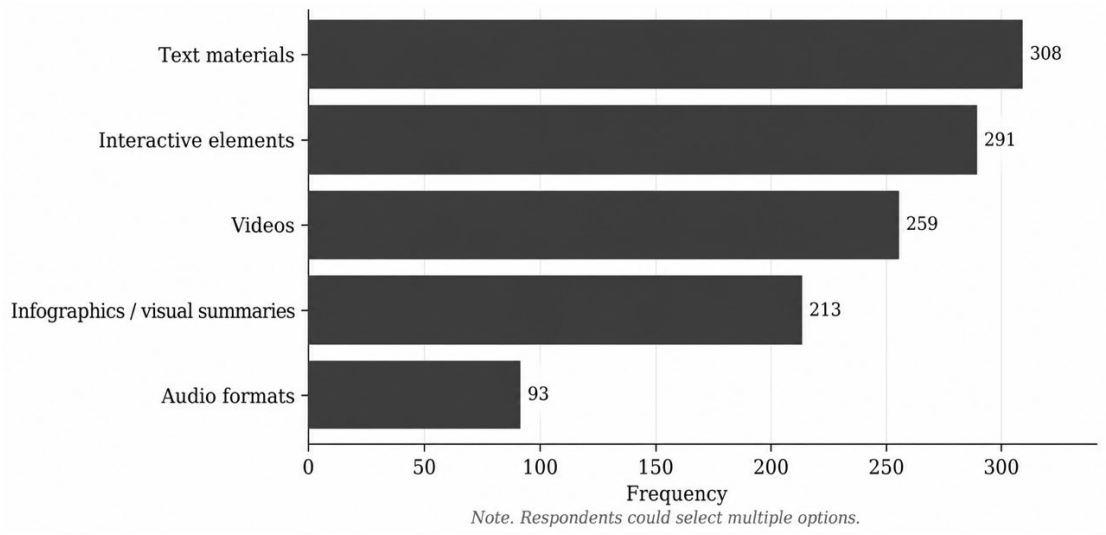


Figure 8: Most benefical content formats

In response to the open-ended questions (16 and 18), respondents overwhelmingly (128 mentions) emphasized the need to download materials (PDFs) for offline study and printing. Other frequent requests included:

- Gamification: Motivation through progress indicators, point collection, or a "roadmap" (inspired by the Duolingo app, which 26 respondents cited as a positive example).
- Flexibility: The ability to pause a lesson and return exactly to where the user left off.
- Practicality: Immediate feedback after completing a test and the option to consult with a tutor via a discussion forum.

When analyzing experiences with other systems (Question 17), the existing platforms were criticized for being overly theoretical and "outdated" (e.g., older courses in Moodle). In contrast, users appreciate modern, fast-paced, and thematically well-defined courses (e.g., projects on osveta.gov.cz or Seduo).

4. Discussion

This study provides empirical insights into the specific user requirements of Czech librarians in the areas of cognitive ergonomics and the interaction design of educational platforms. Given the significantly skewed age structure in this sector (Pillerová, 2025), the findings critically reflect theoretical frameworks related to cognitive aging and technostress (Tarafdar et al., 2019; Whitenton & Nielsen, 2019).

One of the most significant findings in the area of visual ergonomics (RQ1) is the respondents' strong emphasis on interface minimalism, simplicity (Figure 1) and high color contrast (Questions 9 and 19). These preferences directly correlate with Richard Mayer's multimedia principles (2020), specifically the principle of coherence, according to which people learn better when redundant and distracting elements are eliminated from the interface (referred to by respondents as "aggressive colors" or "flashing animations"). For older adults, who experience a natural decline in selective attention, visual restraint acts as a primary protective factor against premature cognitive exhaustion (Sweller, 2011). The observed preference for light mode (68.1%) presented in Figure 3 with the option to manually switch to dark mode (identified in the supplementary text section) aligns with the recommendations of Whitenton and Nielsen (2019) for the design of web interfaces intended for older users, where fixed dark modes can impair readability for people with visual impairments (e.g., astigmatism).

In terms of interaction architecture (RQ2), the research highlighted the phenomenon of technological uncertainty. Although the comprehensibility of standardized icons was high (Figure 6), the fact that 60.3% of users requested an accompanying text label (Question 5) indicates a need for so-called cognitive scaffolding. This converges with the review evidence of Amouzadeh et al. (2025), in which explicit labelling and error tolerance emerged as recurring requirements for users aged 60 and over; the present data extend that evidence from consumer and health applications to the domain of professional continuing education. An isolated icon (e.g., the pencil icon for editing, which showed a drop in comprehensibility to 57.2%, as demonstrated in Figure 6) causes older users to fear making a mistake, which demonstrably reduces their technological self-efficacy and induces technostress (Compeau & Higgins, 1995; Tarafdar et al., 2019). The proposed five-section structure of the main menu (Figure 5), which received overwhelming support (87.1%), together with the request for the

implementation of full-text search (Question 9), confirms that a shallow and transparent information architecture is key for this target group. The inspiration drawn from the Seznam.cz portal in the open-ended responses points to the validity of Jakob's Law of Usability – users prefer interfaces that function the same way as systems they are already familiar with (Whitenton & Nielsen, 2019).

The didactic dimension (RQ3) revealed a clear rejection of robust and rigid educational modules of the type found in traditional LMS platforms like Moodle, which respondents criticized in their open-ended responses for being confusing and outdated (Question 17). The overwhelming consensus on the implementation of microlearning in Figure 7 (88.1% for sessions lasting 5–10 minutes) reflects librarians' high workload and limited working memory capacity when absorbing complex technological content (Hug, 2007). The magnitude of this preference exceeds the levels of acceptance reported by Richardson et al. (2023) in a comparable older workforce, which may reflect the higher baseline information literacy of library professionals. Conversely, the demand for downloadable PDF materials identified here has no counterpart in that study and appears to be profession-specific — plausibly a reflection of the professional habitus of staff whose daily work is organised around print collections. Requests for personalized difficulty levels and automatic adaptation of content format (text vs. video as depicted in figure 8) confirm the theoretical premises of adaptive learning (Brusilovsky, 2001) as a tool for inclusive skill development. A significant and unforeseen finding in the questionnaire (derived from open-ended questions 16 and 18) was the overwhelming demand for downloadable PDF materials for printing. As this preference was not reported by Richardson et al. (2023), it may represent a profession-specific characteristic reflecting library staff's habitual reliance on print-based resources during learning.

4.1 Limitations of the Research and Future Directions

The main limitation of this study is its reliance on self-reported data (respondents' subjective preferences as indicated in the questionnaire). In theory, users may prefer different features than those they actually demonstrate during practical interaction with the software. However, this limitation will be eliminated in the immediately following phase of the project. The empirical data obtained and relevant user feedback (including requests for gamification modeled after the Duolingo app) serve as the direct basis for creating an interactive model and the subsequent high-fidelity prototype of the application. This limitation is consistent with the state of the field: reviews of both adaptive learning systems (Kabudi, Pappas and Olsen, 2021) and age-friendly interface design (Amouzadeh et al., 2025) identify the scarcity of longitudinal usability testing as a systemic weakness rather than a shortcoming specific to this study.

The second limitation is potential self-selection bias in the research sample. Although the questionnaire was distributed widely through the Library Institute of the National Library of the Czech Republic, respondents with higher levels of digital literacy and a positive attitude toward technology may have been more willing to complete the electronic questionnaire. The results may thus underestimate the extent of technological uncertainty and technostress among the least digitally proficient segment of the library population, precisely the group for which the application is primarily intended.

5. Conclusion

This article has successfully achieved its stated objective, which was to define and analyze the specific user, ergonomic, and pedagogical requirements of Czech library staff regarding the design of an educational microlearning application for developing digital competencies. Thanks to the widespread distribution of the questionnaire through the Library Institute of the National Library, a robust dataset was obtained from N = 395 respondents, which exhibits a high degree of representativeness with respect to the actual demographic composition of the Czech library sector.

The research clearly demonstrated that generic, visually saturated, and text-heavy e-learning platforms cannot be used for the effective education of a demographically older and operationally busy population of librarians. A successful design must strictly adhere to the principles of cognitive ergonomics, inclusive Human-Centered Design, and accessibility standards (WCAG 2.2). The key parameters for the successful implementation of the application in practice are:

- Visual simplicity: A minimalist interface in light mode with high contrast, free of distracting dynamic elements.
- Cognitive ease of navigation: A shallow menu structure (5 sections), graphic icons supplemented with explicit text labels, and the integration of robust full-text search.

- Microlearning pedagogy: Fragmentation of course material into short, 5- to 10-minute blocks that allow for flexible, self-paced study without disrupting the institution's operations.
- Adaptive personalization: Difficulty level selection based on an initial diagnostic assessment, choice of content format (text/video), and support for offline study (PDF downloads).

The findings represent a significant contribution to the discussion on methods of professional education for adults in information institutions. The findings of this study were directly applied in the design of the conceptual architecture of the system under development and serve as an empirical basis for the creation of a visual prototype of an application that will genuinely contribute to the systematic and stress-free development of digital competencies (in accordance with the DigComp 3.0 framework) in library practice in the digital age.

Acknowledgements

This research was funded by the Technology Agency of the Czech Republic (TAČR) under the SIGMA Programme through project No. TQ23000252, titled "Personalized Educational Platform for Developing Digital Competencies of Librarians".

Ethics Declaration

Participation in the research was voluntary. All participants were informed about the purpose of the study and agreed that the collected data could be processed and used for research purposes. The questionnaire responses were collected and processed in accordance with applicable data protection regulations and were used anonymously in the analysis and presentation of the results.

AI Declaration

DeepL was used for initial translation support during manuscript preparation. All translated text was subsequently reviewed, revised, and where necessary modified by the author. No AI tools were used for generating research content, analysis, or conclusions.

References

Amouzadeh, E., Dianat, I., Faradmal, J. and Babamiri, M. (2025) "Optimizing Mobile App Design for Older Adults: Systematic Review of Age-Friendly Design", Aging Clinical and Experimental Research, Vol 37, Article 248. doi:10.1007/s40520-025-03157-7.

Borbély, M. (2022) "Public Library Digital Competency Mapping 2019: A Survey on Digital Skills of Library Professionals With Different Qualifications", Frontiers in Education, Vol 7, Article 909502. doi:10.3389/feduc.2022.909502.

Brusilovsky, P. (2001) "Adaptive Hypermedia", User Modeling and User-Adapted Interaction, Vol 11, No. 1–2, pp 87–110. doi:10.1023/A:1011143116306.

Compeau, Deborah R. and Higgins, Christopher A. (1995) "Computer Self-Efficacy: Development of a Measure and Initial Test", MIS Quarterly, Vol 19, No. 2, pp. 189–211. https://doi.org/10.2307/249688

Cosgrove, J. and Cachia, R. (2025) DigComp 3.0: European Digital Competence Framework – Fifth Edition, JRC144121, Publications Office of the European Union, Luxembourg. doi:10.2760/0001149.

Hug, T. (ed.) (2007) Didactics of Microlearning: Concepts, Discourses and Examples, Waxmann, Münster.

Kabudi, T., Pappas, I. and Olsen, D.H. (2021) "AI-enabled Adaptive Learning Systems: A Systematic Mapping of the Literature", Computers and Education: Artificial Intelligence, Vol 2, Article 100017. doi:10.1016/j.caeai.2021.100017.

Mayer, R.E. (2020) Multimedia Learning, 3rd ed., Cambridge University Press, Cambridge.

Moore, R.L., Hwang, W. and Moses, J.D. (2024) "A Systematic Review of Mobile-Based Microlearning in Adult Learner Contexts", Educational Technology & Society, Vol 27, No. 1. 10.30191/ETS.202401_27(1).SP02

Pashler, Harold, McDaniel, Mark, Rohrer, Doug and Bjork, Robert (2008) "Learning Styles: Concepts and Evidence", Psychological Science in the Public Interest, Vol 9, No. 3, pp. 105–119.

Pillerová, V. (2025) "Jak se mění knihovnická profese? Výsledky pravidelného průzkumu", Knihovna: knihovnická revue, Vol 36, No. 2, pp 45–65.

Richardson, M.X., Aytar, O., Hess-Wiktor, K. and Wamala-Andersson, S. (2023) "Digital Microlearning for Training and Competency Development of Older Adult Care Personnel: Mixed Methods Intervention Study", JMIR Medical Education, Vol 9, e45177. doi:10.2196/45177.

Shahzad, K., Khan, S.A., Javed, Y. and Iqbal, A. (2023) "E-Learning for Continuing Professional Development of University Librarians: A Systematic Review", Sustainability, Vol 15, No. 1, Article 849. doi:10.3390/su15010849.

Sweller, John (2011) "Cognitive Load Theory", in Mestre, José P. and Ross, Brian H. (eds.) Psychology of Learning and Motivation, Vol. 55, Academic Press, pp. 37–76. doi:10.1016/B978-0-12-387691-1.00002-8

Tarafdar, M., Cooper, C.L. and Stich, J.-F. (2019) "The Technostress Trifecta – Techno Eustress, Techno Distress and Design", Information Systems Journal, Vol 29, No. 1, pp 6–42. doi:10.1111/isj.12169.

Whitenton, Kathryn and Nielsen, Jakob (2019) Senior Citizens (Ages 65 and Older) on the Web: UX Research Report, 4th ed., Nielsen Norman Group. Available at: https://www.nngroup.com/reports/senior-citizens-on-the-web/

W3C (2023) Web Content Accessibility Guidelines (WCAG) 2.2, World Wide Web Consortium. Available at: https://www.w3.org/TR/WCAG22/

Zhang, J. and West, R.E. (2020) "Designing Microlearning Instruction for Professional Development Through a Competency Based Approach", TechTrends, Vol 64, No. 2, pp 310–318. doi:10.1007/s11528-019-00449-4.

Transforming Teaching Through Blackboard Ultra Adoption: Lecturer Training, Pedagogical Integration and Sustainable LMS use in Higher Education

Elzette van Niekerk and Adriana Botha
University of Fort Hare, East London, South Africa

evanniekerk@ufh.ac.za
adriana@natsphere.com

Abstract: The increasing integration of learning management systems (LMSs) into higher education has intensified the need for lecturers to develop technological competence alongside the pedagogical and disciplinary knowledge required for meaningful digital teaching. This study examined lecturers' perceptions of Blackboard Learn Ultra training, their self-reported implementation of tools and strategies introduced during training, and the conditions they perceived as supportive of continued LMS use at a South African university. The empirical study used a post-training, cross-sectional survey design. An online questionnaire containing closed- and open-ended items was completed by 51 lecturers from a sampling frame of 200 lecturers who had participated in institutional Blackboard development initiatives. Quantitative data were analysed using descriptive statistics, with an exploratory principal component analysis (PCA), correlations and group comparisons reported cautiously because of the modest sample size relative to the 19 Likert items. Qualitative responses were analysed inductively using thematic analysis. TPACK was used as an interpretive framework rather than as a measure of change over time. Results showed that 84.3% of participants reported implementing Blackboard tools or strategies and 88.2% reported moderate or extensive integration. High ratings were recorded for confidence in using Blackboard (M = 4.45), relevance to teaching context (M = 4.45), and understanding how to integrate Blackboard into teaching (M = 4.31). The exploratory PCA identified three components explaining 75.28% of the variance; detailed PCA and group-comparison evidence is provided in Tables 1 and 2. Qualitative findings highlighted hands-on practice, assessment and grading, interactive learning, content organisation, time for consolidation, discipline-specific application and continuing support. The findings describe lecturers' perceptions and self-reported implementation following training; they do not establish causal effects of training or developmental change over time.

Keywords: Blackboard Learn Ultra; Learning Management System; Lecturer Training; TPACK; Professional Development; Technology-Enhanced Learning; Higher Education

1. Introduction

Learning management systems (LMSs) have become central to higher education, supporting content delivery, communication, assessment, feedback and collaboration. However, institutional access does not guarantee sustained or pedagogically meaningful use. Research on LMS adoption indicates that lecturers' use is shaped by factors such as perceived usefulness, ease of use, self-efficacy and institutional support (Matarirano et al., 2021; Almogren, 2022; Parnealsivam & Khalid, 2024).

The use of an LMS does not in itself demonstrate pedagogically meaningful technology integration. From a TPACK perspective, effective integration requires lecturers to connect technological knowledge with pedagogical and disciplinary knowledge when selecting and applying digital tools to teaching, learning and assessment (Mishra & Koehler, 2006; Mujallid, 2021).

In a South African study, Matarirano et al. (2021) found that Blackboard adoption was interalia influenced by behavioural intention, perceived usefulness, self-efficacy and other individual factors, while training and technical support were important institutional enablers. Similarly, Almogren (2022) found that continued Blackboard use was associated with perceived usefulness, ease of use and satisfaction. Lecturers may nevertheless experience technical difficulties, resistance to change and uncertainty about platform functionality despite receiving training (Parnealsivam & Khalid, 2024).

Meaningful LMS adoption therefore extends beyond access to technology or initial acceptance. Professional development, opportunities for practical application and continuing institutional support are important conditions for lecturers to use LMS tools purposefully within their teaching contexts (Huang et al., 2024; Fütterer et al., 2024; Parnealsivam & Khalid, 2024).

Professional development is therefore central to LMS adoption, but should extend beyond once-off technical training. Technology-focused professional development can strengthen knowledge, skills, collaboration and active learning (Huang et al., 2024), while technology integration develops over time and is shaped by contextual factors (Fütterer et al., 2024). The TPACK framework provides a useful lens for understanding this progression,

emphasising that effective technology integration requires the interaction of technological, pedagogical and content knowledge rather than technological proficiency alone (Mishra & Koehler, 2006; Mujallid, 2021).

Blackboard adoption should not be considered only in terms of whether lecturers report using the LMS, but also in relation to the purposes for which its tools are used. Administrative activities such as uploading notes, announcements and marks indicate platform use, whereas interactive learning activities, formative assessment, feedback, collaboration and purposeful alignment between technology, pedagogy and disciplinary content provide stronger indications of pedagogically oriented use (Mishra & Koehler, 2006; Mujallid, 2021).

At the university examined in this study, Blackboard training formed part of an institutional capacity-development strategy supporting the transition to Blackboard Learn Ultra. The study investigated lecturers' perceptions of the training, their self-reported implementation of Blackboard tools and practices, the challenges they experienced, and the support they considered necessary for continued implementation. It was guided by the question: What are lecturers' perceptions of Blackboard Learn Ultra training and their self-reported implementation of Blackboard tools and practices, and what conditions do they identify as supportive of continued and pedagogically purposeful LMS use?

2. Literature Review and Conceptual Framework

2.1 From LMS Acceptance to Sustained Adoption

Technology-acceptance research helps explain lecturers' LMS adoption. Matarirano et al. (2021) found that Blackboard use was influenced by behavioural intention, perceived usefulness, subjective norm, self-efficacy, enjoyment, anxiety and job relevance, demonstrating that adoption is shaped by cognitive, affective and contextual factors. Almogren (2022) extends this perspective by showing that continued Blackboard use is associated with actual use, satisfaction, perceived usefulness and ease of use. Adoption should therefore extend beyond initial acceptance or compliance to sustained and educationally meaningful use.

However, training alone does not guarantee fuller utilisation. Parnealsivam and Khalid (2024) found that technical difficulties and incomplete understanding of Blackboard continued to constrain lecturers despite prior training, highlighting the importance of ongoing professional development, management support and context-sensitive implementation. Similarly, Mella-Norambuena et al. (2025) found that LMS adoption research remains dominated by TAM and frequently focuses on behavioural intention rather than actual use. This distinction is important for the present study, which examines lecturers' self-reported implementation following training while recognising that self-report evidence cannot establish the quality, frequency or sustainability of actual LMS use.

2.2 Professional Development for Technology Integration

Professional development can build lecturers' confidence and technical skills, but meaningful integration also requires opportunities to practise, contextualise and extend learning. Huang et al. (2024) highlight the importance of active learning, collaboration, expert support, feedback and reflection, while Fütterer et al. (2024) show that technology integration develops during professional development and is shaped by contextual conditions. Introductory LMS training is therefore important but may be insufficient on its own; lecturers also require support for pedagogically purposeful applications such as assessment, student engagement and course design (Parnealsivam & Khalid, 2024). This is particularly relevant in African higher education, where LMS integration is influenced by accessibility, usability and broader implementation challenges (Maluleke & Maake, 2025). South African evidence further highlights timely support, clear communication and realistic consideration of lecturer workload as important conditions for successful LMS implementation (Biccard & Makgato-Khunou, 2024).

2.3 TPACK as the Interpretive Framework

TPACK comprises TK, PK and CK and their intersections: technological pedagogical knowledge (TPK), technological content knowledge (TCK), pedagogical content knowledge (PCK) and integrated TPACK (Mishra & Koehler, 2006). The framework exposes a limitation of purely technical LMS training. A lecturer may know how to create a test, configure the Gradebook or open a discussion forum without knowing when that tool is pedagogically appropriate, how it should align with outcomes, or how it can support the knowledge practices of a particular discipline.

Mujallid (2021) argues that higher education professional development should integrate technological competence with pedagogical knowledge rather than treating online teaching as the transfer of conventional practices to a digital platform. In the present study, TPACK is used as an interpretive rather than a psychometric framework. Navigation, tool operation and configuration primarily indicate TK; alignment of Blackboard activities with learning outcomes and assessment indicates TPK; and discipline-specific application indicates movement towards an integrated TPACK. The framework therefore helps distinguish platform proficiency from pedagogically purposeful adoption.

3. Methodology

3.1 Design and Participants

The study was conducted within the context of a multi-year institutional Blackboard capacity-development initiative. However, the empirical study used a post-training, cross-sectional survey design, capturing lecturers' perceptions and self-reported Blackboard implementation at a single point in time. The sampling frame comprised 200 lecturers who had completed one of the university's Blackboard-related training initiatives between 2022 and 2025, of whom 51 participated (25.5% response rate). The design was intended to describe reported experiences and implementation following training rather than to measure change over time or establish a causal effect of training on Blackboard adoption. Women constituted 62.7% of the sample and men 37.3%. Participants represented six faculties: Social Sciences and Humanities (27.5%), Science and Agriculture (19.6%), Management and Commerce (17.6%), Health Sciences (17.6%), Education (11.8%) and Law (5.9%). Of those who reported campus, 60.4% were based at the East London Campus and 39.6% at the Alice Campus.

Participants had completed either the self-paced and facilitated Digital Teaching and Learning Series (DTLS) skills programme (51.0%) or in person Blackboard Learn Ultra training (49.0%). Prior Blackboard experience ranged from none (19.6%) and basic (41.2%) to intermediate (33.3%) and advanced (5.9%).

3.2 Instrument, Ethics and Analysis

Data were collected through an online questionnaire comprising demographic and training-background items, implementation questions, 19 five-point Likert-scale statements and three open-ended questions. The Likert items examined training relevance and clarity, hands-on opportunities, confidence, Blackboard usefulness, pedagogical integration, assessment alignment, student engagement, tool selection, content presentation and post-training support. The open-ended questions asked participants to identify the most valuable aspects of training, areas requiring improvement and recommendations for future initiatives. Participation was voluntary, and respondents were informed of the study purpose and confidential treatment of their responses.

Quantitative data were analysed using descriptive statistics. An exploratory principal component analysis (PCA) with Varimax rotation was conducted on the 19 Likert items. Sampling adequacy was assessed using the Kaiser-Meyer-Olkin statistic and Bartlett's test of sphericity, and internal consistency was assessed using Cronbach's alpha. Given the modest sample size (n = 51) relative to the 19 items, the PCA is treated as exploratory rather than as a definitive validation of the factor structure. Table 1 reports the available component loadings, eigenvalues, variance explained, factor means, standard deviations and reliability coefficients. Pearson correlations were examined among the extracted components, and exploratory group comparisons were conducted by training type; the available descriptive and significance statistics are reported in Table 2. These analyses are interpreted cautiously and are not used to establish comparative effectiveness or causal effects of either training approach. The open-ended questionnaire responses were analysed inductively using thematic analysis. Responses were read repeatedly to identify recurring ideas, which were coded and grouped into broader categories and higher-order themes. Themes were reviewed against the original responses to ensure that they remained grounded in participants' accounts. As individual responses could reflect more than one theme, qualitative frequencies represent coding occurrences rather than mutually exclusive participant totals. TPACK was subsequently used as an interpretive framework rather than to predetermine the themes.

To supplement the survey-based evidence with institutional behavioural context, annual Blackboard Course Tool Usage reports for 2022-2026 were examined descriptively. These aggregate reports record the number of modules in which selected Blackboard tools were used. The analytics were not linked to individual survey respondents, training type, frequency of use, or quality of pedagogical implementation and were therefore used as contextual evidence of institutional utilisation rather than as an outcome measure of training effectiveness.

4. Results

4.1 Implementation and Perceptions

Blackboard implementation following training was high. Forty-three participants (84.3%) reported implementing tools or strategies learned during training, six (11.8%) reported partial implementation, and two (3.9%) reported no implementation. Regarding overall integration, 54.9% reported moderate use and 33.3% extensive use, while 11.8% reported limited integration. In addition, 93.9% indicated that they would recommend the training to colleagues.

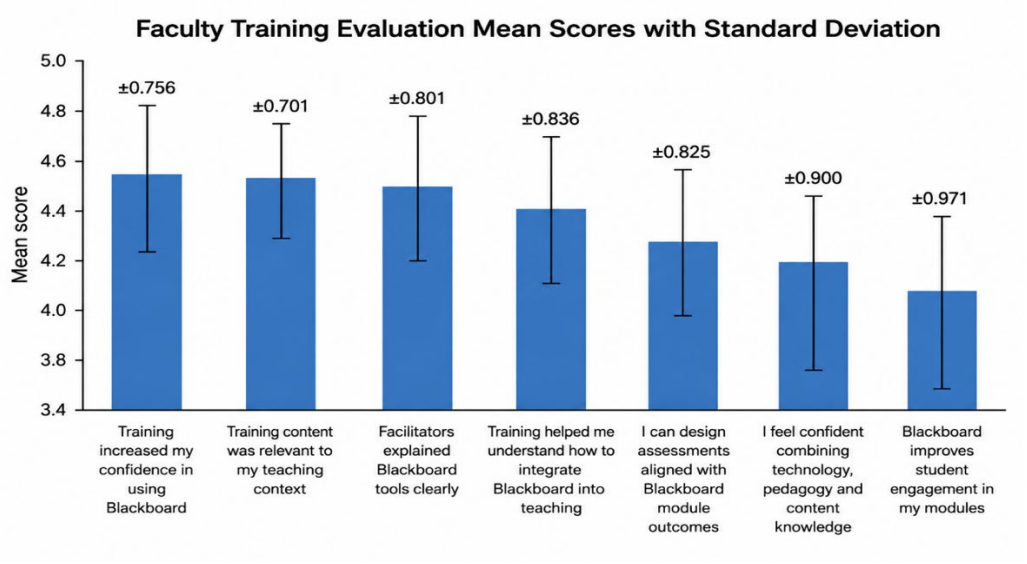


Figure 1: Selected descriptive results for the 19-item questionnaire.

The strongest ratings concerned confidence, contextual relevance and facilitator clarity. The comparatively lower ratings for student engagement and confidence in combining technology, pedagogy and content are analytically important: lecturers were positive about the training, but perceived pedagogical impact was less pronounced than training confidence and relevance.

4.2 Institutional Blackboard tool utilisation, 2022-2026

Institutional Course Tool Usage reports provide additional behavioural context for the self-reported implementation findings. Recorded utilisation expanded markedly across the annual reports. Folder use increased from 4 recorded modules in 2022 to 1,767 in 2026; File from 3 to 1,218; Assessment from 1 to 900; and Web Link from 1 to 231. Learning Module use was first recorded in the comparable data in 2024 (4 modules) and reached 272 in 2026, while Discussion increased from 2 modules in 2023 to 257 in 2026. Survey (47 modules) and Journal (14 modules) were reported in the 2026 dataset. Table 3 and Figure 2 summarise these patterns.

Table 1: Recorded Blackboard tool utilisation by tool, 2022-2026

Tool	2022	2023	2024	2025	2026
Folder	4	6	15	322	1,767
File	3	6	12	261	1,218
Assessment	1	6	15	228	900
Learning Module	0	0	4	45	272
Discussion	0	2	6	53	257
Web Link	1	1	4	62	231
Survey	—	—	—	—	47
Journal	—	—	—	—	14

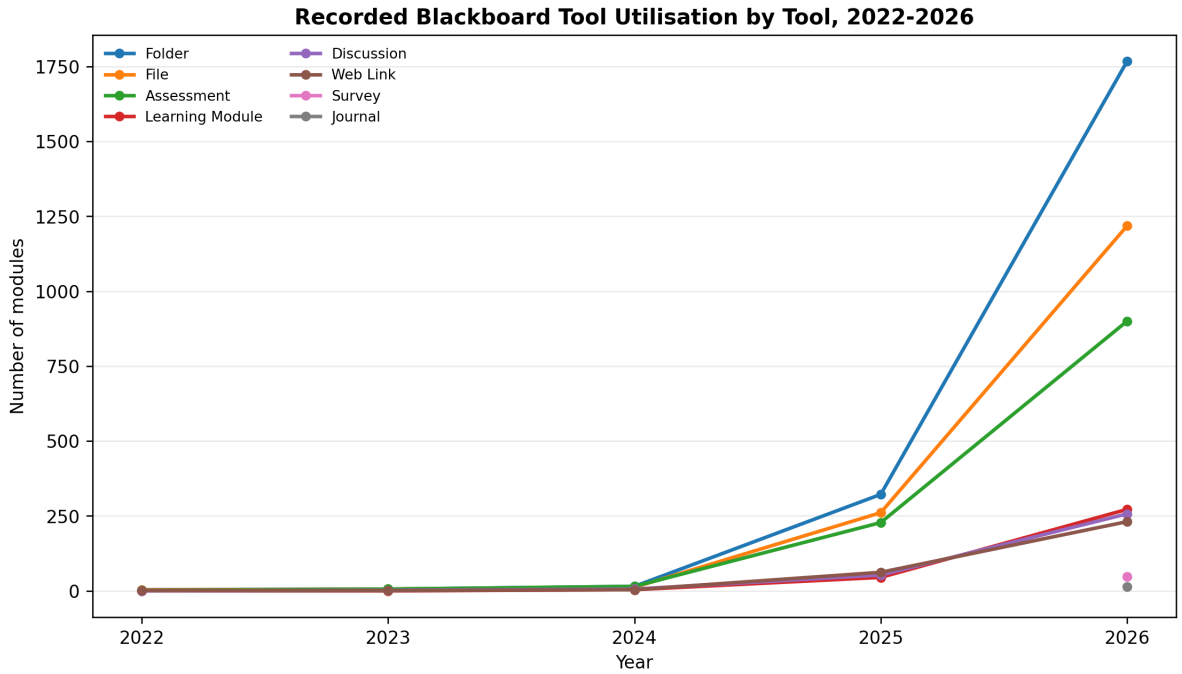


Figure 2: Recorded Blackboard tool utilisation by tool, 2022-2026

Note: Figures represent the number of modules in which each Blackboard tool was recorded as used. Survey and Journal were not reported in the comparable 2022-2025 datasets. Differences in the scope of the annual datasets should be considered when interpreting change over time.

4.3 Factor structure and group differences

The exploratory PCA produced a KMO value of .894 and a significant Bartlett's test of sphericity, $\chi^2(171) = 915.93$, $p < .001$. A three-component solution was examined and accounted for 75.28% of the total variance. Factor 1, Blackboard Training Skills Development, included items related to teaching effectiveness, student engagement, active learning, monitoring progress, assessment innovation, tool selection, content presentation and support. Factor 2, Blackboard Training Experience, reflected clarity, pacing, facilitation, practical opportunity and confidence. Factor 3, Blackboard Training Retention and Transfer, represented contextual relevance, integration into teaching, alignment with outcomes and assessment, and confidence combining technology, pedagogy and content. The detailed component evidence is presented in Table 1. Because the sample is small relative to the number of items, this structure should be regarded as exploratory and interpreted with caution rather than as a confirmed measurement model.

The three exploratory components demonstrated high internal consistency (α = .945, .909 and .919 respectively). They were strongly and positively correlated: Skills Development with Training Experience ($r = .680$, $p < .001$), Skills Development with Retention and Transfer ($r = .797$, $p < .001$), and Training Experience with Retention and Transfer ($r = .799$, $p < .001$). The strength of these correlations also suggests conceptual overlap among the components and reinforces the need for cautious interpretation.

Table 2: Exploratory PCA evidence for the 19 Likert-scale items

Item	Factor loading		
	1	2	3
10. Blackboard improves student engagement in my modules.	.830		
11. Blackboard supports active and collaborative learning.	.796		
12. Blackboard helps me monitor student learning progress more effectively.	.766		
9. Blackboard makes my teaching more effective.	.706		
15. I can use Blackboard to present course content in engaging ways.	.696		
16. I can integrate Blackboard activities that support student participation and interaction.	.686		
19. I feel adequately supported after completing the training.	.681		
14. I can select Blackboard tools that support my teaching strategies effectively.	.679		
13. Blackboard enables me to implement innovative assessment practices.	.619		
2. The training sessions were easy to follow and clear.		.819	
4. The pace of the training supported my learning needs.		.777	
3. The facilitators explained Blackboard tools clearly.		.744	
5. The training provided sufficient opportunities for hands-on practice.		.727	
6. The training increased my confidence in using Blackboard.		.625	
8. The training helped me align online teaching activities with learning outcomes and assessments.			.770
17. I can design assessments using Blackboard that align with module outcomes.			.764
18. I feel confident combining technology, pedagogy, and content knowledge in my teaching practice.			.667
7. The training helped me understand how to integrate Blackboard into my teaching practice.			.637
1. The training content was relevant to my teaching context.			.537
Eigen Value	11.72	1.73	.844
Variance Explained	30.6	22.8	21.9
Factor Mean	3.99	4.28	4.18
Factor Standard Deviation	.744	.691	.732
Reliability	.945	.909	.919

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Note: Extraction method: Principal Component Analysis; rotation: Varimax with Kaiser Normalization; rotation converged in 7 iterations. Blank cells reflect loadings not displayed in the supplied rotated component matrix. Given n = 51 for 19 items, the solution is treated as exploratory.

Participants had completed either DTLS training or Blackboard Learn Ultra training. As shown in Table 2, the exploratory comparison indicated a higher Skills Development score for DTLS participants (M = 37.92, SD = 4.11) than for Blackboard Learn Ultra participants (M = 33.92, SD = 8.20), p = .031. No statistically significant training-type differences were observed for Training Experience (p = .305) or Retention and Transfer (p = .147). Because participants were not randomly assigned to training approaches and no pre-training baseline or behavioural LMS measure was available, these results should not be interpreted as evidence that one training approach was more effective than the other. They are reported as exploratory group differences only.

Table 2: Exploratory training-type group comparisons for the three PCA-derived factors

	Training Course	Mean	Std. Deviation	Std. Error Mean	Sign
Factor 1: Blackboard Training Skills Development	DTLS Training	37.92	4.11	.807	.031
	BB Learn Training	33.92	8.20	1.64	
Factor 2: Blackboard Training Experience	DTLS Training	21.92	2.46	.483	.305
	BB Learn Training	20.92	4.24	.848	
Factor 3: Blackboard Training Retention and Transfer	DTLS Training	21.65	2.54	.499	.147
	BB Learn Training	20.16	4.46	.893	

Note: Group comparisons are exploratory. Participants were not randomly assigned to training approaches, no pre-training baseline was available, and the results do not establish comparative effectiveness or causality.

4.4 Qualitative Findings

Inductive analysis of the open-ended responses produced five higher-order themes. These themes represent interrelated dimensions of lecturers' reported experiences of Blackboard training and implementation rather than sequential stages of adoption. They were derived from participants' responses rather than imposed from TPACK; TPACK was subsequently used to interpret their pedagogical significance.



Figure 3: Themes identified in the findings

Firstly, participants valued learning by doing. Assessment creation, question development, rubrics, Gradebook, sandbox activities and Capstone tasks were repeatedly identified as valuable. Secondly, assessment was simultaneously a strength and a continuing development need: lecturers valued assessment training but requested more time for Gradebook navigation, overall grade calculations and marks management. Thirdly, participants consistently argued that training should not be a once-off event. They recommended annual

refreshers, follow-up workshops, advanced sessions, short instructional videos, manuals and responsive technical support. And fourthly, participants increasingly framed their needs in pedagogical rather than technical terms. They valued discussions, quizzes, polls, live assessments, conditional release and multimedia for student participation. Still, they requested more discipline-specific examples, stronger alignment with learning outcomes and greater attention to integrating technology with pedagogy. Fifth, institutional conditions affected transfer. Blackboard - ITS integration, system stability, workload, protected time, and accessible faculty-level support emerged as practical conditions for sustained use.

5. Discussion

The findings indicate a positive relationship between lecturers' reported training experiences and their self-reported Blackboard implementation, but the study design does not establish that training caused the reported implementation. The high implementation rate shows that most respondents reported applying tools or strategies covered during training, while the exploratory correlations indicate associations among Training Experience, Skills Development, and Retention and Transfer. The qualitative findings further identify practical, pedagogical and institutional conditions that participants perceived as important for continued implementation.

5.1 Adoption, Continued Use and Institutional Support

The findings are consistent with Matarirano et al. (2021), who showed that Blackboard adoption among South African lecturers is shaped by perceived usefulness, ease of use, self-efficacy and related acceptance factors. In the present study, lecturers rated Blackboard's contribution to teaching effectiveness positively, while implementation was high. Yet adoption was not frictionless. Requests for continuing technical support, improved system integration and opportunities to revisit learning indicate that competence alone does not determine sustained use.

The institutional Course Tool Usage reports (See Fig 2) add an important behavioural layer to these perceptions. Across the annual reports, recorded use broadened from predominantly basic content-management functions towards substantially greater use of Assessment, Learning Module and Discussion tools by 2026. This pattern is consistent with wider institutional uptake and diversification of Blackboard functionality. It does not, however, demonstrate that the training programmes caused the increase, because the annual analytics differ in scope and cannot be linked to the individual lecturers who completed the survey or to a specific training route.

This pattern also aligns with Almogren (2022), whose study highlights the importance of perceived usefulness, ease of use, satisfaction and actual use in lecturers' intention to continue using Blackboard. In the present study, participants' responses suggest that opportunities to apply training to authentic teaching tasks and access support for unfamiliar or complex functions are valued features of professional development. These findings should be interpreted as lecturers' perceptions rather than evidence that such features caused continued LMS use.

Parnealsivam and Khalid (2024) similarly found that extensive training does not necessarily lead to full exploitation of Blackboard functionality when lecturers experience technical challenges, incomplete understanding or resistance to change. Their recommendation for ongoing training and management support closely mirrors the current participants' calls for refresher development, just-in-time resources and accessible post-training assistance. The convergence across studies strengthens the argument that LMS adoption should be treated as an institutional capability-development process rather than an individual compliance event.

5.2 TPACK: From Technical Competence to Pedagogical Integration

TPACK provides a useful interpretive lens for the difference between confidence in using Blackboard and confidence in combining technology, pedagogy and content. Participants reported confidence in platform-related activities such as navigation, content creation, assessment configuration and interactive tools. However, the lower mean for integrated technology-pedagogy-content confidence (M = 3.90), together with requests for discipline-specific examples, suggests that respondents perceived pedagogical integration as more challenging than technical platform use. Because TPACK was not measured longitudinally, these findings should not be interpreted as evidence of progression from TK to integrated TPACK over time.

The growth in recorded Assessment, Learning Module and Discussion use is also relevant to the TPACK interpretation. These tools can support assessment, structured learning and interaction beyond basic content distribution. Their presence in a module does not by itself demonstrate integrated TPACK, but the diversification

of tool use provides institutional context for participants' reports that they were applying Blackboard to assessment, participation and course organisation.

This finding supports Mujallid's (2021) argument that professional development for online teaching should integrate technological and pedagogical knowledge. It also explains why participants wanted training built around authentic academic scenarios rather than additional lists of Blackboard functions. For example, redesigning a discipline-specific assessment for formative feedback, using conditional release to scaffold preparation, or constructing a discussion that requires disciplinary reasoning would require lecturers to connect TK with PK and CK.

The implication is significant: the objective of professional development should not be maximal feature exposure. It should be the intentional and purposeful selection and application of LMS affordances to teaching problems. In this sense, adoption becomes pedagogical when lecturers can justify why a tool is appropriate, align it with learning outcomes and assessment, and adapt it to the knowledge practices of their discipline.

5.3 Professional Development as a Continuing Process

Participants' strongest recommendations concerned time, practice and continuity. This is consistent with Huang et al. (2024), who identify active learning, collaboration, expert support, feedback and reflection as important features of technology-enabled professional development. It is also compatible with Fütterer et al. (2024), who show that technology integration changes during professional development and is sensitive to contextual conditions.

The exploratory Skills Development difference between DTLS and Blackboard Learn Ultra participants should not be interpreted as evidence that one programme was more effective. Participants were not randomly assigned, there was no pre-training baseline, and the groups may have differed in prior experience, duration, pedagogical emphasis or opportunities for practice. The result is therefore best treated as a signal for future investigation. Future comparative evaluation could examine whether different combinations of self-paced learning, facilitated guidance, hands-on practice and discipline-specific application are associated with different outcomes using larger samples, baseline measures and behavioural LMS evidence.

5.4 Integrating the Five Research Themes

The five themes should be understood as interrelated dimensions of lecturers' experiences rather than as sequential phases of adoption. Together, they highlight practical digital competence, assessment and administrative capability, continuing professional development, pedagogical integration, and an enabling institutional environment. From a TPACK perspective, the findings suggest that Blackboard professional development should extend beyond technical proficiency by creating opportunities for lecturers to apply technology purposefully within teaching, assessment and disciplinary contexts. Given the cross-sectional design, these dimensions are not presented as a developmental pathway or as evidence that lecturers progress through predetermined stages over time.

6. Conclusion and Recommendations

This study found that Blackboard Ultra training was positively perceived and accompanied by substantial self-reported implementation among respondents. The quantitative findings were complemented by five themes: developing practical digital competence; enhancing assessment and administrative capability; sustaining learning through continuous professional development; moving from technical training to pedagogical integration; and creating an enabling institutional environment. These themes represent interrelated dimensions of lecturers' reported experiences rather than stages through which lecturers were shown to progress.

Through a TPACK lens, the findings point to the value of professional development that connects technological competence with pedagogy and disciplinary content. Future development should retain hands-on sandbox activities; differentiate introductory, intermediate and advanced opportunities; provide refresher and just-in-time resources; organise workshops around authentic disciplinary teaching and assessment problems; provide time for practice and reflection; and strengthen Gradebook, Blackboard-ITS and faculty-level support. Future evaluation should assess application using evidence beyond training attendance and self-report.

The study is limited by its single-institution, post-training cross-sectional design, a response of 51 lecturers from a sampling frame of 200, and the absence of a pre-training baseline or comparison condition. Although aggregate institutional Course Tool Usage reports for 2022-2026 provide behavioural context and show substantial growth and diversification in recorded Blackboard tool utilisation, these analytics cannot be linked to individual survey respondents, training routes, frequency of use or pedagogical quality. Reported implementation therefore cannot be assumed to represent the depth or sustainability of actual course use, and neither the survey nor the aggregate analytics establish that training caused the observed institutional pattern. The exploratory PCA is also limited by the modest sample size relative to the 19 items, and its factor structure and group comparisons should therefore be interpreted cautiously. Future research should use larger samples and combine baseline and follow-up measurement with participant-linked LMS analytics, course-design review and student evidence. Within these limitations, the findings identify lecturer-perceived characteristics and institutional conditions that may inform future Blackboard professional development.

Ethics Declaration

Ethical approval was obtained from the University of Fort Hare Inter-Faculties Research Ethics Committee (IFHREC-26/05-0264). The anonymous SurveyMonkey questionnaire was distributed electronically to lecturers who had completed Blackboard training between 2021 and 2025, with multiple reminders sent. An information sheet and consent form outlined the study purpose, voluntary participation, confidentiality, anonymity and the right to withdraw. Facilitators and supervisors could not identify participants, and findings were reported only in aggregate. Anonymised data were archived electronically at the University for research purposes.

AI Declaration

AI tools, including Elicit and ChatGPT, were used to assist with literature searching, language refinement and editorial support. Nano Banana was used in the creation or modification of visual material. The authors reviewed and edited all AI-assisted outputs and take responsibility for the manuscript. References, analyses and factual claims were checked by the researchers against the underlying sources and study data

Acknowledgement

The authors would like to acknowledge Prof Rudolph Leon van Niekerk from the University of Fort Hare for his valuable assistance with the statistical analysis and interpretation of the quantitative and qualitative data.

References

Almogren, A. S. (2022) Art education lecturers' intention to continue using the Blackboard during and after the COVID-19 pandemic: An empirical investigation into the UTAUT and TAM model. Frontiers in Psychology, 13, 944335. https://doi.org/10.3389/fpsyg.2022.944335

Biccard, P., & Makgato-Khunou, P. (2024) Lecturers' perceptions of a learning management system migration at an open distance learning institution. International Journal of E-Learning & Distance Education/Revue internationale du e-learning et la formation à distance, 38(2). https://doi.org/10.55667/ijede.2024.v38.i2.1300

Fütterer, T., Backfisch, I., & Lachner, A. (2024) Teachers' trajectories of technology integration during participation in an online professional development program. Zeitschrift für Erziehungswissenschaft, 27, 769–801. https://doi.org/10.1007/s11618-024-01251-6

Huang, L., Liang, M., Xiong, Y., Wu, X., & Lim, C. P. (2024) A systematic review of technology-enabled teacher professional development during COVID-19 pandemic. Computers & Education, 223, 105168. https://doi.org/10.1016/j.compedu.2024.105168

Matarirano, O., Jere, N. R., Sibanda, H. S., & Panicker, M. (2021) Antecedents of Blackboard adoption by lecturers at a South African higher education institution—Extending GETAMEL. International Journal of Emerging Technologies in Learning, 16(1), 60–79. https://doi.org/10.3991/ijet.v16i01.16821

Mishra, P., & Koehler, M. J. (2006) Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mujallid, A. (2021) Instructors' readiness to teach online: A review of TPACK standards in online professional development programmes in higher education. International Journal of Learning, Teaching and Educational Research, 20(7), 135–150. https://doi.org/10.26803/ijlter.20.7.8

Parnealsivam, J. A/P, & Khalid, F. B. (2024) Benefits and challenges of using Blackboard® with strategies to overcome them. International Journal of Academic Research in Business and Social Sciences, 14(8), 819–838. https://doi.org/10.6007/IJARBSS/v14-i8/22450

Mella-Norambuena, J., Chiappe, A., & Badilla-Quintana, M. G. (2025) Theoretical and empirical models underlying the teaching use of LMS platforms in higher education: A systematic review. Journal of Computers in Education, 12, 997–1023. https://doi.org/10.1007/s40692-024-00336-9

Maluleke, A. F., & Maake, G. (2025) Examining the impact of learning management systems in African higher education: A systematic review. African Journal of Inter/Multidisciplinary Studies, 7(1). https://doi.org/10.51415/ajims.v7i1.1550

Factors of non-Fiction Texts Comprehension Among Primary School Pupils in the Czech Republic

Dana Vicherková, Josef Malach, Kateřina Malachová, Martin Kolář and Veronika Murinová
University of Ostrava, Ostrava, Czech Republic

dana.vicherkova@osu.cz
josef.malach@osu.cz
katerina.malachova@osu.cz
L20487@student.osu.cz
D23704@osu.cz

Abstract: In the context of Society 4.0, the ability to comprehend electronic forms of everyday non-fiction texts is increasingly important. International PISA research (2009, 2018) has repeatedly identified lower levels of reading literacy among Czech fifteen-year-old boys. The demand for comprehending both formal and informal texts with professional content, such as technical, economic, and media texts, continues to rise. Online media risks, particularly those associated with social networking, constitute considerable threats to primary and secondary school pupils in both educational and family environments. Reading comprehension should also support cognitively and affectively balanced communication. This research examines the roles and models of reading by primary school pupils in the Moravian-Silesian region of the Czech Republic. The qualitative study, using semi-structured interviews with 14 pupils, aimed to identify the types of electronic non-fiction texts preferred by pupils, the e-texts included in recommended secondary school reading lists, the instructional methods employed to develop digital, informational, and reading competencies, and the factors influencing pupils' comprehension of non-fiction texts. The results show that Czech fifteen-year-old pupils frequently find it difficult to interpret visually presented information and tend to prefer texts written in informal abbreviations. Media education is frequently not integrated as a distinct component of the primary school curriculum. This study forms a component of a broader research initiative under the Just Transformation Operational Programme, specifically the REFRESH project (Research Excellence for Region Sustainability and High-tech Industries), registration number CZ10.03.01/00/22_003/0000048.

Keywords: Reading and Digital Literacy, Communication, Text Comprehension, Everyday, non-Fiction Text, Primary/Secondary School Pupils

1. Introduction

In the Czech and broader European context, Society 4.0 in the 2020s is characterised by an information-based, digital era shaped by accelerated lifestyles and work patterns.

In both educational settings and daily life, there is a clear demand for media safety in digital environments, encompassing both formal and informal communication. The processes of digitalisation, automation, and robotisation, driven by advances in artificial intelligence, have both positive and negative effects on everyday life across generations and contribute to growing educational needs. Simultaneously, the need for adaptability is increasing as the labour market rapidly evolves and certain professions in traditional industrial sectors disappear. “A person is influenceable by the social and cultural environment to which they belong.” (Nakonečný, 2014, s. 137).

In the Czech Republic, educational practices in primary and secondary schools increasingly emphasise engagement with traditional artistic texts. Media education for these pupils is frequently undervalued, with greater attention given to informal communication activities.

The Moravian-Silesian region remains shaped by a long industrial tradition, especially hard coal mining and mechanical engineering. Employees of industrial and mining companies and their children form a social community marked by demanding physical labour, psychological strain and specific attitudes, values, lifestyles, educational needs and career adaptability. Janků (2014) recorded interviews with OKD miners employed between 1955 and 1999; these recordings were later published by the Mining Museum.

Mining education in the Moravian region developed from the late 19th century through mining schools and, after 1948, vocational and industrial mining schools. Today, OKD/DIAMO is transforming into a technologically advanced enterprise focused on automation, digitalisation of internal strategies and key data systems, and competitiveness in foreign markets. Digitalisation, robotisation and artificial intelligence therefore also influence educational trends.

The transformation of the Moravian-Silesian coal region brings new visions of regional identity, including restoration of natural environments, environmental ecology and readiness for change. This shift supports environmentally responsible behaviour, positive attitudes toward alternative energy sources and changes in human capital.

In the global, European and Czech regional context, pupils increasingly need functional literacy, including reading, mathematical, financial and environmental literacy, especially when working with different types of texts and environmental attitudes, values, beliefs and behaviour. According to Kroufek and Činčera (2021), environmental literacy includes pupils' environmental knowledge, attitudes, values and beliefs. Czech research has been conducted by Svobodová and Kroufek (2018), while Erdogan and Ok (2011), Spínola (2023), and Johnson and Činčera (2023) contributed to the international "Green School" research program.

The REFRESH research project—Research Excellence for Region Sustainability and High-tech Industries—aims to bridge science and educational practice. Two research teams at the Faculty of Education, University of Ostrava, are engaged in the REFRESH project. One cohort addresses primary education in MSK; the second examines OKD/DIAMO employees' children through quantitative and qualitative research on education, career orientation, functional literacy and media safety.

2. Traditional and Digital Readership as Educational Capital

Traditional paper textbooks (specialised texts) across professional fields remain the primary and most prevalent means of information transfer in Czech primary and secondary educational practice in the 2020s. The harmonisation of cognitive and affective reading comprehension development is central to reading competency within interactions between communicators and communicants across various types of textual communication. Both formal and informal reading underpin the development of individual communication competencies. "Language competences represent the cultural capital of every person and have an impact on educational opportunities, job opportunities, and social status" (Janus-Sitarz and Nowak, 2014, p. 9). The accelerated pace of 21st-century society, in a global context, underscores the necessity of reading diverse texts in both print and electronic formats. The information age demands greater dynamism in working with text, including searching, sorting, evaluating, processing, and understanding information. According to Prázová et al (2014, p. 9), “Reading literacy influences a child’s success in the entire educational process.”

Contemporary readers must possess universal communicative competence and continually build communication capital as a form of educational capital. In the digital age, processes of text administration have evolved to fulfil official communication functions, including managerial, professional, and administrative purposes. There has been a marked increase in the number of language users engaging in administrative texting (Kraus, 1997, p. 92). Elements of communication manipulation, technologization, and marketisation are increasingly present in the digital public text space.

3. Digital Competency of Pupils and Media Safety

In the Czech Republic, binding curricular documents for primary school education are divided into two main categories:

a) "Framework Educational Program" (RVP ZV) at the national level of education
b) "School Educational Program" (ŠVP) at the school level of education

Within the curriculum for primary and secondary school pupils in the Czech Republic, particularly those aged 10 to 18, the objective is to develop social sciences with an emphasis on language education, as well as information and digital strategies. These strategies focus on using technological tools and on understanding safe communication practices in media environments.

4. Pupil Digital Competency and Reading

Reading today is closely linked to the digital environment. As a result, reading activities contribute to the development of both reading and digital competencies.

The risk of vulnerability to disinformation and manipulative content in media and online environments is increasing, particularly among primary school pupils. Schools must therefore counter digital manipulation by developing strategies for searching, processing, comparing and evaluating information. The evaluation of media content is crucial because it concerns the relevance and truthfulness of information in relation to social reality.

Digital and media literacy defence in school practice includes work with formal and informal texts that may contain manipulative or disinformation content. Educational approaches should help pupils identify risk profiles and problematic behaviour in electronic environments, while strengthening critical thinking and the resilience of both individuals and learning communities.

4.1 Reading Roles and Reading Patterns

Literary education creates a stimulating environment for harmonising reader roles and shaping reading patterns when engaging with various types of texts. In the Czech educational context, literary education is grounded in theoretical concepts of teaching literature as a science and its historical dimensions. Currently, emphasis is placed on developing readers from an early (preschool) age, with a focus on the aesthetic and emotional-educational communication functions of texts, considering both content (what the text communicates) and form (how information is conveyed).

Reading manifestations constitute an essential component of functional literacy, influencing both professional success and everyday application. The Czech Institute of Reading highlights the critical state of Czech pupils' reading skills and the level of reading competences among primary and secondary school pupils as "indispensable for the pupils' further study path and their application in real life" (ČŠI, 2019, p. 34).

4.2 Types of Texts in the Educational Reality of Primary Schools, Reading and Literacy

In primary and secondary education, expository texts in print, electronic and digital formats are most common. Höflerová (2011, p. 27) characterises the teaching text as a professional teaching tool: “The teaching text is derived from the professional text in function and style…” Kostečka (2002, p. 184) notes that today “the mixing and interweaving of elements of various functional styles and stylistic procedures is quite common,” reflecting the complexity of the (post)modern world. Expository texts have a long spoken and printed tradition since printing (Gutenberg, 1440). In contemporary schools, electronic texts often combine functional styles. Karlík, Nekula, and Pleskalová (2002, p. 450) define electronic text as “a non-linear way of organising text during computer processing of optical and audio information.” NPI (2024) defines digital competence within the RVP ZV revision as the ability to work with technologies, creativity, digital and formal security, and theoretical and applied knowledge; it includes using technologies safely, confidently, critically and creatively in work, learning, leisure and civic life.

Non-fiction texts encompass several functional styles, including expository, narrative, descriptive, reasoning, reflective, dramatising, polemical, monological, and discussion forms. Educators are not only users of various communication texts but also active transmitters of information in both face-to-face and digital or electronic formats. In the educational context of the 20th and 21st centuries, Internet reading has become one of the most common reading activities, serving diverse purposes. The PIRLS (Progress in International Reading Literacy Study, 2021, p. 12) research highlighted the dominant role of the Internet in the transmission and sharing of electronic data. Reading in the Internet environment “is one of the main ways in which pupils obtain or share information.” (PIRLS, 2021, p. 12).

Reading functions as a lifelong communication and information pathway linking generations in family, work and leisure contexts. In education, it is both a natural and formal act of communication involving cooperation, understanding, listening, text selection, reading purpose and reading experience. “The literacy of the population is considered one of the basic indicators of a nation’s education and a factor in the country’s economic development…”; it also supports social participation and is important in the context of globalisation (Doležalová, 2009, pp. 223-224).

In reading education, work with information sources is changing, and written expression reflects language specificity, language norms, literary activity, opinion formation and critical thinking. Akimov's research (2019) examined understanding of Internet texts and the influence of Internet experience, using semantic and receptive experiments, semantic-pragmatic interpretation, content analysis and subjectively oriented scaling with 716 respondents. More than half underestimated the complexity of Internet texts, and a 15% correlation was confirmed between Internet experience and the quality of understanding. Hník (2016) examined attitudes toward reading and school reading experience, indicating the need to study motivation, strategies, variability and competence in different types of artistic and non-artistic texts.

According to the National Report of the Czech School Inspectorate (NZ ČŠI, 2023, p. 5), reading skills are characterised in educational practice as an essential prerequisite "for the effective acquisition and mastery of many other skills, including, for example, mathematical or scientific skills."

The ability to recognise manipulative text content is influenced by a range of factors, including diagnostic, cognitive, technological, pedagogical, psychological, social, media, communication, and interpersonal elements.

5. Teaching Methods for Developing Critical Thinking, Digital and Reading Skills

Methods that enhance critical thinking skills (RWCT methods) are among the teaching methods that develop digital and reading competencies. One effective approach is the Five-Leaf Method, which uses a five-line graphical structure to clarify the characteristics of a selected concept, such as the word "school." This method provides a clear and engaging way to understand key concepts, resulting in rapid comprehension of the selected term.

Grecmanová and Urbanovská (2007, p. 15) examined methods that foster critical thinking, including the ability to express opinions on selected phenomena, formulate questions, think logically, defend arguments, and develop creative thinking.

Critical thinking methods represent an effective strategy for supporting functional literacy, including reading, digital information literacy, environmental literacy, and mathematical literacy across diverse age groups. Digital literacy is integral to the UNICEF framework, which aims to equip children and adolescents with the skills necessary for success in school, the workforce, and their broader lives. (UNICEF, 2025)

Research by the Czech Institute of Secondary Education (2019, p. 34) identified a gap in Czech primary and secondary education regarding the promotion of independence, critical thinking, and shared responsibility for educational outcomes. This gap is evident in the limited use of diverse non-fiction texts and strategies that facilitate comprehension of various text types. The Czech education system pays insufficient attention to engaging with different types of everyday texts, such as by encouraging pupils to question opinions presented in texts (VZ ČŠI, 2023, p. 97). There is also a notable absence of instruction in reading techniques, including digital strategies and skills such as skimming and scanning.

6. Research Methodology

Researchers from the 2nd REFRESH cohort conducted qualitative research using semi-structured interviews with 15-year-old pupils from the lower-secondary schools and multi-year grammar schools. These pupils come from families whose parents were employed in heavy industry at OKD/DIAMO companies in the 1920s. Currently, OKD is undergoing a transformation into a company focused on coal abatement and the development of new technologies in the Moravian-Silesian Region of the Czech Republic.

The research is based on the literacy-focused report entitled "National PISA Report 2022," published as a planned output of the project "Data-analytical support for the assessment and management of the education system of the Czech Republic." This report builds upon findings in mathematical, reading, and science literacy published by the Czech School Inspectorate (2023). The report identifies a significant issue among pupils from diverse socio-economic backgrounds. The sample of children of OKD/DIAMO employees is considered an at-risk community due to their presumed lower motivation to read and pursue further study at secondary or higher education institutions in the Czech Republic.

The qualitative research aimed to identify the types of non-fiction texts used by children of OKD/DIAMO employees; whether, how often and for what purpose they read electronic or digital non-fiction texts; and how these activities relate to everyday, study-related or occasional reading. It also focused on pupils from families with an active relationship to different types of texts, including artistic/non-artistic, printed/electronic and professional/advertising texts.

Another objective was to determine whether pupils read mainly for compulsory study or according to other preferences, what types of electronic non-fiction texts they prefer, which non-fiction texts appear in school reading lists, what teaching methods support their comprehension, and what factors influence comprehension from their perspective. The study also examined whether media education is taught as a separate subject or within other subjects, and how pupils perceive reading competence, digital technologies and preferred types of technology.

Another aim of the research was to identify selected educational characteristics of children, reading interest, reader profiling, and preference for types of non-fiction texts for everyday life. The research task was also to specify key educational/reading characteristics, as well as the professional orientation, of children of miners and employees of OKD/DIAMO during the period of OKD's transformation and the transformation of society 4.0.

A traditional or non-traditional reader can be understood through interest in artistic or non-artistic texts, preference for print, electronic, digital or hybrid texts, and manifestations of reading and digital literacy. Another criterion is whether texts include an image (illustration, graph, drawing or symbol), informative text, or alternative text such as “Alt text” in an e-commerce environment.

The research method was a semi-structured interview. The answers were transcribed verbatim, analysed through open coding and categorised into profiles of traditional reader and non-traditional reader (digi/profiling). Respondents then reflected on whether they considered themselves traditional or non-traditional readers and formulated characteristics of both profiles in relation to school and home reading environments. Connections were sought between the codes and the profiles of a traditional pupil-reader and a non-traditional pupil-non-reader. Primary schools were characterised by textbook-based instruction and an emphasis on verified information sources for media safety reasons. Pupils expressed interest in a school media-safety regime; at home, media safety is usually ensured by adults, although three pupils stated that parents or legal representatives may be less media literate than the child.

Semi-structured interviews with 14 respondents (fifteen-year-old pupils of lower-secondary schools in the Moravian-Silesian Region) were carried out in person in the classroom of the Faculty of Education of the University of Ostrava in the period from December 11, 2025, to March 3, 2026.

The research survey was intentional. The research sample consisted of 8 boys and 6 girls (14-year-old Czech pupils in 8th grade from two primary schools in Ostrava, in the Moravian-Silesian Region).

7. Description and Analysis of Semi-Structured Interviews with Selected Primary School Pupils

Research sample:

- Primary school pupil (14-15 years old) – traditional primary school reader
- Primary school pupil (14-15 years old) – non-traditional primary school reader

7.1 Code Categorisation

During the interviews with selected primary school pupils, categorised codes were created:

- 1A Category of characteristics of a primary school pupil (traditional reader aged 14-15)
- 2A Category of characteristics of a primary school pupil (non-traditional reader aged 14-15)
- K – code designation of characteristics of traditional and non-traditional readers

7.2 Codes Focusing on the Characteristics of the Traditional Reader

K1: regularity of reading due to interest in school reading, K2: priority of reading printed books, K3: priority of reading environment, K4: library visitor (city, public library), K5: number of (printed) books read in 1 year, K6: number of e-books read/in 1 year, K7: reading for pleasure, K8: reading for duty, K9: reading information with pictures/illustrations, K10: reading audio text, K11: priority of media education, K12: understanding information in visual form, K13: motivation to read poetry, K14: motivation to read prose, K15: relationship to sci-fi, K16: priority of texts written informally, K17: reading combined texts, K18: interest in fantasy, K19: interest in comics, K20: interest in fairy tales, K21: interest in reading Czech literature, K22: interest in reading foreign literature, K23: interest in poetry/poems, K24: interest in music, folk, K25: authorship of poems, I am the author of poems, K26: drawing while reading, K27: factors motivating reading – most often, K28: interest in digital text, K29: interest in creativity, K30: popularity of oral folk literature, K31: interest in the environment, K32: interest in media safety, K33: taking notes while learning from a textbook, K34: preference for writing information on a notebook or PC at school.

7.3 Descriptive Analysis of the Traditional Reader (1A)

1A Characteristics of a primary school pupil (traditional reader aged 14-15):

Codes focusing on the characteristics of the traditional reader (samples of respondents' answers)

K1: regularity of reading due to interest in school reading (“only when I have to”, “for study obligations”), K2: priority of printed books (“I prefer paper”, “I like the rustling of pages”), K3: preferred reading environment (“home”, “house”, “apartment”), K4: library visitor (“yes”, “of course”), K5: number of printed books read in 1 year (“4–5”, “rather 5”), K6: number of e-books read in 1 year (“1–3”), K7: reading for pleasure (“yes”, “mostly yes”, “sometimes”), K8: reading for duty (“yes, I have to”), K9: reading information with pictures/illustrations (“yes”, “priority”), K10: reading audio text (“no”, “not often”, “sometimes”, “I miss the image of the text”), K11: media education (“mandatory”, “useful for orientation in media text”), K12: understanding information in visual form (“probably yes”, “not very much”, “I don’t understand images”), K13: motivation to read poetry (“no”, “rather not”, “waste of time”), K14: motivation to read prose (“yes”, “depending on the theme and plot”), K15: relationship to sci-fi (“rather don’t know”, “no”, “depending on the mood”), K16: informal texts (“no”, “I don’t mind”, “rather not”), K17: combined texts (“rather not”), K18: fantasy (“rarely”, “little”), K19: comics (“yes”, “enthusiastically”), K20: fairy tales (“yes”, “always”, “only some”), K21: Czech literature (“yes”, “probably yes”, “not much”), K22: foreign literature (“sometimes”, “I prefer Czech literature”), K23: poetry/poems (“no”, “rather not”), K24: music/folk (“no”, “sometimes”, “only modern”).

Table 1: Preferences of a traditional reader

Method of reading for pleasure	**Printed**	**Fairy Tales**	**Pictures**	**Sci-Fi**
	10	9	8	4
Method of reading for learning (school reading)	**Digital Text**	**Combination of Digital/Paper**	**Paper**	
	11	12	8	
Favorite form of text for reading out of interest	**Paper**	**Digital**		
	10	4		

The traditional Czech reader (tab 1.) most often reads school reading as a school obligation, not out of interest. He prefers printed text and fairy tales to read. He likes text with pictures, comics (not sci-fi), he likes to read at home and in the library, not at school. He prefers digital texts for learning text.

7.4 Traditional Reader (From the Perspective of Primary School Pupils)

K1: Regularity of Reading

- Question: Do you regularly read the recommended reading by school? From Jana's perspective: *"I read school reading only when I have to."* On the other hand, Eva stated that she reads school reading *"for study obligations"*.

K2: Priority of Reading Printed Books

- Question: What type of text do you prefer for reading? From Petr's answer, it is clear that he prefers traditional paper text for reading*: "I prefer paper, a sheet of paper for reading"*, and this is due to the popularity of paper text: ... *"I like the rustling of the pages"*

K3: Priority of Reading Environment

- Question: What environment do you prefer for reading a text? Karel: *“I prefer reading in a familiar, home environment”* (i.e. home, house, apartment). On the other hand, Petr prefers to read in a (school) library/study room environment and is a regular visitor to the school (and city) library.

K5: Number of (Printed) Books Read in 1 year

- Question: How many books do you read in 1 school year (10 months)? Pavel reads 6 books in a year.

Eva answered that the number of books read for pleasure in 1 year is 6 books (of which 4 are literary texts, 2 are non-fiction texts (specialised books with an economic focus). Petr reads 28 books in a year, " I am a bookworm, I read everything that is around me, on the bus, at home, at school, in the morning, at night."

7.5 Descriptive Analysis of the Non-Traditional Reader (2A) media"), K4: Classical Library use ("Rather Not", "Irregular Visitor"), K5: e-books Read in 1 School Year ("2–12", "not)

2A Characteristics of a primary school pupil (non-traditional reader aged 14-15):

Codes focusing on the characteristics of the non-traditional reader (samples of respondents' answers)

K1: irregular school-recommended reading ("mostly irregularly", "not often, rarely"), K2: priority of non-printed books ("digital text is a priority"), K3: flexible reading environment ("anywhere, whenever I read and listen to many"), K6: hybrid texts read in 1 year ("many", "can't count them", "probably a lot"), K7: reading for duty ("sometimes", "when I have to", "out of obligation"), K8: reading for pleasure ("yes, I choose the texts", "I like reading with pictures/illustrations", "no, it's boring"), K9: audio texts ("sound is important", "pleasant", "important for life"), K10: media education as a separate subject ("rather not", "unnecessary"), K11: visual information ("yes", "graphs, tables, geometric shapes"), K12: poetry ("rather yes", "lyrical motifs"), K13: prose ("rather no, depending on focus"), K14: sci-fi ("yes", "technology", "visions of the future"), K15: informal texts ("yes", "we save time"), K16: combined written/audio/video texts ("yes", "variety is necessary"), K17: comics ("yes, I like it"), K18: fairy tales ("yes", especially combinations of paper text and film adaptation), K19: foreign literature ("yes", "I read in several languages"), K20: poems ("yes", "folk poetry", "I like reading texts and singing"), K21: robotisation ("technical process", "helper", "invention", "danger", "secret").

Examples of selected pupil responses to questions in a semi-structured interview:

7.6 Non-Traditional Reader (From the Perspective of Primary School Pupils)

K1: lack of interest in reading the recommended school reading material

- Question: Do you regularly read the recommended reading by the school? From Anna's perspective: *"I don't read the reading recommended by the school."* Jiří stated that he considers school reading unnecessary.

K2: priority of reading types of texts

- Question: What type of text do you prefer for reading? It is clear from Josef's answer that he prefers non-traditional hybrid texts for reading: *"I have a preference for the hybrid form of text - it is more practical, more accessible, more interesting."*

K3: priority of reading environment

- Question: What environment do you prefer for reading a text? Alena: *"I like to read in nature, for example, in the garden, especially in spring, when the birds sing, and nature wakes up."* Erik prefers to read at night with a flashlight in the attic. Both Alena and Erik visit the city library and the *"reading marathon"* activity organised by the school once a year on the weekend.

K5, 6: number of books read in 1 school year

- Question: How many different types of books do you read in 1 school year (10 months)? Pavel: reads 12 e-books per year and hybrid texts every day; he doesn't count them because *"they are short texts for everyday information".*

Felix answered that he reads hybrid and various digital texts created by AI every day, so 90 texts per year. Alena answered: *"I read more than 25 books per year, mainly prose (novels, short stories, fables, novellas).*

8. Discussion and Conclusions

According to the National Report of the Czech School Inspectorate (NZ ČŠI, 2023, p. 5), reading skills are foundational for acquiring and managing other skills, including mathematical and scientific competencies. Karaarslan's research (2023-2024) examined digital reading, motivation, attitude, reading comprehension and vocabulary acquisition among 23 Turkish primary school pupils learning English as a foreign language. Digital reading positively influenced comprehension across text types and increased engagement; after a six-week intervention, the average motivation score rose from 57.78 to 68.17.

The school-leaving canon was analysed by Hník (2016, pp. 234–241). Ramsay and Sperling (2015) investigated factors influencing text comprehension experimentally, while Hlebová and Vojteková (2018) focused on diagnosing and stimulating pupils' reading competence, particularly in the Roma community. The interviews

indicate that children of mining employees primarily read literature recommended by schools, prefer paper texts for pleasure and electronic or digital texts for learning. The research identified traditional and non-traditional reader characteristics and factors influencing profiles: motivation, text type, video/film/audio materials, media education and safety, formal and informal education, digital and reading competencies, media literacy and collaboration with artificial intelligence.

The use of a non-standardised research tool constitutes both a limitation and a potential advantage of this study. A notable advantage is the targeted examination of traditional and non-traditional reading, alongside the investigation of curiosity about everyday life as a means to improve adaptability in the labour market within a rapidly changing technological and digital environment. Additionally, the relationship to reading demonstrates a broader pursuit of knowledge.

Future research in the field of reading should investigate the factors that influence academic success among adult pupils, particularly within the context of preparing future teachers to serve as reading role models.

McGeown et al. (2020) identified differences in reading motivation and preferences for various text types among 28 Scottish children aged 9 to 11 years. Respondents reported reading for reasons such as seeking peace and relaxation, maintaining routine, experiencing joy and cheerfulness, and escaping reality. In addition to printed books.

Malach and Vicherková (2020) investigated the characteristics of traditional and non-traditional students in tertiary education, with a focus on locus of control. Rotter (1966) explored the relationships among locus of control, professional, managerial, and academic success, as well as life satisfaction among students. Fazey and Fazey (2001) emphasized the dynamic interplay of factors affecting learning autonomy, motivation to study, and both internal and external control throughout the study period.

The qualitative findings suggest that categorising primary school pupils as traditional or non-traditional readers is influenced by internal and external factors linked to pupils' psychological characteristics and instability during puberty. Traditional readers do not always prioritise school-assigned reading and often read out of obligation, whereas non-traditional readers combine school-recommended texts with personally selected texts, especially those relevant to everyday life.

Acknowledgement

This contribution is a part of the research initiative under the Operational Programme Just Transformation, specifically the REFRESH (Research Excellence for Region Sustainability and High-tech Industries) project, registration number CZ10.03.01/00/22_003/0000048. The authors express gratitude for the support.

Ethics Declaration

Ethical clearance is not required for the research referred to in this paper.

AI Declaration

An artificial intelligence tool was not used to create the presented article.

References

Abrahamsson, K. (Ed.) (1987). *Implementing Recurrent Education in Sweden: On Reform Strategies of Swedish Adult Education.* A Selection of Papers from International Conference: Serving the Adult Learner: New Roles and Changing Relationships of Adult and Higher Education. Stockholm, Sweden, May 20-22.

Akimova, N., & Oleksandrenko, K. (2019). Impact of the Internet Using Experience on the Peculiarities of the Internet Texts Understanding. [cit. 2024-05-07] Online. *Psycholingvistika, 26(1), s. 11-36.*

Doležalová, J. (2009). Gramotnost. In J. Průcha (Ed.), Pedagogická encyklopedie (s. 223 - 229). [Literacy. In J. Průcha (Ed.), Pedagogical Encyclopedia (pp. 223 - 229).]

Erdogan, M., & Ok, A. (2011). An Assessment of Turkish Young Pupils' Environmental Literacy: A nationwide survey. *International Journal of Science Education.*

Fazey, D. M. A., & Fazey, J. A. (2001). The potential for autonomy in learning: Perceptions of competence, motivation and locus of control in first-year undergraduate students. Studies in Higher Education, 26(3), 345–361. https://doi-org.ezproxy.muni.cz/10.1080/03075070120076309

Grecmanová, H., & Urbanovská, E. (2007). *Aktivizační metody ve výuce, prostředek ŠVP*. Olomouc: Hanex. [*Activation methods in teaching, a tool for the SEP*. Olomouc: Hanex.]

Hlebová, B., & Vojteková, G. (2018). Diagnostika a stimulácia čitateľskej kompetencie žiakov z marginalizovaných rómskych komunít. Prešov: Vydavateľstvo Prešovskej univerzity. [Diagnostics and stimulation of reading competence of pupils from marginalized Roma communities. Prešov: Prešov University Publishing House.]

Hník, O. (2016). Školní literární kánon. *Český jazyk a literatura*, 66(5), 234–241. [School literary canon. *Czech language and literature*, 66(5), 234–241.]

Höflerová, E. (2011). Elektronický učební text. VŠB – technická univerzita Ostrava. [Electronic learning text. VŠB – Ostrava University of Technology.]

Janků, P. (2014). Cesty k hornictví a hornictvím. Ostrava: Klub přátel hornického muzea. [Paths to mining and mining. Ostrava: Club of Friends of the Mining Museum.]

Janus-Sitarz, A. a Nowak, E. (2014). Język ma siłę sprawczą, czyli o kształceniu kompetencji komunikacyjnych. In A. Janus-Sitarz, & E. Nowak (Eds.), Szkolna polonistyka zanurzona w języku (s. 9–14). Kraków: Universitas [Language has a driving force, which means developing communicative competences. In A. Janus-Sitarz and E. Nowak (Eds.), School Polish studies immersed in language (pp. 9–14). Kraków: Universitas]

Johnson, B., & Činčera, J. (2023). Relationships between outdoor environmental education program characteristics and children's environmental values and behaviors. *Journal of Adventure Education and Outdoor Learning*, *23*(2), 184–201.

Karlík, P., Nekula, M., & Pleskalová, J. (eds.) (2002). Encyklopedický slovník češtiny. Praha 2002, s. 450 [Karlík, P., Nekula, M. and Pleskalová, J. (eds.) (2002). Encyclopedic dictionary of the Czech language. Prague 2002, p. 450]

Kostečka, J. (2002). Český jazyk pro 3. ročník gymnázií. Praha: SPN – pedagogické nakladatelství. [Czech language for the 3rd year of grammar schools. Prague: SPN – pedagogical publishing house.]

Kraus, J. (1997). Jazyk hospodářských dokumentů a písemností. In Český jazyk na přelomu tisíciletí. K vydání připravil František Daneš a kol. Praha: Academia, s. 92-96. [Language of business documents and documents. In The Czech language at the turn of the millennium. Prepared for publication by František Daneš et al. Prague: Academia, pp. 92-96.]

Kroufek, R., & Činčera, J. (2021). Metodický rámec pro environmentální gramotnost ve školách. Souhrnná zpráva. Praha: Ministerstvo životního prostředí. [Methodological framework for environmental literacy in schools. Summary report. Prague: Ministry of the Environment.]

Malach, J., Vicherková, D., & Barot, T. (2020). Locus of control analysed with regards to non-traditional tertiary students of education. *Studia paedagogica*. **25**(4), s. 135-159.

McGeown, S., Bonsall, J. Andries, V., Howarth, D., & Wilkinson, K. (2020). Understanding Reading Motivation across Different Text Types: Qualitative Insights from Children. Journal of Research in Reading, 43(4), 550-608.

M. Ramsay, C., & A. Sperling, R. (2015). Reading Perspective: Can It Improve Middle School Students' Comprehension of Informational Text?.*The Journal of Educational Research*, *108*(2), 81-94.

Nakonečný, M. (2014). *Motivace chování*. Praha: Triton. [*Behavioral Motivation*. Prague: Triton.]

PIRLS (2021). *Mezinárodní šetření PIRLS 2021: Národní zpráva.* (online 16.05.2023). Praha: Česká školní inspekce. [*International PIRLS Survey 2021: National Report*. (online 16.05.2023). Prague: Czech School Inspectorate.] Available from: https://www.csicr.cz/cz/Dokumenty/Publikace-a-ostatni-vystupy/Mezinarodni-setreni-PIRLS-2021-%E2%80%93-Narodni-zprava.

Prázová, I., Homolová K., Landová, H., & Richter, V. (2014). *České děti jako čtenáři*. Brno: Host. [*Czech children as readers*. Brno: Host.]

Rotter, A. (1966). Generalized expectancies for internal versus external control of reinforcements. Psychological Monographs: General and Applied, 80(1), 1–28.

Spínola, H. (2023). Environmental attitudes among students at the University of Madeira, Portugal. International Journal of Development Education and Global Learning, 15(1), 56-68.

Svobodová, S., & Kroufek, R. (2018). Možnosti využití škály MSELS pro testování environmentální gramotnosti na základních školách v České republice. Scientia in educatione, 9(2). [Possibilities of using the MSELS scale for testing environmental literacy in primary schools in the Czech Republic. Scientia in educatione, 9(2).]

UNICEF (2025). *Digital literacy for children: 10 things you need to know* [online 30. 9. 2025] Retrieved from: https://www.unicef.org/innocenti/documents/digital-literacy-children-10-things-you-need-know

Výroční zpráva České školní inspekce (VZ ČŠI) za školní rok 2019/2020. [Annual report of the Czech School Inspectorate for the 2019/2020 school year]. Retrieved from: https://www.csicr.cz/cz/Archiv/Verejne-zakazky/Vyhlaseni-verejnych-zakazek-v-roce-2020/Tisk-Vyrocni-zpravy-2019-2020.

Výroční zpráva České školní inspekce (VZ ČŠI) za školní rok 2022/2023. [Annual report of the Czech School Inspectorate for the 2022/2023 school year]. Retrieved from: https://VZ 2023. https://www.csicr.cz/cz/Aktuality/Kvalita-vzdelavani-ve-skolnim-roce-2022-2023-%E2%80%93-vyr.

AI, Instructor, Observer, and Peer Feedback in Online Microteaching: A Multi-Rater Comparison Study

Veronika Vrábelová, Zuzana Sándorová, Zdena Kráľová, and Elena Kováčiková
Department of English Language and Culture, Constantine the Philosopher University in Nitra, Nitra, Slovakia

veronika.vrabelova@ukf.sk
zsandorova@ukf.sk
zkralova@ukf.sk
ekovacikova@ukf.sk

Abstract: Feedback is central to the effectiveness of microteaching as a teacher training method, yet how different feedback sources compare when evaluating the same performance is an underdeveloped area. In online environments, peer feedback, instructor assessment, external observation, and AI-generated evaluation can all operate alongside one another. Understanding which sources produce meaningful and consistent judgements has practical consequences for teacher educators. This paper presents a case study conducted at Constantine the Philosopher University in Nitra, Slovakia, within a master's-level course on the Didactics of Distance Online Education. Students of the course each delivered an online microteaching session via Google Meet, incorporating at least one AI tool connected to their lesson objective. All sessions were recorded and re-analysed by an external observer. Five rater sources evaluated each microteaching against a shared four-criterion rubric: course peers, the course instructor, the external observer, and two AI models applied under three distinct evaluation conditions: isolated, sequential, and simultaneous. The study investigates how these rater sources compare in their assessments of the same teaching events and whether AI-generated feedback converges with trained human judgement depending on evaluation structure. Findings reveal differences between rater sources. Peer feedback shows a pattern of inflation and an inverse relationship with expert judgement, raising questions about its reliability as a standalone formative instrument. AI evaluation is highly sensitive to evaluation conditions. Isolated assessment produces variable and unreliable results. Sequential evaluation, in which all presenters are assessed in a single continuous conversation, converges closely with the judgements of trained human raters. This convergence was confirmed across both AI models tested. These findings have implications for feedback design in online teacher education. Rater source is as consequential as rubric design. Peer feedback requires pedagogical framing and should not function as a standalone assessment instrument. AI can serve as a supplementary tool when deployment conditions are understood in advance. Pedagogical expert judgement remains the validity benchmark against which any automated or peer-based system must be measured. The study contributes empirical data to a field where AI tools are being adopted faster than the research can evaluate them.

Keywords: Online Microteaching, AI Assessment, Peer Feedback, Multi-Rater Comparison, EFL Teacher Education

1. Introduction

Large Language Models (LLM) propose the possibility of being used as supplementary assessors of student performance. Instructors managing large cohorts can now turn to models such as ChatGPT (OpenAI, 2026) or Gemini (Google, 2025) to help with generating evaluations and reducing the time pressure that comes at scale. Whether these tools can reliably fulfil this role still depends on factors that are not fully explored.

This study investigates that question in pre-service EFL teacher education. Twenty-two online microteaching sessions were evaluated by nine raters: an Instructor, an Observer, Peers, and two AI models. All of them were assessing identical performances against a shared rubric. Three AI evaluation conditions were tested: Isolated, Sequential, and Simultaneous. Peer feedback is examined alongside AI evaluation, as it raises parallel reliability concern and research shows peer scores tend toward leniency due to familiarity with other students (Falchikov and Goldfinch, 2000; Tomal, 2010).

The overreaching aim is to determine whether AI models can serve as dependable supplementary raters and to identify the conditions under which their evaluations most closely align with expert human judgement. All raters used a four-criterion rubric on a five-point Likert scale covering: Overall Effectiveness, Clarity of Instruction, Engagement, and Use of Technology. Three research questions guide this study:

> (RQ1): How do different rater sources compare in their assessments of the same microteaching when evaluated against a shared rubric?
>
> (RQ2): To what extent does peer feedback converge with expert judgement, and what patterns characterise peer rating behaviour?

(RQ3): How does evaluation condition (Isolated, Sequential, Simultaneous) affect the consistency and accuracy of AI-generated feedback relative to human raters?

2. Literature Review

2.1 Background

Microteaching was formalised by Allen and Ryan (1969) as a teacher training method that reduces classroom complexity into a short teaching episode delivered to a small group, followed by structured observation and feedback. The main principle is that isolating a specific teaching skill in a controlled environment allows for targeted development before student teachers enter real-life classrooms. The method has since been adapted widely across teacher education programmes internationally, including online microteaching where videoconferencing platforms replace traditional classroom settings (Tomal, 2010).

Feedback is the core instrument that makes microteaching functional and an important learning experience. Nicol and Macfarlane-Dick (2006) argue that effective feedback must be specific and credible. This is relevant to peer feedback specifically as students cannot provide meaningful and evaluative feedback without being prepared to apply assessment criteria first. In microteaching courses, feedback typically comes from the instructor and from peers but with AI come new possibilities, though Zhang et al. (2026) found that while AI-generated feedback enhanced pre-service teachers' self-efficacy comparably to teacher feedback, it was rated significantly lower in credibility and usefulness.

2.2 Current Position in Research

The reliability of peer feedback in higher education has been thoroughly researched. In a meta-analysis of peer assessment studies, Falchikov and Goldfinch (2000) found that peer scores tend to be inflated compared to instructor scores, with the gap widening as personal familiarity increases. Tomal (2010) documented the same pattern specifically for microteaching contexts, where pre-service teachers tend to assign socially affirming scores. These tendencies appear across contexts and a question how they compare against AI raters gains relevance. Usher (2025) found that peer and instructor grades aligned with each other, while AI assigned higher grades. Usher and Faraon (2025) refined this picture further, showing that the relationship is dependent on quality. In the study, ChatGPT's alignment with instructors improved as project quality increased, while peer-instructor alignment was strongest for lower-quality work. Students themselves perceived ChatGPT as more generous, neutral, consistent, and less attentive. This positions leniency as an aspect that shifts based on what is being evaluated and by whom.

Understanding why AI raters behave as they do requires a closer look beyond the actual score. Seo et al. (2025) identified systematic biases in LLM evaluation, including position bias, knowledge bias, and format bias. Each of these three biases are capable of distorting the scores regardless of the student's performance. The same study found that five-point Likert scale produced inconsistent AI output and models gravitated towards extremes scores. Yavuz et al. (2025) reported strong Intraclass Correlation Coefficient scores for ChatGPT and Bard against experienced EFL teachers but noted that reliability weakened on non-objective criteria such as style and expression, suggesting that AI consistency is criterion dependent. Usher (2025) also found that AI provided more detailed and structured feedback than peers, although sometimes with irrelevant or contradictory content. All these findings suggest that AI evaluation has a unique behaviour that is not strictly lenient or reliable, it is shaped by what the rubric asks it to score and how that judgement is structured.

Evaluation conditions became crucial in Zhuo et al.'s study (2024), where they mention their ProSA framework. They found that prompt sensitivity is caused especially in subjective evaluation contexts, which helps to explain why the same rubric can produce different results depending on how and when it is presented in the model. Further layer was added by Song et al. (2025), that prior conversational history functions as in-context learning, meaning that each additional student a model evaluates in sequence changes the context against which the next is judged. They also identified model-specific positional bias. According to their research, GPT-4o favoured responses presented earlier in a sequence, while Gemini favoured those presented later.

AI evaluation is not stabilised and produces inconsistent results. Its outputs are shaped by model architecture, evaluation structure, prompt design, and conversational history. All of these factors interact with each another in ways that we are slowly beginning to understand. This study addresses the question of how these factors play out in real educational assessment context, alongside peer, instructor, and observer judgement.

3. Methodology

3.1 Participants and Setting

The study was conducted at Constantine the Philosopher University in Nitra within the master student level course *Didactics of Distance Online Education* in the EFL teacher education programme. Twenty-two students participated during the winter semester of 2024/2025 academic year. From the cohort, 19 students were fourth year, and 3 were fifth year students. As part of the course, each had to deliver a 15-minute microteaching via Google Meet with the instructions to target their classmates, as C1 students, and to incorporate at least one AI tool pedagogically connected to their lesson objective. At the beginning of the course, all participants were informed about the nature of the study and signed informed consents. All sessions were recorded with OBS software and later analysed for detailed transcript construction. Throughout the whole study, all information of the participants was anonymised, and we refer to them simply as Presenter Number (Pn).

3.2 The Evaluation Rubric

All raters used the same four-criterion rubric, presented in full in the Figure 1. The rubric covers Overall Effectiveness (pedagogical coherence, achievement of learning objective, contextual appropriateness, and time management), Clarity of Instruction (precision, accessibility, and Multimodal delivery), Engagement (active participation and learner involvement), and Use of Technology (integration, functionality, and pedagogical value of AI and digital tools used). Each criterion is scored on a 1-5 Likert Scale with detailed descriptions at every level.

Category	Criteria Focus	1.	2.	3.	4.	5.
Overall Effectiveness	Achievement of learning objective Pedagogical coherence Contextual appropriateness Time management	Learning objectives are absent or unattainable within the time range Activities are incoherent, disconnected Topic is inappropriate for C1 academic context	Learning objective is visible but unmet Time management is poor Activities have minimal logical connection to the objective Topic is of limited sustainability for C1 academic learners	Learning objectives are evident and partially achieved Time is managed with some inconsistencies Most activities relate to the objective with rationale provided Topic is generally appropriate	Learning objectives are clearly visible and sustainably achieved Time is well-managed Activities are purposefully connected with clear rationale Topic is appropriate for C1 academic learners Comprehension is monitored	Learning objectives are explicitly stated Objectives can be fully achieved Time is managed All activities are purposefully organised Topic is suited for C1 academic level Learner comprehension is monitored
Clarity of Instruction	Precision Accessibility Multimodal delivery	Instructions are absent Incomprehensible Verbal without visual support Repeated clarification is needed	Instructions are partially clear Visual support is inconsistent or absent Unnecessary complexity Requests for clarification	Instructions are generally clear Instructions are occasionally supported both verbally and visually Most participants can proceed without further clarification	Instructions are clear Instructions are verbal and visible on the screen Students can engage with tasks with little need for clarification	Instructions are precise Instructions are delivered verbally and visible on the screen No clarification is needed Students can engage with activity immediately
Engagement	Active participation Learner involvement	Participants remain passive No facilitation strategies Activities require no meaningful participation Learner engagement is not monitored	Engagement is limited to minority Facilitation attempts are minima Activities allow passive participation Monitoring is absent or superficial	Most participants engage adequately The presenter employs basic facilitation strategies Activities encourage some active participation Monitoring occurs periodically	Most participants can engage consistently The presenter uses facilitation strategies Activities require meaningful participation Learner engagement is monitored	All students can engage throughout Presenter continuously facilitates involvement All activities create meaningful engagement Participation is monitored systematically
Use of Technology	Integration Functionality Pedagogical value	Technology is absent Technology disrupts the learning process Tools have not been tested Tools do not add value to activities	Technology is employed but bears little connection to the learning objective Technical difficulties No advantage over traditional instruction	Technology supports learning objectives in a basic manner Tools function without major disruption Some advantage over traditional instruction	Technology is well-integrated Technology meaningfully supports the learning objective Tools function reliably Prior test of tools was done Digital use offers advantage over traditional instruction	Technology is seamlessly integrated Technology enriches the learning objective All tools have been pre-tested and function correctly Technology can enable learning experience that is not possible in traditional instruction

Figure 1: The Evaluation Rubric

3.3 Human Raters and Peer Data Collection

Two raters scored all 22 performances, course instructor and an external observer. Both re-analysed all recordings, transcripts, and scores accessed after the course ended and again a year later for the purpose of this study. The external observer is also one of the authors of this study and to manage this positionality, the re-analysis was introduced from video recordings to ensure that scores were not influenced by memory or context of the course itself.

The peer feedback form conducted via Google Forms used the same criteria and was distributed to all participants except the presenter immediately after they finished their microteaching. Session attendance naturally varied between 19 and 22 participants.

3.4 AI Evaluation Conditions

Two AI models were used to collect data in the form of identical feedback, ChatGPT GPT-5.3 Instant (OpenAI) and Gemini 3 Flash (Google). Both were accessed through paid subscriptions purchased for this study, using the standard conversational type as the most employed among practitioners. Deep-thinking or reasoning mode were strategically not employed for the same reason. The evaluation procedure was identical for both models but varied based on conditions. Three evaluation conditions were applied:

- **Isolated condition:** each transcript with prompt was submitted in a completely new chat session, with no prior context. This was repeated three times per presenter to assess within-condition consistency. Evaluation rubric was included each time with the transcript.
- **Sequential condition:** all 22 participants were submitted one by one within the same continuous conversation which allowed the model to retain all prior evaluations as context. This was repeated as three separate full sessions. Evaluation rubric and the prompt were pasted only once at the beginning of the chat.
- **Simultaneous condition:** all 22 transcripts and evaluation rubric were submitted together in a single prompt. This was also repeated three times.

In all three types of conditions identical prompt was used, except Simultaneous condition, which required to change it from singular to plural form as a technical necessity (*"evaluate student – evaluate students")*. An initial test without these instructions showed that the model returned feedback for only the first student. All AI scores reported in the findings represent means across the three runs per condition.

3.5 Statistical Analysis

Intraclass Correlation Coefficients (ICC) for absolute agreement (A,1) assessed reliability between instructor and observer under a two-way mixed effects model. ICC was selected because both raters used a rubric with numeric scale from 1 to 5. Spearman correlations were computed against all sources using means across all four criteria. A Friedman test compared all nine-rater source and Wilcoxon signed-rank post-hoc comparisons applied to the six AI condition comparisons of interest. All AI scores are means across three runs per conditions. Each statistical method was strategically chosen to answer following questions: What were the absolute scores given? Who was systematically too strict or lenient? What is the overall agreement level? Are the drops and jumps between specific AI conditions real? Does the AI's internal student ranking mimics Instructor and Observer? Statistical analyses were conducted using IBM SPSS Statistics (Version 32).

This study is exploratory and it does not seek to generalise findings across institutions or AI systems. Its main purpose is to surface patterns and generate hypotheses that future and larger-scale research can test in more detail.

4. Findings

4.1 Human Benchmark Reliability

The Instructor and Observer agreed well enough to be taken as reliable benchmark. Their ICC(A,1) was .858, 95% CI [.791, .904], $p < .001$, placing it in the good range (Koo and Li, 2016). They gave identical scores on 68.2% of all cells and differed at maximum only by one point on 94.3%. Their combined mean was 3.19 and it also serves as a benchmark for the statistical analysis. Instructor and Observer used the Likert scale of 1-5 almost fully, with their means minimum scaling from 1.25 to maximum 5.00.

4.1.1 (RQ1) Rater Source Comparison

A Friedman test across all nine rater sources confirmed statistically significant differences, chi-squared (8)=68,983, p<.001, Kendall's W =.392. This reflects moderate agreement across all nine rater sources on the ordering of presenters. It is worth mentioning that it is still closer to the lower end of the scale which can mean that some sources were misaligned with the rest. This can be clearly seen with GPT Isolated. The differences between raters are evident and not random and the data in Table 1 supports this claim. The combination and further division of Instructor and Observer mean created Benchmark for this study.

Table 1: Descriptive statistics, Friedman mean ranks, Spearman r against Instructor and Observer (N=22)

Rater Source	Mean	Standard Deviation	Minimal mean	Maximal mean	Friedman rank	Spearman r vs. Benchmark ((Inst.+Obs.)/2)	p
Instructor	3.16	1.111	1.25	5.00	4.34	-	-
Observer	3.22	1.073	1.50	5.00	4.73	-	-
Peer Feedback	4.27	0.495	2.68	4.70	7.82	-.300	.175 (ns)
GPT Isolated	2.68	0.591	1.42	3.58	1.93	.774***	<.001
GPT Sequential	3.23	0.700	1.92	4.50	5.23	.797***	<.001
GPT Simultaneous	3.10	0.704	1.83	4.25	4.18	.840***	<.001
Gemini Isolated	3.22	0.782	1.67	4.25	4.86	.745***	<.001
Gemini Sequential	3.20	0.747	1.83	4.50	4.82	.687***	<.001
Gemini Simultaneous	3.47	0.801	2.00	4.75	7.09	.761***	<.001

4.2 (RQ2) Peer Feedback

The peer feedback showed interesting divergence from other rater sources. Visible from Table 1, their average mean sits one point higher in comparison to both Instructor and Observer, as well as all 6 AI raters. Their standard deviation of 0.495 is the smallest, which shows their uniform rating style. Interestingly enough, even though peers rated their classmates highly, they never reached complete agreement to give someone mean of 5.00. The Spearman correlation did not reach significance against the Benchmark (r = -.300, p= .175), and it correlated negatively even with all AI raters as well. Two presenters are worth mentioning, specifically P21 and P22. Both are fifth-year students. Their peer mean scores were 3.18 and 2.68, the two lowest in the entire group against average a mean of 5.00 and 4.13 from the Benchmark. Qualitative open-ended feedback, although not included for the statistical purpose of this study, clarifies possible reasons. Peers had following suggestions for improvement for both presenters (P21 and P22): "*Given that it's a micro teaching presentation I feel like it should have been some kind of presentation of some knowledge on the topic.*" Another student stated:" *Longer introduction, not just the activity*," or "*Honestly just re-do it all. It was boring, not engaging, and again he could have checked if those who were using tablets, phones, etc. could upload document into Google Doc.*" Third fifth-year student who followed more formulaic microteaching session similar to the rest of the cohort scored significantly higher (Peer mean = 4.25; Benchmark=4.38).

4.3 (RQ3) AI Evaluation Conditions

Our claim states that AI conditions are important for final evaluation, which can be seen in Table 2. For each condition and each run, the same prompt, evaluation rubric, and transcripts were used, and yet it produced different scores. GPT general mean across three conditions varied from 2.68 to 3.23 to 3.10, Gemini produced more similar results, but means were still not identical.

GPT Isolated produced the lowest mean rank in the Friedman analysis at 1.93. The lowest mean was 1.42 and the highest 3.58. A student who received full score (5/5) from the Benchmark was able to receive only 3.58 in the GPT Isolated condition. Data implies that GPT Isolated treated each presenter's transcript strictly against the evaluation rubric.

GPT Sequential behaved differently. The average mean of 3.23 is very close to the average Benchmark mean of 3.19, and its Spearman correlation against the Benchmark was r = .797, p<.001. In this case, the maximum mean was 4.50, GPT Sequential was able to recognise strong presenters and score them adequately. The Wilcoxon test confirmed that sequential scored 20 out of 22 students higher than isolated (Z= -3.916, p< .001) and Simultaneous scored higher 18 out of 22 students than Isolated (Z= -3.567, p< .001). Comparison between GPT Sequential and Simultaneous showed that Sequential scored 14 out of 22 students higher (Z= -2.302, p< .001), which is the weakest of the three GPT comparisons. Ranking emerges: GPT Isolated < GPT Simultaneous < GPT Sequential.

Gemini conditions produced different results in comparison to GPT. Gemini Isolated and Gemini Sequential were quite well aligned with the Benchmark. Gemini Isolated produced a mean of 3.22, a distance of only +0.03 from the Benchmark of 3.19, and Sequential produced a mean of 3.20, a distance of +0.01. The comparison between them produced a null result (Z= -0.599, p= .549). With only two ties, scores were moving up and down for the same number of students. A possible explanation is that the two conditions approached individual criteria differently, but this claim would require a further analysis to test each criterion separately, and therefore, it remains as a limitation of this study.

Gemini Simultaneous produced repeating inflation as it scored higher than Isolated for 17 out of 22 students (Z= -3.101, p= .002) and higher than Sequential for 21 out of 22 students (Z= -4.020, p< .001). Its mean and Friedman mean rank place it well above the Observer and Instructor scores.

From these findings, GPT Isolated and Gemini Simultaneous seem to be at opposite ends of the spectrum. GPT Isolated is too strict, and Gemini Simultaneous more lenient. Table 2 presents Wilcoxon results, with each condition presented in a shortened version (ChatGPT Isolated = GPT Iso, etc.). Z^a is based on negative ranks, and Z^b is based on positive ranks.

Table 2: Wilcoxon signed-rank results for AI condition comparison

Comparison	Negative Ranks	Positive Ranks	Ties	Z	p
GPTSeq>GPTIso	1	20	1	-3.916[a]	< .001
GPTSim>GPTIso	1	18	3	-3.567[a]	< .001
GPTSeq>GPTSim	14	5	3	-2.302[b]	.021
GemSeq vs GemIso	10	10	2	-0.599[b]	.549
GemSim>GemIso	3	17	2	-3.101[a]	.002
GemSim>GemSeq	0	21	1	-4.020[a]	< .001

Table 3 shows all runs for each condition with the standard deviation and distance from the Benchmark. The standard deviations report on how much the average mean varied between runs. GPT Isolated was the most stable yet had the biggest distance from the Benchmark (-0.51). GPT Sequential, Gemini Sequential, and Gemini Isolated were very close to the Benchmark.

Table 3: AI scores across three runs per condition

Condition	Run 1	Run 2	Run 3	Mean	SD	Distance from the Benchmark (3.19)
ChatGPT Isolated	2.68	2.69	2.66	2.68	.017	-0.51
ChatGPT Sequential	3.19	3.19	3.30	3.23	.059	+0.04
ChatGPT Simultaneous	3.05	3.06	3.19	3.10	.082	-0.09
Gemini Isolated	3.20	3.26	3.20	3.22	.035	+0.03
Gemini Sequential	3.20	3.12	3.26	3.20	.070	+0.01
Gemini Simultaneous	3.42	3.51	3.48	3.47	.046	+0.28

5. Discussion

The main finding of this study is that rater source matters significantly. Although at first, this may not sound like a new finding, after close examination of all rater sources against each other, especially all AI raters, we can see how vast the differences are. All nine raters used the same four-criterion rubric and evaluated the same 22 microteaching transcripts, but the score differences were systematic. Kendall's W of .392 indicates that the nine sources shared a moderate common signal about which students performed better and which worse. This means that the evaluation system as a whole was moderately coherent, and that two sources, peer feedback, and GPT Isolated in particular, were pulling in directions that weakened the overall agreement. Understanding why each source behaved as it did is as important as observing that they differed.

5.1 Peer Feedback

Peer feedback showed almost no variance across students, with a mean of 4.27 (Benchmark mean=3.19) and a standard deviation of only .495. Peers gave nearly identical high scores to almost everyone regardless of the actual performance. The P21 and P22 cases serve as evidence. Both were fifth-year students who received very high scores from the Instructor and Observer but the two lowest peer scores in the cohort. This is consistent with Falchikov and Goldfinch (2000), who found that peer scores inflate in comparison to instructor scores and that personal familiarity between raters and presenters widens the gap. The negative Spearman correlation between peers and the human Benchmark, which extended across all six AI conditions too, suggests peers were scoring highly and in a direction that diverges from all other rater sources. Since these presenters are pre-service EFL teachers who will assess their students, it is important to note that genuine understanding of performance criteria requires structured development (Nicol and Macfarlane-Dick, 2006). Evaluation exercises and post-observation reflection where peer scores are compared directly against instructor's ratings are the kinds of activities that help to build this skill.

5.2 AI Rater Findings

The AI findings reveal that evaluation results produce different scores depending on the condition. Without cohort context, models apply the rubric as an absolute standard with no reference to how the group performed relative to each other. GPT Isolated shows this clearly, as it produced consistent deflation and a ceiling that never reached what the Instructor gave. The mechanism for why Sequential condition corrects this is that processing transcripts one by one within the same conversation provides the model with a frame of reference. It learns what the specific cohort looks like before scoring. This is similar to Song et al. (2025), who found that prior conversational history functions as in-context learning for GPT models specifically.

Gemini behaved differently, which asks for further explanation. Gemini Isolated was already close to the human Benchmark without any cohort context, and the comparison between Gemini Sequential and Isolated produced a null result ($p = .549$). This suggests that Gemini calibrates differently in Isolation, possibly because it can be more sensitive to the evaluation rubric and prompts than GPT is. The asymmetry in Simultaneous condition is also quite revealing. GPT Simultaneous improved on GPT Isolated, which suggests that having all transcripts present in one prompt provided some comparative context. Gemini Simultaneous inflated scores for 21 out of 22 presenters above Sequential levels, which produced the strongest Wilcoxon result. A possible hypothesis is that Gemini's higher context sensitivity makes it vulnerable to the structure of a simultaneous prompt. When all 22 transcripts were approached at once, it may have read the amount of content as a signal that this is a high-performing group and scored according to that. This makes its Isolated condition more calibrated but Simultaneous more vulnerable. Although it remains as a hypothesis and requires further evidence.

GPT Simultaneous produced the highest Spearman correlation out of all AI rates across conditions ($r = .840$), which indicated strong rank-order agreement with the Benchmark. It reflects a correlation but does not prove the accuracy of absolute scores. GPT Simultaneous remained below the Benchmark in mean score and was outranked by GPT Sequential in direct Wilcoxon comparison, making Sequential more suitable. This is worth mentioning as it poses an interesting addition to our findings.

For future study, each criterion should be analysed separately. The current analysis worked at the presenter level, averaging the mean across all four criteria. Conditions may have differed not only in overall score but also in how much importance they assigned to specific criteria, which could explain some of the individual-level variation between runs. Knowing which criterion is most sensitive could help instructors use these tools more precisely.

5.3 Limitations

This study has several limitations. It is a single case study at one institution with 22 presenters, which limits generalisability. AI models evaluated written transcripts, while human raters had access to both video recordings and transcripts. The three AI conditions were not structurally identical, as the Simultaneous required a slightly different framing, changing the noun for "*presenter*" from singular to "*presenters*" in plural. The study tested two specific model versions at a specific timeframe, and since both models are updated continuously, the same study may produce different results in future.

6. Conclusion

This study compared nine rater sources evaluating the same 22 microteaching sessions with an identical rubric. Rater sources had a statistically significant effect on scores. Peer feedback inflated scores and could not differentiate between stronger and weaker performances. It correlated negatively with Instructor and Observer Benchmark but also with all AI conditions. Peers were ranking the cohort in the opposite direction from every trained human or AI raters. Sequential AI evaluation aligned with the human Benchmark for both ChatGPT and Gemini. GPT Isolated underperformed, while Gemini Isolated was already similar to the Benchmark. Gemini Simultaneous produced inflation, while GPT Simultaneous performed moderately.

The practical recommendation is that Sequential evaluation for both models proved to be the safest. It is the only condition that produced results close to the human Benchmark for both models. It avoided the inflation of Gemini Simultaneous and the strict scoring of GPT Isolated. Without understanding of how evaluation conditions affect AI output, using AI feedback risks adding to instructors' workload. It may also create an unfair environment for students who are assessed differently depending on how the tool was set up and what condition was used. The trained pedagogical professionals should remain the primary evaluators of student performance, especially where feedback needs to be given to complex performance. AI models in their current form are best used sparingly as a supplementary source that can support their judgement or help manage large cohorts. Until AI models are calibrated specifically for pedagogical evaluation and tested to perform reliably, instructors should treat their output as one of the possibilities but not as a final decision. Given the exploratory nature of this study, all recommendations should be treated as working hypotheses.

Ethics Declaration

Signed consent forms were collected at the beginning of the course, and all participants were informed about the nature of the study. All data were anonymised and no personal information was necessary for conducting this research.

AI Declaration

AI models were used as evaluators of microteaching transcripts and formed central component of the research. AI assistance was used during the writing process to check for grammatical errors.

Acknowledgments

The current study is part of the KEGA 027UKF-4/2025 project titled *Didactics of Online and Distance English as a Foreign Language Teaching - Textbook and Multimedia Support Design*. It also forms part of the UGA project V/16/2026 titled *Využívanie umelej inteligencie v dištančnom vyučovaní anglického jazyka.*

References

Allen, D.W. and Ryan, K. (1969) *Microteaching*. Reading, MA: Addison-Wesley.

Falchikov, N. and Goldfinch, J. (2000) "Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks", *Review of Educational Research*, Vol 70, No. 3, pp. 287-322.

Google (2025) *Gemini 3 Flash,* Large language model, https://gemini.google.com/app [Accessed: 28 March 2026]

Koo, T.K. and Li, M.Y. (2016) "A guideline of selecting and reporting intraclass correlation coefficients for reliability research", *Journal of Chiropractic Medicine*, Vol 15, No. 2, pp. 155-163, https://doi.org/10.1016/j.jcm.2016.02.012.

Nicol, D.J. and Macfarlane-Dick, D. (2006) "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice", *Studies in Higher Education*, Vol 31, No. 2, pp. 199-218, https://doi.org/10.1080/03075070600572090.

OpenAI (2026) *ChatGPT GPT-5.3 Instant*, Large language model, https://chatgpt.com/ [Accessed: 28 March 2026]

Seo, H., Hwang, T., Jung, J. et al. (2025) "Large language models as evaluators in education: verification of feedback consistency and accuracy", *Applied Sciences*, Vol 15, No. 2, p. 671, https://doi.org/10.3390/app15020671.

Song, M., Zheng, M. and Luo, X. (2025) "Can many-shot in-context learning help LLMs as evaluators? A preliminary empirical study", in: *Proceedings of the 31st International Conference on Computational Linguistics.* Abu Dhabi: Association for Computational Linguistics, pp. 8232-8241, https://aclanthology.org/2025.coling-main.548/.

Tomal, D.R. (2010) *Action research for educators.* 2nd edn, Rowman & Littlefield, Lanham, MD.

Usher, M. (2025) "Generative AI vs. instructor vs. peer assessments: a comparison of grading and feedback in higher education", *Assessment & Evaluation in Higher Education*, Vol 50, No. 6, pp. 912-927 https://doi.org/10.1080/02602938.2025.2487495.

Usher, M. and Faraon, M. (2025) "Who grades best? Comparing ChatGPT, peer, and instructor evaluations across varying levels of student project quality", *Assessment & Evaluation in Higher Education,* pp. 1-20, https://doi.org/10.1080/02602938.2025.2588682.

Yavuz, F., Çelik, Ö. and Yavaş Çelik, G. (2025) "Utilizing large language models for EFL essay grading: an examination of reliability and validity in rubric-based assessments", *British Journal of Educational Technology*, Vol 56, pp. 150-166, https://doi.org/10.1111/bjet.13494.

Zhang, C., Hu, M., Wu, W., Chen, Y., Wang, K., Gao, T. and Fu, Q. (2026) "AI versus teacher feedback in developing pre-service teachers' teaching self-efficacy: a quasi-experimental study in simulated teaching", *Teaching and Teacher Education*, Vol 172, article 105355, https://doi.org/10.1016/j.tate.2025.105355.

Zhuo, J., Zhang, S., Fang, X., Duan, H., Lin, D. and Chen, K. (2024) "ProSA: assessing and understanding the prompt sensitivity of LLMs", in: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Association for Computational Linguistics, Miami, FL, pp. 1950–1976, https://aclanthology.org/2024.findings-emnlp.108/.

How Gaming Activities Shape Lower Secondary School Pupils' Understanding of Artificial Intelligence

Magdalena Zavodna, Hana Ulrichova and Katerina Kostolanyova
University of Ostrava, Faculty of Education, Department of Information and Comunication Technologies, Ostrava, Czech Republic

Magdalena.Zavodna@osu.cz
Hana.Ulrichova@osu.cz
Katerina.Kostolanyova@osu.cz

Abstract: As artificial intelligence (AI) becomes increasingly integrated into everyday life, there is a growing need to develop not only technical skills among lower secondary school pupils, but above all an understanding of how AI works and the ability to use it critically and effectively. The aim of this study was to determine whether incorporating selected game-based activities, specifically escape rooms and activities focused on "learning" a simple artificial intelligence model related to ocean cleanup and the creation of fake news in the Canva app using AI tools, leads to a better understanding of how AI works and improves pupils' ability to use these technologies effectively. The research was conducted as a pedagogical experiment at a lower secondary school in the Czech Republic. A total of 56 pupils aged 14 to 15 were involved in the study, divided into one control group and two experimental groups. The experiment took place over three class periods as part of the computer science curriculum, with data collected through pre-tests and post-tests, analysis of pupils' outputs, and a self-evaluation questionnaire in which pupils assessed individual activities and the extent of their impact on their understanding of the subject matter and motivation to learn. The results show that the pupils in the experimental group achieved significant improvement compared to the control group. These pupils demonstrated a greater understanding of the basic principles of how artificial intelligence works, as well as an improved ability to effectively use AI tools when solving problems. At the same time, a higher level of motivation to learn was observed. The strongest motivation was identified in the activity focused on creating fake news. The findings suggest that the implementation of game elements into computer science instruction represents an effective approach to developing AI literacy in lower secondary school. The study thus contributes to expanding knowledge about the possibilities of didactically utilizing game-oriented activities in the field of artificial intelligence education and offers insights for further research and pedagogical practice.

Keywords: Pedagogical Experiment, Lower Secondary School, Artificial Intelligence, Escape Room, Educational Apps.

1. Introduction

The rapid integration of AI into society requires an urgent redefinition of basic education (UNESCO, 2021). Traditional digital literacy, which focuses on software and hardware control, is insufficient for modern algorithmic systems (Long and Magerko, 2020). Pedagogical focus must shift toward complex competencies addressing both the cognitive and ethical aspects of human-AI coexistence (Ng et al., 2021).

We live in an era where AI has ceased to be the domain of science fiction literature and specialized laboratories and has become an integral part of everyday reality (Xiaoyu et al, 2025). The younger generation interacts with these technologies constantly, often without a deeper understanding of how they work (Long and Magerko, 2020)

However, while youth intuitively consume AI technologies, they often lack a deeper understanding of how they work, their limitations, and their risks (Long and Magerko, 2020). Educational systems must transform basic technical instruction into comprehensive AI literacy that encompasses technological understanding, effective use, and critical thinking (Long and Magerko, 2020; Ng et al., 2021; Touretzky et al., 2019).

Most AI education research targets college or high school settings, leaving a shortage of studies focused on lower secondary pupils (Sanusi, 2021). Highly abstract concepts like machine learning or data bias are difficult for younger students to grasp. This problem is highly relevant to the Czech lower secondary education system, which is currently revising its Framework Educational Programs (FEP) to emphasize "new computer science." Current teaching practices lack empirical data from real-world pedagogical experiments, leaving most recommendations purely theoretical.

The aim of this research is therefore to verify the impact of activities incorporating gamification elements on pupils' understanding of AI-related content. Based on this objective, the main research question was formulated: *What impact does teaching with gamification elements have on understanding how AI works*? Subsequently, the hypothesis was established: *Teaching with gamification elements leads to a greater understanding of the*

principles of how AI works. A sub-research question further focuses on: *How do pupils perceive the motivational effect of the proposed activities*?

The significance and usefulness of this study are thus evident on three key levels. The study provides empirical evidence on how game-based design (escape games, interactive simulations, creative digital production) affects both the cognitive (understanding) and affective (intrinsic motivation) components of learning. It offers an immediately applicable toolkit for computer science teachers, featuring three validated educational scenarios: a digital escape game, an AI learning simulation, and the critical creation of fake news in the Canva app. It fosters critical AI literacy, teaching pupils to deconstruct AI-generated content (e.g., recognizing fake news) to defend against disinformation, manipulation, and the loss of digital autonomy. The development of critical AI literacy is a key weapon against disinformation, manipulation, and the loss of digital autonomy. By showing how to teach pupils to deconstruct AI-generated content (e.g., recognize fake news), the study directly contributes to building society's resilience against the risks associated with the massive rise of generative models.

1.1 Artificial Intelligence in Lower Secondary Education

According to Long and Magerko (2020), "AI literacy" extends beyond programming; it is a set of competencies enabling individuals to critically evaluate, communicate with, and utilize AI as a problem-solving tool. This framework consists of three dimensions:

- Cognitive: Understanding algorithmic principles, data structures, and machine learning limits.
- Instrumental: Effectively deploying AI tools (e.g., prompt formulation).
- Ethical-critical: Reflecting on privacy, algorithmic bias, and socioeconomic impacts (Ng et al, 2021).

Critical thinking serves here as a fundamental defensive and productive mechanism. Given that generative models exhibit an inherent tendency toward so-called confabulations or hallucinations, students' ability to subject AI outputs to rigorous verification becomes a fundamental prerequisite for digital autonomy (Fulsher et al, 2026).

The integration of AI-related topics into official educational programs is undergoing significant evolution both globally and locally. International strategic frameworks, such as UNESCO's (2021) recommendations for AI in education, explicitly call on national governments to incorporate AI instruction into the core curricula of elementary and secondary schools. The goal is to eliminate the threat of a new type of digital divide that would isolate the technologically illiterate population.

In the Czech Republic, this is driving the revision of the FEP and the introduction of "new computer science" (Kostolanyova et al, 2023). This shift moves instruction away from basic user support toward computational thinking. Computational thinking acts as a pedagogical bridge; teaching machine learning is most effective when students engage empirically with datasets rather than abstract mathematical theory (Sanusi, 2021).

Students aged 12 to 15 operate within Piaget's stage of formal logical operations, allowing them to handle abstract concepts and social causality (Inhelder and Piaget, 1958). While younger children learn through anthropomorphic games (Druga et al, 2019), adolescents require authentic learning where they can actively deconstruct algorithm architectures, such as social media recommendation systems.

However, educators must overcome two distinct barriers in this age group:

- Technocentric skepticism: Believing the material is mathematically inaccessible.
- Uncritical technological optimism: Accepting AI-generated outputs without cognitive effort.

1.2 The Potential of Gamification in Education

In contemporary education, we are increasingly faced with the need to transform traditional teaching methods so that they align with digital reality and enhance students' intrinsic motivation. One of the most prominent trends addressing this need is the implementation of game principles into the educational process, in other words, gamification.

In academic settings, gamification is most commonly defined as the use of game design elements in non-game contexts (Deterding et al, 2011). In the context of education, this is not merely about "playing games" in class, but about the deliberate structuring of instruction using mechanisms such as points, badges, leaderboards, and above all, narrative, immediate feedback, and the opportunity to learn from mistakes without fatal consequences (Kapp, 2012).

According to Werbach and Hunter (2012), gamification operates on three levels:

- Dynamics: Broader concepts such as narrative, emotion, and the gradual revelation of information.
- Mechanics: Processes that drive the activity forward (challenges, cooperation, competition).
- Components: Specific tools (points, levels, tasks).

In education, gamification has a strong psychological foundation. According to self-determination theory (Ryan and Deci, 2000), game elements increase intrinsic motivation by fulfilling three basic human needs: autonomy (the ability to make choices), competence (experiencing success in overcoming challenges), and relatedness (collaboration with peers).

A specific and recently highly popular form of gamification is Educational Escape Rooms (EER). They are based on commercial escape rooms, but their primary goal is not merely entertainment, but rather the achievement of specific educational objectives (Veldkamp et al, 2020).

An escape room in education provides an environment where students collaborate to solve a series of puzzles and tasks within a limited time to achieve a specific goal (e.g., “escaping” the room or solving a mystery).

Fotaris and Mastoras (2019) emphasize that escape games transform the student’s role from a passive recipient of information to an active problem-solver. The main benefits of EER include:

- Development of soft skills: They significantly stimulate communication, teamwork, and critical thinking.
- Immediate feedback: Students immediately see whether their solution is correct (the key fits, the code works), which supports so-called natural learning from mistakes.
- Linking theory with practice: Tasks require the application of knowledge in a simulated but realistic context.

Combining gamification with AI instruction in lower secondary school appears to be a highly effective teaching strategy. AI concepts (such as machine learning, algorithms, or pattern recognition) are often abstract and difficult for students to grasp. As Gee (2003) notes in the context of game-based learning theory, good games create so-called “zones of proximal development,” where students are able, with modest support, to solve tasks that would be too complex for them on a purely theoretical level. An escape game or gamified activity (e.g., training a model to clean up the oceans or detect fake news) gives abstract algorithms tangible meaning. Students thus do not merely consume theory about AI, but experience its functionality in practice, which directly leads to a critical understanding of technologies, the core of digital literacy (Ng et al, 2021).

1.3 A Didactic Approach to Selected AI Concepts

The effective implementation of AI in lower secondary education requires the translation of complex algorithmic concepts into age-appropriate teaching models. Contemporary pedagogy therefore prioritizes active and game-based learning, which allows students to understand how complex systems work through experimentation and interaction.

The development of generative AI is shifting the nature of digital creation toward co-creation between humans and AI (Xiaoyu et al, 2025; Lin and Tan, 2025). The integration of generative tools into graphic editors, such as Canva, helps students overcome technical barriers and focus on the concept and content of their creations. Studies confirm that the use of text-to-image generators increases motivation, develops language and logical skills, and supports students’ creative self-efficacy (Xiaoyu et al, 2025; Lin and Tan, 2025). A key role is played by the creation of precise prompts, which guide students toward more accurate expression and work with syntax.

However, there is a risk of excessive automation in content creation and a weakening of deeper cognitive engagement. Effective teaching must therefore include reflection, critical evaluation of outputs, and discussion of ethical issues, copyright, and plagiarism (Long and Magerko, 2020; Touretzky et al, 2019).

One of the most pressing challenges today is also the misuse of generative AI to create disinformation and deepfakes. Traditional media education focused solely on text analysis is no longer sufficient (Fulsher et al, 2026). Therefore, a “prebunking” strategy is gaining traction in education, that is, preventive “vaccination” against disinformation, in which students learn the principles of manipulative content in a safe environment (Ecker et al, 2022; Fulsher et al, 2026).

Research shows that the most effective approach is the guided creation of simulated fake news by the students themselves (Fulsher et al, 2026). When creating manipulative news using AI tools, students become aware of the psychological mechanisms of disinformation and the ease with which it spreads. Subsequent fact-checking and reflection strengthen critical thinking, skepticism toward unverified information, and long-term information resilience (Ecker et al, 2022).

2. Methodology

2.1 Research Design

To achieve the study's objectives, a pedagogical experimental research design was chosen, consisting of one control group and two experimental groups. The independent variable was the type of instruction (traditional vs. instruction with gamification elements), while the dependent variable was the level of understanding of artificial intelligence principles, operationalized through a test score. The research used a pretest–posttest design, which allowed for tracking changes in pupils' knowledge before and after the intervention.

2.2 Research Sample

The research sample consisted of a total of 56 pupils aged 14 to 15. The pupils were divided into three classes. The control group (Con) consisted of 20 pupils, experimental group 1 (Ex1) consisted of 17 pupils, and experimental group 2 (Ex2) consisted of 19 pupils.

The groups were considered comparable in terms of previous performance in computer science and socioeconomic background, based on school records and the teacher's assessment.

2.3 Course of the Intervention

The experiment took place over three consecutive 45-minute computer science classes. The individual groups differed in the extent and form of gamification elements used; see Table 1. The control group underwent traditional instruction based primarily on lecture-style teaching and independent student work without gamification elements (lecture + worksheets). The experimental groups further differed from one another in the versions of the activities used. In the Ex1 group, only partial gamification was implemented in the form of an online escape game; however, the "ocean cleanup" activity was carried out here in the same way as in the Con group, namely in the form of a worksheet. The Ex2 group underwent fully gamified instruction (all activities in an interactive format). This arrangement made it possible to compare different levels of gamification in instruction.

Table 1: Comparison of the activities used in all groups

Activity	Control Group (Con)	Experimental Group 1	Experimental Group 2
Escape game	lecture-style teaching	✓	✓
AI learning simulation	worksheet	worksheet	✓
The creation of fake news	✓	✓	✓

The first class began with a pre-test to assess the pupils' initial level of knowledge and skills in the field of AI. This was followed by an activity called "AI Lab," in which pupils were tasked with completing an online escape game. In the first part of the second lesson, pupils tried their hand at AI training using the HourOfCode web app in an activity focused on "Cleaning Up the Ocean," or completed a worksheet on the same topic. In the second part of the lesson, pupils were given a final assignment: working in pairs, they were to create a "fake news" newspaper article in the Canva app using AI tools to generate fake images. In the final lesson, pupils had time to complete their own fake news piece and then took a post-test and a motivational questionnaire. The instruction for all groups aligned with the theme "How New Technologies Work Around Me" according to the current FEP and School Educational Program.

2.4 Description of the Activities Used

The intervention program in the experimental groups was based on the pedagogical concept of game-based learning and consisted of three key thematic activities, each supported by a specific software tool.

2.4.1 Digital Escape Game "AI Lab"

The first activity focused on introducing basic AI terminology and concepts. For this purpose, an educational escape game was created in the Genially app. During the game, pupils completed individual tasks focused on how AI works, detecting AI outputs, and prompt engineering. For correctly solving each task, pupils received a piece of code. At the end of this activity, they were prompted to unlock the final phase of the game (stabilizing the AI in the lab) and, by entering the code, they discovered whether they had made any mistakes while completing the individual tasks.

2.4.2 Machine Learning Simulation "Cleaning Up the Ocean"

The HourOfCode educational app was used to help pupils understand the principles of training AI models and the issue of data bias. Pupils took on the role of AI programmers and manually classified objects on the screen (fish vs. trash), thereby creating a training dataset for AI machine learning. They then analyzed the success of their model and deliberately tested how insufficient or one-sidedly biased data (e.g., training only on plastic bottles) would affect AI's ability to correctly identify other types of trash. This activity directly visualized the AI learning process.

2.4.3 Creating "Fake News"

The final activity linked the technical functioning of generative AI with the development of critical thinking and media literacy. Pupils were tasked with creating believable-looking fake news in the Canva graphic editor using integrated AI image generators. In the final phase of the activity, a guided deconstruction of these outputs took place, during which pupils analyzed specific artifacts of the generated content (e.g., syntactic anomalies, anatomical errors in the generated images) and discussed the ethical risks associated with the misuse of AI.

2.5 Research Tools

Pupils' cognitive knowledge was measured using an identical pretest and posttest consisting of 10 items (7 multiple-choice questions and 3 open-ended questions requiring a brief response). This was a knowledge test focused on understanding the principles of AI, recognizing AI outputs, and basic application skills. The content validity of the test was verified through expert assessment by two experts in computer science education.

The research also included a questionnaire that pupils completed at the end of the final class session. The questionnaire contained 7 items on a 5-point Likert scale (strongly disagree to strongly agree) and 3 closed-ended questions with yes/no options. The questionnaire assessed the pupils' level of motivation, the subjective benefits of the given game elements for learning, the clarity of the activities, and their preferred type of activities.

2.6 Data Processing

Quantitative data from the pretests and posttests were analyzed statistically using the PAST4 software. First, the normality of the data distribution was tested using the Shapiro-Wilk test. Since the data were normally distributed, a parametric paired t-test was used to compare the results across groups (Con, Ex1, Ex2).

3. Results

3.1 Comparison of Pretests and Posttests

Before proceeding to the actual testing of the null hypotheses, basic descriptive statistics (sample means and standard deviations) were calculated for the pretest and posttest results, as well as for the absolute educational gain (defined as the difference between the posttest and pretest scores). The research sample consisted of a total of 56 respondents divided into three groups: experimental group 1 (N = 17), experimental group 2 (N = 19), and the control group (N = 20). The maximum possible score on both tests was 10. A detailed descriptive overview is summarized in Table 2.

Table 2: Average score

Group	Pretest M (SD)	Posttest M (SD)	Average Profit (SD)
Ex1	5,82 (1,29)	8,94 (1,30)	+3,12 (1,93)
Ex2	5,05 (1,61)	8,42 (1,12)	+3,37 (1,77)
Con	5,75 (1,68)	8,00 (1,56)	+2,25 (2,38)

To select the appropriate statistical method (parametric vs. nonparametric tests), the normality of the distribution of differences (gains) was tested for each group using the Shapiro-Wilk test. The significance level was set at the standard 0.05. If $p > 0.05$, the data are normally distributed. No statistically significant deviation from a normal distribution was found in any of the three groups (see Table 3). The normality of the data was thus confirmed, and a paired t-test could be used for further analysis (see Tables 4, 5, and 6).

Table 3: Data normality test

Group	Ex1	Ex2	Con
N	17	19	20
Shapiro-Wilk W	0,9498	0,9392	0,9051
p(normal)	0,4526	0,2555	0,05148

Table 4: Ex1 paired t-test

t test			
Mean difference:	3,1176	95% conf.:	(2,1239 4,1113)
t:	-6,651	p (same mean):	5,57E-06
Exact:	-	p (same mean):	0,00012207

Table 5: Ex2 paired t-test

t test			
Mean difference:	3,3684	95% conf.:	(2,5151 4,2218)
t:	-8,2931	p (same mean):	1,46E-07
Exact:	-	p (same mean):	3,81E-06

Table 6: Con paired t-test

t test			
Mean difference:	2,25	95% conf.:	(1,115 3,385)
t:	-4,1491	p (same mean):	0,00054513
Exact:	-	p (same mean):	0,0014038

The aim of the main phase of the analysis was to verify whether respondents in the individual groups demonstrated a demonstrable increase in knowledge compared to the control group after completing the experimental instruction. For the Ex1 group, the average score increased from 5.82 on the pretest to 8.94 on the posttest. A paired t-test (Table 4) confirmed that this average increase of +3.12 points is statistically highly significant ($t = 6.65$; $p < 0.001$). The calculated Cohen's d value of 1.61 indicates, according to standard criteria, an extremely large effect of the intervention. In the case of the second experimental group (Ex2), a shift was recorded from a baseline value of 5.05 to a final score of 8.42 points. This average gain of +3.37 points proved to be statistically the most significant based on a paired t-test (Table 5) ($t = 8.29$; $p < 0.001$). Cohen's $d = 1.90$ indicates an exceptionally strong scientific effect of the applied method. This finding indicates a strong influence of gamification activities on pupils' acquired knowledge.

In the control group (Con), the mean shifted from 5.75 to 8.00 points (average gain of +2.25). Based on a paired t-test (Table 6), a statistically significant improvement was also observed in this group. Although the effect size expressed by Cohen's $d = 0.94$ still represents a significant effect, it is evident that it is lower than in both experimental groups.

The hypothesis that teaching incorporating gamification elements leads to better mastery of the subject matter and a better understanding of how AI works was confirmed by the results from both experimental groups.

3.2 Evaluation Questionnaire

All pupils in the Ex2 group unanimously agreed that they would like to use this type of activity more often in class. In the Ex1 group, however, this figure was “only” 70.6% of pupils.

Creating their own "fake news" in the Canva app was successful across all groups, with the strongest motivational effect observed in the partially gamified group Ex1 (76.5%) and the fully gamified group Ex2 (68.4%), while in the control group without a gaming context, only 55% of pupils found it enjoyable.

The online escape game was also a big hit with pupils in the experimental groups; 70.6% of pupils in group Ex1 enjoyed it, and 63.2% of pupils in group Ex2 enjoyed it. In group Ex2, however, of all three activities, pupils enjoyed teaching AI about ocean cleanup the most, at 78.9%.

Pupils in the fully gamified environment (Ex2) rated the activities as highly understandable (78.9%) and easy to complete (68.4%). A full 73.7% of them reported that these games helped them understand the material better, and 57.9% believe they will remember it longer as a result. In the Ex1 group, which had a lower level of interactivity, this subjective effect was not as pronounced. 58.8% of pupils rated the activities as understandable, and 52.9% as easy to complete. Only 41.2% of pupils reported that these activities helped them better understand the subject matter and remember it longer (52.9%).

4. Conclusion

The main research hypothesis, that teaching incorporating gamification elements leads to a better understanding of how AI works, was unequivocally confirmed by the results of the pedagogical experiment. The data clearly show that both experimental gamified teaching methods demonstrate high and consistent effectiveness. The implementation of both types of interventions (Ex1 and Ex2) led to a demonstrable and robust increase in pupils’ knowledge, as evidenced by the extremely high effect sizes. When comparing Ex1 and Ex2, the collected data do not allow us to designate either method as statistically superior. The Ex2 group did show a slightly higher growth rate (net gain), which could have been caused by a higher number of gamification activities or simply by knowledge differences among pupils. However, all these minor deviations are fully within the limits of natural sampling variability. From a pedagogical perspective, both experimental methods can be considered comparably successful alternatives that demonstrably outperform the standard growth observed in the control group.

From the perspective of pedagogical practice, a key finding is that, based on the collected data, no statistically significant difference in performance was found between partial (Ex1) and full (Ex2) gamification. Minor deviations in growth are fully within the bounds of natural sample variability. For teachers, this leads to an optimistic conclusion: even the inclusion of just one game element (e.g., online escape games) can dramatically increase the effectiveness of instruction compared to conventional lectures with worksheets.

The sub-research question focused on how pupils perceive the motivational effect of the proposed activities. The results confirmed that the implementation of game elements has a strong psychological impact on intrinsic motivation, comprehension of the material, and the integration of theory and practice. The activity focused on machine learning simulation, “Ocean Cleanup,” became the absolute favorite of the fully gamified Ex2 group. This simulation clearly visualized the AI learning process and the risks associated with one-sided, biased data for the pupils.

Teaching with gamification elements also had a significant motivational effect in the experimental groups and a positive influence on easier comprehension of the subject matter and its longer retention, which is also confirmed by data from other studies focused on game-based learning (e.g., Veldkamp et al., 2020; Gee, 2003; Kapp, 2012).

The study offers an immediately applicable toolkit for computer science teachers, directly supporting the Czech Republic's “new computer science” curricular shift toward developing computational thinking. By creating simulated fake news in a safe classroom environment, pupils uncover the psychological triggers of disinformation and learn to rigorously verify AI outputs, building long-term societal resilience against generative models. Integrating game elements and digitally relevant scenarios into lower secondary instruction allows

adolescents to successfully deconstruct complex technological concepts. This shifts them from naive consumers into critical editors of their own informational reality.

This study also has its limitations, including, for example, a small sample size (56 pupils) from a single school, their specific age range (14–15 years), the short duration of the intervention, and the absence of a follow-up assessment, which made it impossible to objectively verify long-term retention of the material; instead, only the immediate improvement in knowledge and the pupils' subjective opinions could be assessed.

Ethics Declaration

The research was conducted in a lower secondary school setting with prior written informed consent obtained from the parents or legal guardians of participating pupils. No additional formal ethical clearance was required for this research.

AI Declaration

ChatGPT was used solely for language editing and grammatical refinement. The ideas, structure, interpretations, and conclusions presented in this manuscript are entirely the authors' original work.

References

Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pp 9–15. https://doi.org/10.1145/2181037.2181040

Druga, S., Vu, S. T., Likhith, E., and Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*, pp 104–111. https://doi.org/10.1145/3311890.3311904

Ecker, U. K., Lewandowsky, S., Cook, J. et al. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, pp 13–29. https://doi.org/10.1038/s44159-021-00006-y

Fotaris, P. and Mastoras, T. (2019). Escape Rooms for Learning: A Systematic Review. In *13th European Conference on Games Based Learning*, pp 235–243, https://doi.org/10.34190/GBL.19.179

Fulsher, A., Pagkratidou, M., Kendeou, P. (2026). GenAI and misinformation in education: A systematic scoping review of opportunities and challenges. *AI and Society*, Vol 41, No. 1, pp 1373–1385. https://doi.org/10.1007/s00146-025-02536-y

Gee, J. P. (2003). What Video Games Have to Teach Us About Learning and Literacy. *Computers in Entertainment*, pp 1–4. https://doi.org/10.1145/950566.950595

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Basic Books. https://doi.org/10.1037/10034-000

Kapp, K. M. (2013). T*he gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons Inc.

Kostolanyova, K., Cirus, L., Javorcik, T., Simonova, I. (2023). A New Concept of the Informatics Curriculum in the Czech Republic: Teacher Reflection on the First Year of the Transition Period. *Sustainability*, Vol 15, No. 5, pp 1–19. https://doi.org/10.3390/su15054091

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp 1–16. https://doi.org/10.1145/3313831.3376727

Lin, X. and Tan, H. (2025). A Systematic Review of Generative AI in K–12: Mapping Goals, Activities, Roles, and Outcomes via the 3P Model. *Systems*, Vol 13, No. 10, pp 840–865. https://doi.org/10.3390/systems13100840

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W and Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence.* Vol 2, No. 1*,* pp 1–11. https://doi.org/10.1016/j.caeai.2021.100041

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*. Vol 25, No. 1, pp 54–67. https://doi.org/10.1006/ceps.1999.1020

Sanusi, I. T. (2021). Teaching machine learning in K-12 education. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, pp 395–397. https://doi.org/10.1145/3446871.346976

Touretzky, D., Gardner-McCune, C., Martin, F. and Seehorn, D. (2019). Envisioning AI literacy for K-12 What Should Every Child Know about AI?. *Communications of the ACM*, Vol 33, No. 1, 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

UNESCO. (2021). *AI and education: Guidance for policy-makers.* United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54675/PCSP7350

Veldkamp, A., Daemen, J., Teekens, S. et al. (2020). Escape boxes: Bringing escape room experience into the classroom. *British Journal of Educational Technology,* Vol 51, No. 4, pp 1220–1239. https://doi.org/10.1111/bjet.12935

Werbach, K. and Hunter, D. (2012). *For the Win: How Game Thinking can Revolutionize your Business*. Wharton Digital Press.

Xiaoyu, W., Zainuddin, Z., and Hai Leng, C. (2025). Generative Artificial Intelligence in Pedagogy: A Systematic Review of Empirical Studies (2022–2024). *Cogent Education,* Vol 12, No. 1, https://doi.org/10.1080/2331186X.2025.2485499

Toward Intersectionally Fair Educational AI for Learning Analytics and Student Guidance

Stefania Zourlidou, Tai Le Quy, Resmin Hossain and Frank Hopfgartner
University of Koblenz, Koblenz, Germany

zourlidou@uni-koblenz.de
taileguy@uni-koblenz.de
resminhossain@uni-koblenz.de
hopfgartner@uni-koblenz.de

Abstract: Artificial intelligence is increasingly embedded in learning analytics dashboards, adaptive learning environments and student-guidance tools. These systems do more than personalise learning: they rank, filter and prioritise educational opportunities in ways that shape learner visibility, access and future pathways. Fairness evaluations, however, often remain organised around single protected attributes such as gender, ethnicity, socio-economic status or disability. This paper argues that such evaluations are insufficient for educational contexts because they can obscure harms experienced by learners located at the intersection of multiple demographic positions. It therefore advances intersectional fairness auditing as a human-centred requirement for responsible educational AI. The argument is informed by a motivating adjacent-domain case: a study of a transformer-based job-recommendation system found that a system appearing fair attribute by attribute can still conceal substantial unfairness for intersectional subgroups, and that common mitigation methods do not necessarily improve all subgroups evenly. Drawing on recent scholarship on human-centred AI in education, learning analytics auditing, educational recommendation and intersectional fairness, the paper proposes an audit framework for educational AI. The framework extends existing learning analytics audit work by making subgroup visibility, ranking effects, demographic data governance, contestability and post-deployment monitoring explicit. The paper concludes that educational AI should be judged not only by predictive accuracy or personalisation, but also by whether it supports equitable learner futures across intersecting differences and enables learners and educators to question, interpret and contest AI-mediated guidance.

Keywords: Educational AI, Intersectional Fairness, Learning Analytics, Student Guidance, Human-Centred AI, Audit Framework

1. Introduction

Artificial intelligence is now woven into many educational decision-support environments: dashboards highlight learners at risk, recommenders suggest courses and pathways, admissions systems allocate scarce places, and career-guidance tools mediate how students pursue future opportunities. Although often discussed in terms of efficiency and personalisation, they also shape visibility, access and trajectories by ranking, filtering and prioritising some options over others. This makes fairness a central educational concern. Unfairness can influence whether a learner sees a recommendation, is flagged for support, or is channelled towards particular futures. Baker and Hawn (2022) argue that algorithmic bias in education is not simply a matter of isolated model defects, but a broader issue involving design choices, data practices and institutional contexts. Yet much fairness work in education still treats identity one axis at a time. Gender, ethnicity, socio-economic status or disability may each be analysed separately, but learners do not experience educational systems one attribute at a time. A mature student from a minority ethnic background, a disabled student from a low-income household, or a first-generation international student each occupies an intersectional position. Single-axis audits may therefore make an educational AI system appear fair while obscuring patterned harms for smaller or less visible subgroups.

This paper argues that educational AI needs an explicitly intersectional fairness perspective. It builds this argument through a conceptual, case-informed approach. The motivating case comes from a recent master's thesis on a transformer-based job recommendation system. That work found that a system appearing fair on single protected attributes could still generate intersectional unfairness, and that technical mitigation strategies did not improve outcomes evenly across subgroups (Hossain, 2026). This is not introduced as evidence about real hiring practice, but as an adjacent high-stakes recommender case from which education can learn. The paper makes four contributions: it synthesises recent literature on responsible AI in education, fairness in learning analytics and educational recommendation; reframes many educational AI systems as allocating opportunity rather than only personalising; transfers lessons from an intersectional job-recommendation case to learning analytics and student guidance; and proposes an intersectional fairness audit framework foregrounding subgroup visibility, ranking effects, contestability and institutional accountability. Based on the literature and the motivating case, the paper addresses three research questions: (RQ1) How do recent studies on educational AI, learning analytics and student-guidance systems conceptualise fairness, and to what extent

do they address intersectional harms? (RQ2) What transferable lessons from intersectional bias in job-recommendation systems are most relevant for learning analytics and student guidance in higher education? (RQ3) What should an intersectional fairness audit framework for educational AI include if it is to support human-centred and responsible deployment?

2. Recent Literature Review

Recent work on educational AI has moved beyond simple enthusiasm for automation towards stronger attention to human-centred design, stakeholder participation and hybrid intelligence. Alfredo et al. (2024), in a systematic review of human-centred learning analytics and AI in education, show that stakeholder involvement is increasingly recognised as necessary, but remains uneven across the design process. Their review connects safety, reliability and trustworthiness to when and how users are involved. Topali et al. (2025) reinforce this concern from a more design-focused angle. Reviewing 47 empirical studies on human-centred learning analytics and AI in education, they report that stakeholder participation is often present but concentrated in later stages, while ideation, prototyping and evaluation remain less consistently reported. Their absence during early design can normalise narrow assumptions about what counts as a good outcome or recommendation. Cukurova (2025) pushes the conversation further by framing AI in education through the lens of hybrid intelligence. By treating AI not purely as a tool that replaces educational judgement, he argues for a broader understanding in which AI systems extend, challenge and reshape human cognition and practice. Fairness therefore cannot be reduced to outputs alone, it also concerns how humans interpret, act on and contest those outputs. Hooshyar et al. (2025) identify nine persistent challenges in current AI for education and argue that inadequate stakeholder involvement and weak ethical oversight remain central weaknesses. Their proposal for hybrid human-AI systems resonates strongly with a fairness agenda. If educational AI is hybrid and not autonomous, then fairness must include human oversight, interpretability, contestability and governance, not just model optimisation.

Within learning analytics, auditing and governance are becoming more visible themes. Simbeck (2024) proposes a four-phase audit process for learning analytics systems consisting of delimitation, risk-based definition of audit criteria, auditing and assessment, and monitoring and reassurance. This provides an education-specific bridge between general AI ethics principles and the practical evaluation of deployed systems. Importantly, Simbeck centres the learner perspective and argues that learning analytics should be audited for whether they benefit learners and avoid placing them at risk. Baker and Hawn (2022) also remain foundational for educational fairness because they map algorithmic bias across the educational pipeline. Their analysis makes clear that bias can emerge from the framing of educational problems, the collection and labelling of data, the modelling stage, and institutional implementation. This pipeline view aligns closely with the argument of the motivating job-recommendation thesis we use in this study, where unfairness was treated as a property of the broader model-data interaction and not a simple reflection of a single model component. A more application-specific picture emerges in Cabral et al. (2025), whose systematic review of AI-powered learning analytics dashboards found that most systems focus on predicting academic performance, supporting self-regulated learning and providing teacher-facing insights. The review documents rapid application growth while also pointing to recurring research gaps, including limited evidence about classroom deployment and the long-term effects of dashboard use. Dashboards are not neutral windows onto student behaviour, but they structure attention and determine which learners appear visible, urgent or successful.

The relevance of fairness becomes clearer when educational AI is understood as recommendation and guidance infrastructure. Shailya et al. (2025) examine fairness in language-model-based academic recommendations and identify geographic, demographic and economic biases in university and programme suggestions. Their work is especially useful for this paper because it shows bias at the ranking and visibility level: some institutions, countries and learner profiles become more recommendable than others. Weissburg et al. (2025) reach a related conclusion from another angle. Evaluating large language models as personalised educational teachers, they show significant bias in the generation and selection of educational content across demographic groups, including race, ethnicity, sex, gender, disability, income and national origin. This is not a traditional recommender setting, but it shows that bias in educational AI extends beyond ranking systems to personalised explanation, support and feedback. Fairness concerns also reach AI-supported assessment: Zacharis and Papadakis (2025) examine the validity, reliability and fairness of AI-graded university coursework, indicating that responsible educational AI must meet the same fairness standards whether it ranks, generates or grades. Student guidance is also an emerging site of AI-mediated decision making. A systematic review by Sarmurzin et al. (2026) shows that AI is transforming university career counselling from predictive tools towards interactive generative agents, while simultaneously raising concerns about student agency and the platformisation of

guidance. In a related but different context, Ibrahim et al. (2025) present an equity-aware recommender for university admissions, showing how recommendation and allocation methods are increasingly implicated in high-stakes educational decisions. These studies suggest that educational AI increasingly allocates opportunity instead of merely offering information. When a system influences what learners see, which pathways appear plausible and who receives support or visibility, fairness becomes inseparable from educational design.

Intersectionality offers a language for harms that are not visible when systems analyse social categories separately. Crenshaw's foundational work (Crenshaw, 1989) showed how legal and institutional frameworks can marginalise people whose experiences sit at the intersection of multiple structures of disadvantage. In algorithmic systems, Buolamwini and Gebru (2018) provided an influential demonstration of this problem: in commercial gender-classification systems, darker-skinned females were the most misclassified group, with error rates of up to 34.7%, while the maximum error rate for lighter-skinned males was 0.8%. Their study is outside education, but its lesson is directly relevant, as apparently acceptable aggregate performance can conceal substantial harm for intersectional subgroups. Despite these foundational insights, intersectionality remains underdeveloped in educational AI research. Mirsch et al. (2025) explicitly argue that single-axis approaches to fairness in educational AI obscure the experiences of individuals facing intersecting forms of discrimination. Their intervention is important because it makes visible a recurring blind spot, showing that educational fairness conversations often stop at disaggregating by one variable at a time. The motivating work by Hossain (2026) provides a useful adjacent-domain illustration of why this matters. In a transformer-based job recommendation setting, the study found that fairness assessed on a single protected attribute could mask substantial unfairness across combined demographic subgroups, and that commonly used mitigation methods did not produce balanced improvement across all groups. Although the application domain differs from education, the structural similarity is strong. Both job recommendation and educational AI rank opportunities, define positive outcomes, and shape future trajectories under conditions of partial visibility and unequal data representation. Read across these strands, the reviewed studies engage several distinct conceptions of fairness, but unevenly. Performance fairness, the parity of predictive accuracy or error across groups, is the most established concern and underlies pipeline accounts of bias (Baker and Hawn, 2022) and demonstrations that aggregate accuracy hides subgroup disparity (Buolamwini and Gebru, 2018). Representational fairness, concerning who is visible or well served in data and generated content, appears in work on biased educational content (Weissburg et al., 2025). Exposure or ranking fairness, concerning which options are made visible in top-k results, is beginning to be examined in academic recommendation and allocation (Shailya et al., 2025; Ibrahim et al., 2025). Procedural fairness, namely auditing, oversight and reassurance, is articulated most clearly by Simbeck (2024) and Hooshyar et al. (2025), while participatory fairness, the early involvement of learners and educators, is foregrounded in human-centred and hybrid-intelligence accounts (Alfredo et al., 2024; Topali et al., 2025; Cukurova, 2025). By contrast, contestability, whether learners can question or challenge an AI-mediated outcome, is gestured at but rarely operationalised, and intersectional fairness remains thin, raised directly only by Mirsch et al. (2025). In answer to RQ1, recent educational-AI scholarship conceptualises fairness mainly through performance, participation and emerging audit practice, with exposure fairness newly visible, whereas intersectional harm and contestability are the least developed, which is precisely the gap the proposed framework addresses.

Table 1: Selected recent literature and its relevance to this paper

Study	Domain	Main Relevance to this Paper
Alfredo et al. (2024)	Human-centred LA/AI	Shows that user participation is important but uneven across design stages.
Topali et al. (2025)	Human-centred design	Finds stakeholder involvement is common but often late, with evaluation underreported.
Cukurova (2025)	Hybrid intelligence	Positions educational AI as part of human–AI judgement rather than replacement.
Simbeck (2024)	Learning analytics audit	Provides an education-specific audit model focused on learner risk and reassurance.
Cabral et al. (2025)	AI-powered dashboards	Shows expanding dashboard use in prediction, self-regulation and teacher insight.

Study	Domain	Main Relevance to this Paper
Shailya et al. (2025)	Academic recommendations	Demonstrates geographic, demographic and economic bias in educational recommendation rankings.
Weissburg et al. (2025)	Personalised education	Shows bias in LLM-generated educational content across several demographic groups.
Crenshaw (1989)	Intersectionality	Provides the foundational concept for analysing harms that arise at intersections of social categories.
Buolamwini and Gebru (2018)	Algorithmic fairness	Shows how aggregate system performance can conceal substantial disparities for intersectional subgroups.
Mirsch et al. (2025)	Intersectional fairness	Argues that single-axis fairness approaches obscure intersecting discrimination in educational AI.

3. Approach

This paper adopts a conceptual, case-informed design that combines a targeted integrative review of recent scholarship with reflective transfer from an adjacent-domain thesis case. It is not a full systematic review. Instead, we deliberately gathered the studies needed to make sense of intersectional fairness in educational AI, and we explain how we did so, so that readers can judge our choices for themselves. Sources were identified through Scopus, Web of Science, the ACM Digital Library, IEEE Xplore and Google Scholar, supplemented by hand-searching recent proceedings (ACL/NAACL, IJCNLP, PMLR) and the journals Computers and Education: Artificial Intelligence, British Journal of Educational Technology and AI and Ethics. Search terms combined fairness vocabulary (fairness, algorithmic bias, intersectional) with educational and system vocabulary (learning analytics, educational AI, student guidance, academic recommendation, audit). Studies were included if they were peer-reviewed, published between 2024 and 2026, written in English, and addressed at least one of four strands: human-centred and hybrid AI in education, fairness and auditing in learning analytics, educational recommendation and guidance, or intersectional fairness. Work outside these strands and non-peer-reviewed material were excluded, with the single, explicitly flagged exception of the motivating thesis, which is used illustratively. Pre-2024 work was excluded except for foundational sources on intersectionality and algorithmic bias (Crenshaw, 1989; Buolamwini and Gebru, 2018; Baker and Hawn, 2022), retained because they remain conceptually indispensable. The 2024 to 2026 window was chosen because generative and transformer-based models were rapidly integrated into educational AI during this period, and because the most relevant recent reviews and the motivating case are contemporaneous, allowing the transfer of lessons to be situated against current debates rather than an outdated baseline. This recency window applies only to evidence about how current systems behave. The framework's ethical and governance foundations draw on established learning-analytics-ethics scholarship that predates it and remains authoritative, including early analyses of the ethical dilemmas of learning analytics (Slade and Prinsloo, 2013) and checklist-based approaches to trusted, privacy-aware analytics (Drachsler and Greller, 2016). Selected studies were analysed thematically: for each, the domain, the system type and, central to RQ1, how it conceptualises and operationalises fairness were recorded. The motivating case is a master's thesis (Hossain, 2026), openly available under a Creative Commons licence and examined at the University of Koblenz, which audited a transformer-based job-recommendation pipeline. The system matched applicant resumes to O*NET occupational profiles using Sentence-BERT embeddings (all-mpnet-base-v2) and cosine similarity, evaluated as a top-k recommender. Fairness was assessed with three metrics, namely demographic parity, conditional demographic parity and equal opportunity, first for single attributes (gender, race) and then for intersectional subgroups combining gender, race, age and experience level, using a publicly available recruitment dataset of mixed synthetic and real application data. The study found that a system appearing fair attribute by attribute concealed substantial disparities for smaller intersectional subgroups, that two layers of resampling improved fairness unevenly and traded accuracy against parity, and that aggregate top-k fairness masked persistent ranking disadvantage for some groups. The thesis is used here as an illustrative, adjacent-domain case that motivates the framework, not as independent empirical evidence about educational systems. No fairness scores are reproduced, and no claim about education rests on it as if externally validated. The transfer from job recommendation to education is therefore analytical, not literal. Both domains use data-driven systems to rank, filter and narrate future opportunities under conditions of unequal representation, which is what makes the analogy useful. But the two differ, and those differences set limits.

Education carries its own institutional and legal duties, teaching relationships, developmental goals, and expectations about student choice, so the job-recommendation lessons are best treated as questions to ask about educational AI, not answers to assume.

4. From Job Recommendation to Educational AI

4.1 Educational AI as Opportunity Allocation

A central argument of this paper is that many educational AI systems should be understood as opportunity-allocation systems. This section addresses RQ2 by identifying the transferable lessons from intersectional bias in job-recommendation systems that are most relevant for learning analytics and student guidance. They allocate not scarce seats but visibility, salience, intervention and perceived possibility. In education, the consequences of ranking and filtering are normative. A learner may never encounter the course, placement, institution or support pathway that would have mattered most because the system ranked alternatives differently. In both domains, top-k exposure matters, outcome definitions matter, and apparently small ranking effects can have long-term trajectory consequences.

4.2 Why Single-Attribute Fairness is Insufficient

The second argument is that single-attribute fairness analysis is inadequate for educational AI. If a system is audited only for gender parity, it may still disadvantage older women, disabled women, or women from specific ethnic groups. If it is audited only for socio-economic status, it may still reproduce harms for low-income international students or mature learners from minority backgrounds. The point is not merely combinatorial complexity. It is that social disadvantage is often produced at intersections, not along one line at a time. Mirsch et al. (2025) make this point directly for educational AI. The study by Hossain (2026) illustrates how this arises within a recommendation pipeline: when fairness was examined attribute by attribute, the system seemed relatively fair, but when subgroup combinations were evaluated, substantial disparities became visible. This lesson translates well to learning analytics and student guidance because these systems also work with heterogeneous populations in which smaller subgroups are easily drowned out by majority patterns. Institutions are especially prone to missing this because they treat AI as support rather than selection infrastructure, yet support systems still stratify opportunity by shaping who is noticed and encouraged.

4.3 Why Technical Fixes Alone are Insufficient

A third lesson from the motivating case is that bias mitigation should not be treated as a one-step technical remedy. In Hossain (2026), preprocessing interventions improved some fairness outcomes but did not improve all subgroups evenly. Practitioners should therefore not deploy a mitigation method and assume the problem is solved. A mitigation technique can reduce disparity for a visible aggregate group while shifting error, exposure or risk onto a smaller intersectional subgroup. Hooshyar et al. (2025) emphasise inadequate ethical oversight and stakeholder involvement, while Simbeck (2024) argues for ongoing monitoring and reassurance and not just one-off checks. Together these works suggest that responsible educational AI requires socio-technical governance. Fairness must be revisited as systems are updated, populations change and uses drift.

5. An Intersectional Fairness Audit Framework for Educational AI

Drawing on the literature and the motivating case, this paper proposes an intersectional fairness audit framework for educational AI systems used in learning analytics, student guidance, academic recommendation and related decision-support settings. In doing so, the section responds to RQ3 by specifying what an intersectional fairness audit framework for educational AI should include to support human-centred and responsible deployment. The framework builds on Simbeck's (2024) learning analytics audit model and earlier trusted-analytics checklists (Drachsler and Greller, 2016), but it differs in three ways. First, it makes intersectional subgroup visibility an explicit audit object and not just treating fairness as a general risk category. Second, it is ranking-sensitive, recognising that many educational AI systems distribute visibility and exposure rather than only binary decisions. Third, it connects technical evaluation to human-centred practices of explanation, contestability and institutional monitoring. The framework is intended as a practical checklist for design, procurement, deployment and post-deployment review. It does not prescribe a single fairness metric, instead, it asks institutions to make explicit what the system does, who may be affected, what outcomes and rankings matter, and how learners can challenge or reinterpret AI-mediated guidance.

Table 2: Proposed intersectional fairness audit framework

Audit Dimension	Core Question	Lifecycle Point	Illustrative Practices
Context and stakes	What educational decision or guidance function is the system shaping?	Design and procurement	Define the educational use, affected learners and possible trajectory effects before evaluation.
Data representation and demographic governance	Who is visible in the data, and who may be missing, proxied or flattened by aggregation?	Design and data preparation	Review coverage, proxies, subgroup size, data minimisation, consent and small-cell re-identification risks.
Outcome and ranking definition	What counts as a positive outcome, and how does ranking influence exposure?	Design and evaluation	Audit top-k visibility, recommendation depth and which options are systematically foregrounded.
Intersectional evaluation	Do subgroup combinations reveal harms that single-attribute checks miss?	Testing and deployment	Compare aggregate results with ethically appropriate subgroup combinations and report uncertainty for small groups.
Transparency and explanation	Can learners and educators understand the system's purpose, limits and logic?	Deployment	Provide intelligible explanations, model cards or user-facing guidance on system scope and uncertainty.
Human oversight and contestability	Can outputs be reviewed, overridden or contested?	Deployment and use	Build teacher review processes, learner feedback channels and routes for appeal or correction.
Governance and monitoring	Who is responsible for fairness after deployment?	Post-deployment	Assign institutional ownership, document updates and run periodic audits across cohorts and contexts.

To show what the framework changes in practice, consider a recommender that suggests optional first-year modules, where some modules feed into competitive, employability-linked progression routes. A conventional single-axis audit would check overall click-through or completion and perhaps compare recommendations for men and women, conclude that parity holds, and stop. The framework reorganises this audit around educational consequence. Under context and stakes, the team first establishes that the system is not merely informational: by shaping which routes appear plausible, it allocates opportunity. Under data representation, it examines who is present in the training data and at what granularity, and here the motivating case is instructive, because the intersectional cells most exposed to harm, such as the smallest subgroup in Hossain (2026) with only a few dozen records, are exactly those a majority-driven model represents least reliably. This raises an ethical and statistical limit: only subgroup combinations large enough to support a defensible claim should be reported, small-cell counts must be disclosed and not silently pooled, and re-identification risk constrains how finely groups may be cut. Under outcome and ranking definition and intersectional evaluation, the audit inspects top-k exposure across combined groups rather than aggregate rates, since the case showed that aggregate fairness can improve while relative ranking still disadvantages particular subgroups, and that resampling can buy parity at the cost of accuracy, so both must be reported together with uncertainty for small groups. Finally, transparency, human oversight and contestability, and governance require that learners receive intelligible explanations that counter automation bias, that a teacher can override a steer and a learner can challenge it through a defined route, and that a named owner re-audits as cohorts change. A single-axis audit would have passed this system as fair. The intersectional audit instead reveals who stays invisible, where ranking quietly widens gaps in opportunity, and how learners can challenge the results.

6. Discussion

The proposed framework has practical implications for several educational contexts. Returning to RQ1, the literature reviewed in Section 2 conceptualises fairness in educational AI primarily through human-centred design, learning analytics auditing and bias in recommendation or generation, but the reviewed studies address intersectional harms only partially, with most contributions still organised around single protected attributes. The framework proposed here addresses that gap. Across learning analytics dashboards, student-guidance

systems and admissions tools, it shifts attention from overall model performance to whose behaviour is legible, whose needs become visible, how recommendations may narrow aspiration, and the distributive consequences of ranking and allocation.

This paper has limitations. First, it makes a conceptual contribution. It does not report new empirical findings, and the review covers a focused set of recent studies. The framework therefore needs empirical testing in specific educational settings such as learning analytics dashboards, academic guidance chatbots and institutional career services. Future research could operationalise the framework through case studies, co-design workshops, or audits of existing university systems. A further step is to examine how ethical and institutional constraints affect intersectional evaluation, especially where demographic data are incomplete or sensitive. The current educational AI conversation is rapidly expanding, but it risks reproducing a familiar pattern in which fairness language is adopted while single-axis evaluations and weak governance remain in place. Intersectional auditing provides a clearer basis for responsible practice because it asks institutions to identify who benefits, who remains invisible and who can contest the system's effects.

7. Conclusion

Educational AI increasingly ranks, filters and guides learners across key moments of academic and professional development. As these systems expand, fairness must be treated as an educational and institutional concern, not merely a technical feature. Recent literature shows growing awareness of bias and responsibility, but intersectionality remains insufficiently integrated into practice. In response, the paper proposed an intersectional fairness audit framework centred on context, data representation and governance, ranking, intersectional evaluation, transparency, human oversight and monitoring. Educational AI should be judged not only by whether it predicts well or personalises efficiently, but by whether it supports equitable participation across intersecting learner identities.

Acknowledgements

The authors received no specific funding for this work.

Ethics Declaration

This paper is a conceptual study drawing on published literature. It reports no new human-participant research and no new collection of personal data. The motivating case (Hossain, 2026) is a publicly available, independently examined master's thesis released under a Creative Commons licence. Only its published findings are used, and no non-public or author-held personal data were drawn upon. Ethical clearance was therefore not required for the research reported here.

AI Declaration

An AI language model was used solely for language revision. All substantive arguments, interpretations, source selections, and final wording were reviewed and approved by the authors, who take full responsibility for the final manuscript.

References

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D. and Martinez-Maldonado, R. (2024) "Human-centred learning analytics and AI in education: A systematic literature review", *Computers and Education: Artificial Intelligence*, 6, 100215. doi:10.1016/j.caeai.2024.100215.

Baker, R.S. and Hawn, A. (2022) "Algorithmic bias in education", *International Journal of Artificial Intelligence in Education*, 32, pp. 1052–1092. doi:10.1007/s40593-021-00285-9.

Buolamwini, J. and Gebru, T. (2018) "Gender shades: Intersectional accuracy disparities in commercial gender classification", *Proceedings of Machine Learning Research*, 81, pp. 77–91.

Cabral, L., Pinto, R. and Gonçalves, G. (2025) "AI-powered learning analytics dashboards: a systematic review of applications, techniques, and research gaps", *Discover Education*, 4, 525. doi:10.1007/s44217-025-00964-y.

Crenshaw, K. W. (1989) "Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics", *University of Chicago Legal Forum*, 1989, pp. 139–167.

Cukurova, M. (2025) "The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence", *British Journal of Educational Technology*, 56(2), pp. 469–488. doi:10.1111/bjet.13514.

Drachsler, H. and Greller, W. (2016) “Privacy and analytics: it’s a DELICATE issue. A checklist for trusted learning analytics”, in *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK ’16)*, Edinburgh, UK, pp. 89–98. doi:10.1145/2883851.2883893.

Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D. and Azevedo, R. (2025) “Towards responsible AI for education: Hybrid human-AI to confront the elephant in the room”, *Computers and Education: Artificial Intelligence*, 9, 100524. doi:10.1016/j.caeai.2025.100524.

Hossain, R. (2026) *Exploring Intersectional Fairness and Bias Mitigation Techniques in Job Recommendation*. Master’s thesis. University of Koblenz. Available at: https://opus4.kobv.de/opus4-uni-koblenz/frontdoor/index/index/docId/2638 (CC BY-SA 4.0). doi:10.82549/opus4-2638.

Ibrahim, A., Alarood, A. and Alsolami, E. (2025) “An equity aware recommender system for university admissions balancing operational constraints and strategic objectives”, *Scientific Reports*, 15, 39756. doi:10.1038/s41598-025-23116-6.

Mirsch, M., Strube, J. and Leicht-Scholten, C. (2025) “Invisible inequalities - intersectional fairness in educational AI”, *Proceedings of Machine Learning Research*, 294, pp. 403–409.

Sarmurzin, Y., Azhenov, A., Kudysheva, A., Ualikhanova, B., Saparbekova, I. and Besbayev, A. (2026) “Implementation of AI in career counselling for university students: a systematic review”, *Frontiers in Education*, 11, 1787689. doi:10.3389/feduc.2026.1787689.

Shailya, K., Mishra, A.K., Krishnan, G.S. and Ravindran, B. (2025) “Where should I study? Biased language models decide! Evaluating fairness in LMs for academic recommendations”, in *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, Mumbai, India, pp. 2291–2317. doi:10.18653/v1/2025.findings-ijcnlp.141.

Simbeck, K. (2024) “They shall be fair, transparent, and robust: auditing learning analytics systems”, *AI and Ethics*, 4, pp. 555–571. doi:10.1007/s43681-023-00292-7.

Slade, S. and Prinsloo, P. (2013) “Learning analytics: ethical issues and dilemmas”, *American Behavioral Scientist*, 57(10), pp. 1510–1529. doi:10.1177/0002764213479366.

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M.J., Er, E., Khalil, M. and Akçapınar, G. (2025) “Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review”, *Behaviour and Information Technology*, 44(5), pp. 1071–1098. doi:10.1080/0144929X.2024.2345295.

Weissburg, I., Anand, S., Levy, S. and Jeong, H. (2025) “LLMs are biased teachers: Evaluating LLM bias in personalized education”, in *Findings of the Association for Computational Linguistics: NAACL 2025*, Albuquerque, New Mexico, pp. 5665–5713. doi:10.18653/v1/2025.findings-naacl.314.

Zacharis, G. and Papadakis, S. (2025) “Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment”, *Educational Process: International Journal*, 19, e2025591. doi:10.22521/edupij.2025.19.591.

Embedding Responsible AI in a Technical Machine Learning Course

Stefania Zourlidou and Frank Hopfgartner
Institute for Web Science and Technologies, University of Koblenz, Koblenz, Germany

zourlidou@uni-koblenz.de
hopfgartner@uni-koblenz.de

Abstract: Machine learning (ML) courses in higher education usually emphasise mathematical foundations and algorithmic methods, while ethical and societal questions are often treated as separate or optional topics. This paper reports a small-scale practice-based evaluation of a blended design that embedded responsible AI within a Master's-level ML and Data Mining course at a German university in winter semester 2025/26. The design combined fifteen on-campus lectures on core ML topics, including one lecture on fairness as a technical topic, fifteen optional weekly assignments discussed in tutorials, and an eight-module microlearning course on AI literacy and responsible AI. Completion of at least half of the microlearning and a mandatory practical assignment were required for exam eligibility. The study analyses an anonymous end-of-course survey, using only responses from students who explicitly consented to research use (n=35). Student perceptions were generally positive: 78% reported a good balance between theory and practice, 75% found the weekly assignments important for understanding, and 59% judged the microlearning to fit well with the on-campus lectures. For responsible AI, around two thirds reported greater confidence in identifying fairness issues, feeling better prepared to discuss ethics professionally, or perceiving ethics as integrated rather than added on. 75% intended to consider fairness and ethical implications in future ML projects, with no respondent disagreeing. Open responses suggested perceived transfer, with students naming hiring systems, facial recognition, student-risk prediction and credit or loan decision-making, and often using technical vocabulary such as training data, optimisation objectives and evaluation metrics. The paper does not claim causal impact, and findings should be interpreted in light of the small voluntary sample, the single-cohort single-institution setting, and the possibility of social-desirability bias on an end-of-course survey on ethics. Instead, it offers descriptive evidence that a lightweight microlearning layer, combined with a technical fairness lecture, can be a feasible way to connect technical ML learning with ethical reflection in practice. It also identifies revisions for the next iteration: reduce workload tension, add explicit ethical-reflection prompts to assignments, and include more practical coding work.

Keywords: Machine Learning Education, Responsible AI, AI Literacy, Microlearning, Blended Learning, Higher Education

1. Introduction

Machine learning (ML) is now a core component of computer science and data science curricula in higher education. Master's-level ML courses often prioritise mathematical foundations and algorithms such as regression, classification, clustering and neural networks. Ethical and societal questions may be acknowledged as important, but are still frequently relegated to a guest lecture, optional reading list or separate elective. This separates technical competence from ethical awareness. The rapid deployment of ML in consequential domains such as hiring, lending, healthcare and education makes this separation difficult to defend. Documented intersectional accuracy disparities in commercial gender classification (Buolamwini and Gebru, 2018), algorithmic bias in educational ML systems (Baker and Hawn, 2022) and harms in public-service decision-making (Eubanks, 2018) show that fairness, transparency and accountability are not optional add-ons. Policy frameworks such as the EU AI Act (European Parliament and Council of the European Union, 2024), together with AI-literacy frameworks (Long and Magerko, 2020; Ng et al, 2021), increase the importance of preparing graduates to reason about these properties as well as model accuracy. Computer science education likewise calls for ethics to be integrated into technical programmes (Grosz et al, 2019; Saltz et al, 2019; Ferrer et al, 2021). A parallel development in e-learning offers a delivery vehicle. Blended learning combines online and face-to-face elements (Garrison and Vaughan, 2008), while microlearning uses short, focused online units (Hug, 2005; Leong et al, 2021). In this design, microlearning was used to add case studies, societal context and AI literacy while keeping the technical lecture sequence intact. Generative AI has further intensified discussion of responsible AI education (Kasneci et al, 2023).

This paper reports on a blended course design that takes these developments seriously. In the winter semester 2025/26, a Master's-level ML and Data Mining course at a German university combined three components: (i) fifteen on-campus lectures covering technical ML topics from optimisation to neural networks, including a dedicated lecture on fairness, (ii) fifteen optional weekly assignments combining knowledge questions, practical problems and programming tasks, discussed in weekly tutorials, and (iii) an eight-module microlearning course on AI literacy and responsible AI. Partial completion of the microlearning was required for exam eligibility, alongside a mandatory practical assignment. We report on three questions: how students perceive the three-

part design, how they report their framing of ML and confidence in identifying fairness issues, and what ethical examples they describe. The contribution is descriptive and practice-based, not causal.

2. Related Work

Ethics has been recognised in formal computing-curriculum guidance (Joint Task Force on Computing Curricula, 2013), but an isolated "ethics module" is widely seen as insufficient. Grosz et al (2019) describe Harvard's embedded ethics model, in which short ethics sessions are co-taught within technical CS courses. Other authors argue that judgements about data, evaluation metrics and deployment cannot be separated from technical design (Saltz et al, 2019; Ferrer et al, 2021; Borenstein and Howard, 2021). In ML, the fair-ML literature provides quantitative fairness definitions, bias sources and mitigation approaches that can be taught as mathematical and engineering content (Mehrabi et al, 2021), and course reports show how ethics activities can be integrated into programming and CS classrooms (Fiesler et al, 2021; Skirpan et al, 2018).

Blended learning combines online and face-to-face elements within one course design (Garrison and Vaughan, 2008). Microlearning describes short, focused online units, typically a few minutes of video followed by a brief assessment (Hug, 2005; Leong et al, 2021). Such units can support focused objectives when aligned with cognitive-load principles (Sweller, van Merriënboer and Paas, 2019), and they offer a low-friction way to add content layers to existing courses (Sankaranarayanan et al, 2023). AI literacy - the ability to critically evaluate, communicate and collaborate with, and use AI (Long and Magerko, 2020) - has emerged as a distinct competence, with later work refining its operationalisation in formal education (Ng et al, 2021; Casal-Otero et al, 2023). General-public AI literacy does not substitute for the technical literacy needed by CS students, and this design attempts to bridge that gap.

Recent work has moved from broad calls for responsible-AI education to concrete curricular interventions. Aler Tubella, Mora-Cantallops and Nieves (2024) identify competencies and implementation challenges for responsible-AI teaching in higher education. Weichert and Eldardiry (2025) evaluate a two-lecture AI-policy module in a graduate introductory machine-learning course, while Jayapal, Hammerman and Chaspari (2025) report a pilot AI-ethics curriculum embedded in graduate-level machine learning. Although the two curricular interventions focus mainly on discrete modules, the present study examines a broader blended configuration combining a technical fairness lecture with sustained microlearning.

Evaluating educational designs in authentic course settings is methodologically constrained: experimental designs are rarely feasible because of small cohorts, ethical concerns about withholding instruction, and the entanglement of multiple course components. Practice-based and design-based approaches (Reeves, 2006; Anderson and Shattuck, 2012) instead emphasise iterative description of what was done and how it was perceived. We adopt this stance by describing a concrete design and reporting what students said about it, while being explicit about the limits of inference.

3. Course Design

3.1 Course Context

The course is the Master's-level Machine Learning and Data Mining course taught in winter semester 2025/26 at a German university. Its main target programme is a data-science-oriented Master's programme. It also serves students from other Master's programmes, including an applied mathematical modelling programme. Prerequisites include undergraduate linear algebra, calculus, probability and statistics, and basic Python. Assessment is by a written examination. Exam eligibility depends on two requirements, which are passing a mandatory practical assignment and completing at least half of the microlearning course described below.

3.2 Component A: Lectures

Fifteen 90-minute on-campus lectures span the technical ML curriculum: introduction, mathematical optimisation (two lectures), linear regression, linear two-class classification (two lectures), linear multi-class classification, clustering, feature engineering, non-linear feature engineering, dimensionality reduction and principal components, tree-based learners, optimising ML models, fairness, considerations and limits in ML, and neural networks. Lecture 14 ("Fairness, considerations and limits in ML") is positioned within the technical sequence and not as a separate ethics day and treats fairness as a quantitative topic: group-fairness metrics, sources of bias in training data, mitigation strategies, and limitations.

3.3 Component B: Assignments and Tutorials

Fifteen weekly assignments accompany the lectures. They are optional in the sense that they are not marked individually, but students are recommended to complete at least 50% of each. Each assignment combines knowledge questions, conceptual problems, and small programming tasks. Solutions are discussed in weekly tutorials led by teaching assistants. A separate mandatory practical assignment, distinct from the optional weekly ones, must be passed at 50% to qualify for the exam.

3.4 Component C: Microlearning Course "Machine Learning Made Easy"

Component C was a microlearning course hosted on the institution's learning management system. It comprised a pre-course questionnaire, eight content modules, and a post-course assessment. Each module paired a short video with an assessment (about 30-40 minutes' work), for a total of about six to seven hours. The modules were: (1) Everyday Intelligence: What is AI, (2) Data Matters: The Fuel of AI, (3) How Machines Learn from Data, (4) Teaching the Machine, (5) When Machines Judge, (6) Personalisation and Evaluation, (7) Truth, Trust, and Misinformation, and (8) Responsible AI. Content emphasised AI literacy and the societal dimensions of ML, i.e., how data shapes models, where bias enters, how recommendation systems work, how misinformation can be amplified, and what responsible AI practice entails. Completion of at least 50% of the modules was required for exam eligibility.

3.5 Design Rationale

The design uses two entry points to responsible AI. Lecture 14 treats fairness as part of technical modelling. The microlearning adds AI literacy and societal context without crowding out the technical curriculum. The assignments and tutorials provide practice and feedback. Together, the components form a lightweight model for embedding responsible AI in a technical ML course.

4. Methodology

An end-of-course questionnaire comprising 52 numbered items plus an initial research-consent item was administered through the institutional learning management system from late January to mid-March 2026. Responses should therefore be read as end-of-course perceptions and not as a single post-exam measurement. The questionnaire covered demographics, engagement with lectures, tutorials, assignments and microlearning, generative-AI use and perceptions, course design, responsible-AI learning, overall satisfaction, and four open-ended items. Likert items used a five-point scale (Strongly disagree to Strongly agree). The survey was voluntary and anonymous. As the initial item, respondents were asked whether they consented to research use and possible publication of their fully anonymised responses. Only respondents who explicitly consented are analysed here. The survey received 51 valid submissions: 35 consented, 8 did not consent, and 8 left the consent item blank. All analyses are restricted to the consenting subset (n=35). Missing or invalid export values were treated as missing, and percentages were calculated item-wise from available consenting responses. Consequently, valid n varies between items. No direct identifiers were collected and demographic reporting is aggregated.

Of the 35 consenting respondents, 26 were enrolled in the main data-science-oriented Master's programme, 3 in an applied mathematics/modelling programme, 5 in other Master's programmes, and 1 did not answer the programme item. Twenty-five were in their first or second semester. Eighteen had taken at least one prior AI or ML course, fifteen had not, and two did not answer. The modal age category was 23-25 (n=15), followed by 26-30 (n=11), 20-22 (n=6), 31+ (n=2), and one non-response. Twenty-three identified as men, eight as women, three preferred not to self-describe or not to say, and one did not answer. As with any voluntary end-of-course survey, the consenting subset is likely affected by self-selection: students who disengaged from the course or held strongly negative views may be under-represented, and we do not have administrative data to compare the subset against the full active cohort.

Quantitative analysis is descriptive. For each Likert item we report the percentage agreeing (Agree and Strongly agree), neutral, and disagreeing on a 1-5 scale. Percentages may not sum to 100 because of rounding. Where means are reported, they are used as descriptive shorthand only. Likert responses are treated as ordinal throughout. We do not perform inferential tests, both because the descriptive question is the right one for practice-based evaluation at this scale and because the sample is too small for robust subgroup inference. Q40 ("Before this course, I mainly saw ML as a technical field") and Q41 ("After this course, I see ML as a socio-

technical field") are presented side by side as a retrospective comparison of two differently worded constructs. No within-person pre/post effect is inferred.

Qualitative analysis used a lightweight inductive coding approach informed by Braun and Clarke (2006) for Q47 and Q52. The first author coded responses in two passes: an initial pass identified categories close to students' wording, and a second pass consolidated them into the domain and workload/integration groupings reported in Section 5.4. Responses such as "no", "none" and "N/A" were classified as non-substantive and were not included in the thematic categories. No inter-rater coding was undertaken.

5. Findings

5.1 Engagement with the Three Components

Self-reported engagement is summarised in Figure 1. Microlearning completion was concentrated at the top end: 22 of 34 respondents (65%) completed all eight modules and 26 of 34 (76%) completed at least six. Two reported 0-2 modules, suggesting that not all survey respondents had completed the exam-eligibility requirement at the time of response or intended to sit the exam. Weekly assignments showed a wider spread: 11 of 32 respondents completed 12-15, 5 completed 8-11, 11 completed 4-7, and 5 completed 1-3. Weekly hours ranged from 2-4 hours (10 respondents) to 5-7 (13), 8-10 (5), and more than 10 (6). All 33 respondents who answered the mandatory-assignment question reported submitting it (94% of the consenting sample). Lecture and tutorial attendance were mixed.

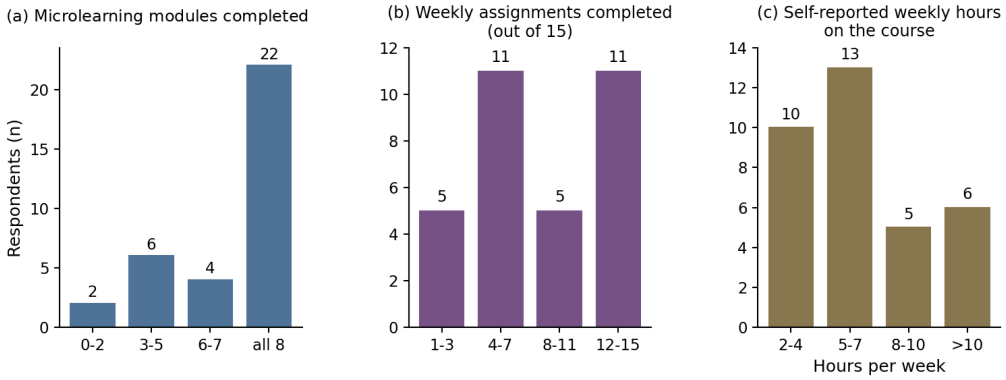


Figure 1: Self-reported engagement with the three course components (consenting respondents, n=35, valid responses per item shown in panels). The microlearning, gated by a 50% exam-eligibility threshold, generated high completion rates, with most respondents completing six or more modules. The optional weekly assignments show the diverse engagement profile typical of non-mandatory work.

5.2 Perceptions of the Course Design (RQ1)

Likert distributions for nine course-design items are shown in Figure 2. The design is broadly endorsed. 78% agreed or strongly agreed that the course provided a good balance between theory and practice (Q26). 75% agreed that the optional weekly assignments were important for their understanding (Q31), the second-highest endorsement of any design item, and 69% agreed that the practical components connected theory and practice (Q28). The microlearning fit item was positive but less emphatic: 59% agreed that the microlearning fitted well with the on-campus lectures (Q33). The tutorials received more measured endorsement (53% agree, 44% neutral, Q32), consistent with the variable attendance reported in Section 5.1.

Two additional items addressed balance directly: Q29 ("too theoretical") split the cohort (44% agree, 41% neutral, 16% disagree), while Q30 ("too tool-focused") was disagreed with by 41% and agreed with by 16%, suggesting that imbalance was felt mainly in the theory direction. Workload (Q34) had the highest disagreement rate among the positively framed design items: 56% agreed it was manageable but 12% disagreed, with qualitative feedback linking this to the lecture-plus-microlearning combination. Confidence outcomes were positive: 62% agreed their confidence in applying ML to new datasets had increased (Q35) and 69% agreed their confidence in explaining ML to non-experts had increased (Q36), with no disagreement on Q35 and only 3% on Q36.

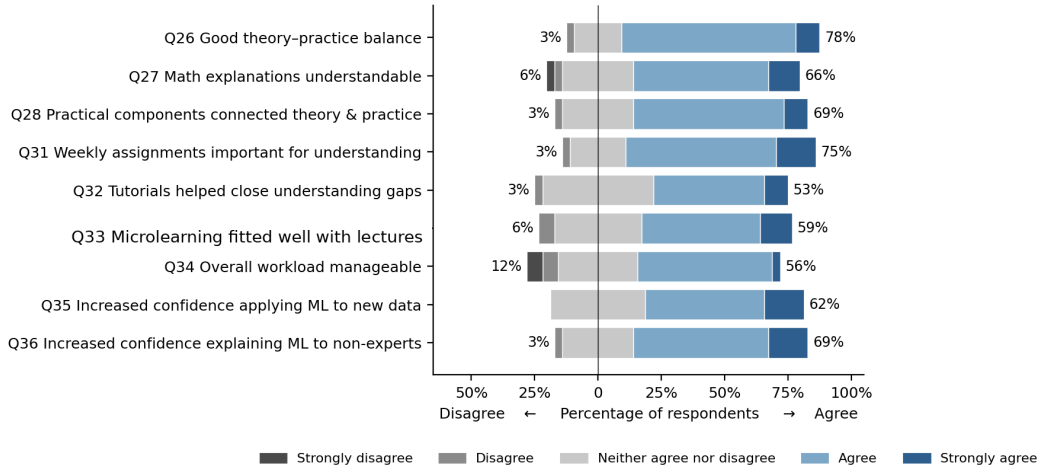


Figure 2: Student perceptions of the blended course design. Percentages on the right show agree and strongly agree. Percentages on the left show strongly disagree and disagree when not negligible (n=32 valid responses per item).

5.3 Perceptions of Responsible AI Learning (RQ2)

Likert distributions for the seven core responsible-AI items are shown in Figure 3. The retrospective comparison of Q40 and Q41 is presented separately in Figure 4. Three observations stand out.

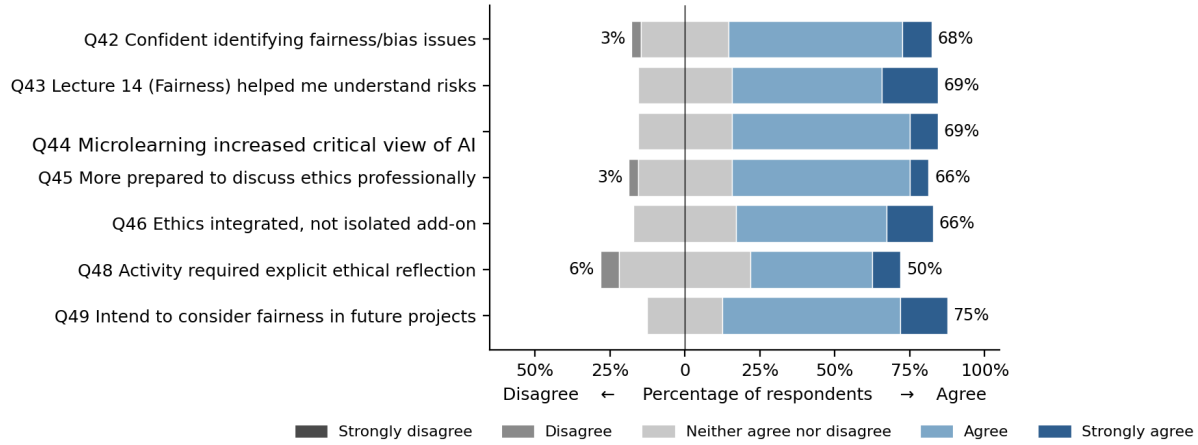


Figure 3: Student perceptions of responsible-AI learning. n=31-32 valid responses per item.

First, the strongest single result in the survey is the future-intention item (Q49): 75% of respondents intend to actively consider fairness and ethical implications in future ML projects, with not a single respondent disagreeing. While stated intentions are not the same as future behaviour, the absence of any disagreement on a voluntary survey suggests that the ethical framing was broadly accepted rather than resisted.

Second, the two responsible-AI components received comparably high endorsement. 69% agreed that Lecture 14 helped them understand ethical risks (Q43), and 69% agreed that the microlearning made them more critical of everyday AI systems (Q44). Around two thirds agreed they were more confident identifying fairness or bias issues (Q42, 68%), better prepared to discuss ethics professionally (Q45, 66%), and that ethics felt integrated rather than isolated (Q46, 66%).

Third, the weakest item in the ethics block is Q48 ("at least one assignment or activity required me to explicitly reflect on ethical or fairness aspects of ML"): only 50% agreed, with 44% neutral. This is a useful diagnostic. Although Lecture 14 and the microlearning carry the ethical content, the weekly assignments do not currently include an explicit reflection prompt. Half the cohort therefore does not perceive an activity that required ethical reflection, even though many of those same respondents endorse Q46 and Q49. The disconnect suggests a concrete revision target: add a short ethical-reflection prompt to one or two weekly assignments.

Figure 4 reports the two retrospective ML-framing items descriptively. Before the course, 53% of respondents agreed they had mainly seen ML as a purely technical field, 31% were neutral and 16% disagreed. After the course, 53% agreed they now see ML as a socio-technical field, 41% were neutral and 6% disagreed. Because Q40 and Q41 assess differently worded constructs, no within-person pre/post change is inferred; the distributions are reported as complementary indicators of students' perceived framing of ML. Both items are also vulnerable to recall bias and socially desirable responding on an end-of-course survey on ethics. The qualitative responses in Section 5.4 provide a separate line of descriptive evidence.

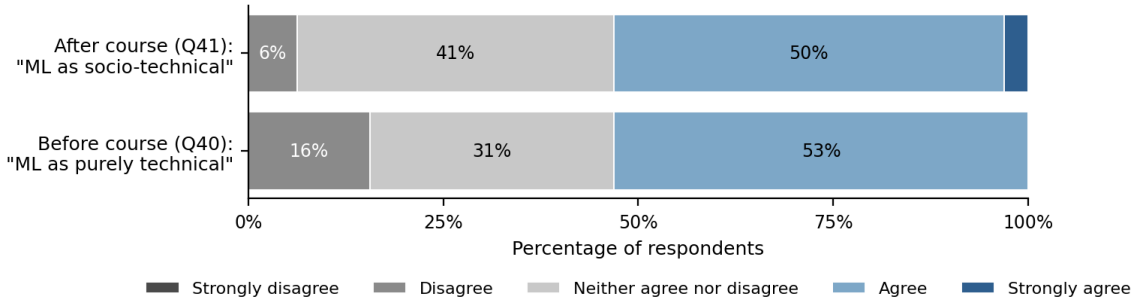


Figure 4: Retrospective comparison of two differently worded ML-framing items (Q40 and Q41), n=32 valid responses per item.

5.4 Qualitative Evidence of Perceived Transfer (RQ3)

Q47 asked students to describe one example, from the course or from their own life, where they now see fairness or ethical issues in an ML or AI system. This open-ended question was used to examine whether students could transfer the responsible-AI themes from the course to concrete examples. Thirteen respondents provided a non-blank entry. Nine entries contained substantive examples, one was unrelated, and three were non-substantive responses or declines, such as "No" and "N/A". Only the nine substantive responses were included in the thematic categorisation. They clustered into four domains: hiring and CV screening (five responses), student-risk prediction in higher education (two), facial recognition (one), and credit or loan decision-making (one). Students described, for example, how historical hiring bias can enter training data and how attendance or platform-activity features can disadvantage students with jobs, caregiving responsibilities or unreliable internet access. Three features are worth noting: first, several examples were linked to course components, including the tutorial hiring case and the microlearning module on personalisation and judgement. Second, the facial-recognition example suggests transfer to public discourse. Third, several responses used technical language from the course, such as training data, optimisation objectives, evaluation metrics, fairness and transparency.

Q52, the open feedback item, drew thirteen responses, seven of them substantive. Of these, five mentioned workload as too high, with three of the five linking this partly to the lecture-plus-microlearning combination. One called for more practical coding work, in line with Q37, and one clearly praised the integration of ethical content. Negative feedback concentrated on workload and the desire for more practical coding work.

5.5 Overall Satisfaction

Overall satisfaction with what was learned was high: 72% agreed or strongly agreed (Q50, mean 3.97 on a 1-5 scale), and 75% would recommend the combination of on-campus lectures, assignments, and microlearning to other students (Q51, mean 4.09). The recommendation item received the highest mean of any Likert item in the survey, with twelve respondents choosing "strongly agree".

6. Discussion

Three results form the core of what was perceived as working. First, 66% agree that ethics felt integrated rather than added on, with zero disagreement. Second, the future-intention finding (75% intend to consider fairness in future projects; 0% disagreement) suggests broad acceptance of the ethical framing, while not demonstrating future behaviour. Third, the qualitative examples (hiring systems, student-risk prediction, facial recognition, and credit or loan decisions) show perceived transfer beyond the specific course examples.

Workload is the clearest design problem. Although a majority report the workload as manageable, 12% disagree (the highest rate among the positively framed course-design items), and qualitative responses attribute the issue partly to the lecture-plus-microlearning combination. The microlearning enabled AI-literacy and societal coverage but added load on top of fifteen lectures. Possible mitigations include reducing the microlearning module count, aligning modules more tightly with the lecture schedule, or making one or two modules optional. A second issue is the perception that no activity required explicit ethical reflection (Q48: only 50% agree). The current design relies on Lecture 14, the microlearning and tutorial discussion to carry ethical reflection. The optional assignments do not include reflection prompts. Adding a short prompt, for example, "Identify one fairness consideration in your solution", to one or two assignments is a low-cost change that targets this gap directly. A third issue, recurring across Q37 and Q52, is the desire for more practical coding work, especially in service of the mathematical material - this is a familiar tension that deserves attention in any next iteration.

Three limitations bound what can be inferred. First, the sample is small (n=35), voluntary and self-selected. Second, Q40 and Q41 are retrospective, differently worded items and are vulnerable to recall bias; they are reported descriptively only. Third, the study is single-cohort, single-institution and single-course, so we cannot claim causal impact or generalisability. A further important caveat is social-desirability bias: an end-of-course survey explicitly framed around ethics may encourage respondents to over-report agreement with ethically aligned statements such as Q49. Open-ended responses were completed outside a controlled setting, and possible use of generative-AI tools could not be ruled out; they are therefore interpreted as submitted examples of perceived transfer rather than independently verified evidence of unaided competence. Additional limitations include instructor effects (the first author was involved in both course delivery and qualitative coding, which may have influenced both the design and the interpretation of responses) and possible LMS-related response bias. Future iterations could add longitudinal follow-up, cross-cohort comparison and interviews on how technical and ethical content interact in students' reasoning.

7. Conclusion

This paper described a blended Machine Learning course design embedding responsible AI through a technical fairness lecture, an eight-module AI-literacy microlearning course, and weekly optional assignments. Among 35 consenting respondents, perceptions were broadly positive and students generated ethical examples from hiring, education, identification, and credit or loan systems using technical vocabulary from the course. We do not claim causal impact. The findings are nevertheless consistent with a workable approach in this context: a fairness lecture plus microlearning wrap-around appears feasible as a way to support integrated responsible-AI teaching inside a technical ML course. The most actionable revision targets, reducing workload tension, adding explicit ethical-reflection prompts, and increasing practical coding work, are concrete and low-cost. Generalisability beyond one course and cohort remains an open question.

Ethics Declaration

The survey was voluntary and anonymous, and explicit research-use consent was gathered as the first item. Only consenting responses are analysed. No direct identifiers were collected; demographic results are reported only in aggregate. Formal institutional ethical review was not sought for this study, which was conducted as a voluntary, anonymous end-of-course evaluation with explicit research-use consent collected as the first survey item.

AI Declaration

An AI language model was used solely for language revision and copy-editing. All substantive arguments, interpretations, source selections, and final wording were reviewed and approved by the authors, who take full responsibility for the final manuscript.

References

Aler Tubella, A., Mora-Cantallops, M. and Nieves, J.C. (2024) "How to teach responsible AI in Higher Education: challenges and opportunities", Ethics and Information Technology, Vol 26, Article 3. doi:10.1007/s10676-023-09733-7.

Anderson, T. and Shattuck, J. (2012) "Design-based research: A decade of progress in education research?", Educational Researcher, Vol 41, No. 1, pp 16–25. doi:10.3102/0013189X11428813.

Baker, R.S. and Hawn, A. (2022) "Algorithmic bias in education", International Journal of Artificial Intelligence in Education, Vol 32, pp 1052–1092. doi:10.1007/s40593-021-00285-9.

Borenstein, J. and Howard, A. (2021) “Emerging challenges in AI and the need for AI ethics education”, AI and Ethics, Vol 1, pp 61–65. doi:10.1007/s43681-020-00002-7.

Braun, V. and Clarke, V. (2006) “Using thematic analysis in psychology”, Qualitative Research in Psychology, Vol 3, No. 2, pp 77–101. doi:10.1191/1478088706qp063oa.

Buolamwini, J. and Gebru, T. (2018) “Gender shades: Intersectional accuracy disparities in commercial gender classification”, in Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, Vol 81, pp 77–91.

Casal-Otero, L. et al (2023) “AI literacy in K-12: a systematic literature review”, International Journal of STEM Education, Vol 10, Article 29. doi:10.1186/s40594-023-00418-7.

Eubanks, V. (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, St. Martin’s Press, New York.

European Parliament and Council of the European Union (2024) “Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)”, Official Journal of the European Union, L 2024/1689, 12 July. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Ferrer, X. et al (2021) “Bias and discrimination in AI: A cross-disciplinary perspective”, IEEE Technology and Society Magazine, Vol 40, No. 2, pp 72–80. doi:10.1109/MTS.2021.3056293.

Fiesler, C. et al (2021) “Integrating ethics into introductory programming classes”, in Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE ’21), ACM, New York, pp 1027–1033. doi:10.1145/3408877.3432510.

Garrison, D.R. and Vaughan, N.D. (2008) Blended Learning in Higher Education: Framework, Principles, and Guidelines, Jossey-Bass, San Francisco.

Grosz, B.J. et al (2019) “Embedded EthiCS: integrating ethics across CS education”, Communications of the ACM, Vol 62, No. 8, pp 54–61. doi:10.1145/3330794.

Hug, T. (2005) “Micro learning and narration. Exploring possibilities of utilization of narrations and storytelling for the designing of micro-units and didactical micro-learning arrangements”, in Proceedings of the Fourth Media in Transition Conference, MIT, Cambridge, MA.

Jayapal, I.V., Hammerman, J.K.L. and Chaspari, T. (2025) “Expanding AI Ethics in Higher Education Technical Curricula: A Study on Perceptions and Learning Outcomes of College Students”, in Proceedings of the 2025 ASEE Annual Conference and Exposition, Montreal, Quebec, Canada. doi:10.18260/1-2--56476.

Joint Task Force on Computing Curricula (2013) Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science, ACM and IEEE Computer Society, New York. doi:10.1145/2534860.

Kasneci, E. et al (2023) “ChatGPT for good? On opportunities and challenges of large language models for education”, Learning and Individual Differences, Vol 103, 102274. doi:10.1016/j.lindif.2023.102274.

Leong, K. et al (2021) “A review of the trend of microlearning”, Journal of Work-Applied Management, Vol 13, No. 1, pp 88–102. doi:10.1108/JWAM-10-2020-0044.

Long, D. and Magerko, B. (2020) “What is AI literacy? Competencies and design considerations”, in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, ACM, New York, pp 1–16. doi:10.1145/3313831.3376727.

Mehrabi, N. et al (2021) “A survey on bias and fairness in machine learning”, ACM Computing Surveys, Vol 54, No. 6, pp 1–35. doi:10.1145/3457607.

Ng, D.T.K. et al (2021) “Conceptualizing AI literacy: An exploratory review”, Computers and Education: Artificial Intelligence, Vol 2, 100041. doi:10.1016/j.caeai.2021.100041.

Reeves, T.C. (2006) “Design research from a technology perspective”, in van den Akker, J. et al (eds.) Educational Design Research, Routledge, London, pp 52–66.

Saltz, J. et al (2019) “Integrating ethics within machine learning courses”, ACM Transactions on Computing Education, Vol 19, No. 4, Article 32, pp 1–26. doi:10.1145/3341164.

Sankaranarayanan, R. et al (2023) “Microlearning in diverse contexts: a bibliometric analysis”, TechTrends, Vol 67, No. 2, pp 260–276. doi:10.1007/s11528-022-00794-x.

Skirpan, M. et al (2018) “Ethics education in context: a case study of novel ethics activities for the CS classroom”, in Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE ’18), ACM, New York, pp 940–945. doi:10.1145/3159450.3159573.

Sweller, J., van Merriënboer, J.J.G. and Paas, F. (2019) “Cognitive architecture and instructional design: 20 years later”, Educational Psychology Review, Vol 31, pp 261–292. doi:10.1007/s10648-019-09465-5.

Weichert, J. and Eldardiry, H. (2025) “Educating a Responsible AI Workforce: Piloting a Curricular Module on AI Policy in a Graduate Machine Learning Course”, in Proceedings of the 2025 ASEE Annual Conference and Exposition, Montreal, Quebec, Canada. doi:10.18260/1-2--56314.

The Effectiveness of Developing and Optimizing a Self-Learning Course to Promote Chinese Learning

Felix Chao
The Independent Learning Centre, The Chinese University of Hong Kong, Hong Kong

felixchao@cuhk.edu.hk

Abstract: Most universities in Hong Kong require that students demonstrate fluency in the Chinese language. In 2013, the Department of Chinese Language and Literature (CHLL) and the Independent Learning Centre (ILC) at the Chinese University of Hong Kong initiated the development of "CHLT 1104 Self-learning Elementary Chinese" to facilitate independent learning for international students who do not possess prior Chinese language education. This paper presents the redesign and optimization of CHLT 1104 Self-learning Elementary Chinese at the Chinese University of Hong Kong from an instructional design perspective. The manuscript is structured according to an ADDIE-informed process: analysis of learner needs and course limitations, design of enhanced learning pathways and platform layout, development of new bilingual materials and trilingual videos, implementation of the revised course in the second semester of 2025–2026, and evaluation via a structured questionnaire and open-ended feedback. The study clarifies the evolution of the course from three original modules to five, articulates the pedagogical rationale for the revised design and learning activities, and clarifies the differences between quantitative questionnaire results and qualitative feedback themes. The evaluation findings indicate significant student approval of the updated platform, while the open comments provide detailed information regarding usability, material organization, and recommendations for further enhancement. Numerous students believe that the newly integrated elements will greatly improve their understanding of the Chinese language and culture. However, feedback indicating the necessity for design enhancement or an extension of the course parameters has also been acknowledged. The design and optimization of this course is considered effective and successful.

Keywords: Self-learning Course, Chinese Learning, Online Learning, Independent Learning, Language Learning

1. Introduction and the Research Question

The need for independent Chinese learning at CUHK is not new. Since the 1980s, CHLL has developed self-learning resources for Chinese language study, created a language knowledge reader, “Selected Essays on Chinese Language,” mainly for students to study independently for the then-mandatory Chinese course. (Chao, 2015) From 2012 to 2015, the content of this book evolved into “The Chinese Grammar Site,” an online learning platform created by Independent Learning (ILC), with a funding from the Teaching Development Grant (TDG). Building on this tradition, CUHK and ILC launched several self-learning Chinese courses, including CHLT 1104 Self-learning Elementary Chinese, an introductory course for beginners. The course combines mandatory lectures, optional consultations, and an online platform to support guided independent learning in three areas: Chinese Characters, Cantonese, and Putonghua. This design reflects the view that self-learning courses should accommodate differences in learners’ self-regulated learning capacity and provide both structure and flexibility (Lin, Gong, and Xu, 2021). CHLT1104 is designed as a self-learning course that allow students to achieve their specified learning goals at their own pace, according to their learning habits and with proper guidance.

The research question of this paper is: Can the optimization project enhance the effectiveness of CHLT 1104 Self-learning Elementary Chinese to promote Chinese learning? This question is answered by explaining the analysis of the original course, the aims of the optimization project, the project's development, the implementation, the evaluation findings, and the following discussion.

2. Analysis of the Original Course

The need for revision of CHLT1104 emerged from both prior course experience and student feedback. Since the launch of CHLT 1104 in 2014, approximately 200 to 250 students have enrolled each semester, and the student population has become increasingly diverse, including local, mainland, and international students with different levels of prior exposure to Chinese. One of the comments collected from the course evaluation in the first semester of 2022, stated that: “I found this course very helpful, but as a native Mandarin speaker, the recommended links and books might be a bit basic. It would be even better if there were learning materials at different levels for native and non-native speakers. However, the teacher's lectures were very clear. Due to the pandemic and online classes, my Zoom sometimes didn't work, and I'm very grateful to all the teachers for their timely feedback via email. Thank you again, and best wishes!” (informant U) Similar comments could be found on course evaluations from more recent years, for example, another student’s comment from first semester of 2023 stated that: “If it’s possible it might be better to classify student's ability before the course lecture.”

(informant X). The comments suggested that certain students needed more explicit guidance for self-learning, while others found certain aspects of the original materials to be too basic and these comments should consider a compelling justification for course enhancement. Therefore, to effectively address the existing diverse needs of students and the future development, ILC and CHLL decided to revise and enhance the complete curriculum. A project entitled "Enriching the Self-Learning Resources of CHLT 1104 Elementary Self-Learning Chinese for International Students," funded by the "Teaching Development and Language Enhancement Grant 2022–2025" at CUHK, was effectively implemented aiming to enhance the content and design of CHLT 1104.

3. Aims of the Optimization Project

The project aims to optimize the content and design of CHLT 1104 Self-learning Elementary Chinese in response to the increasing diversity of its student population. It focuses on enhancing the online platform through written materials, exercises, feedback questions, and multilingual videos to support more effective independent and deep learning in Chinese language, writing system, grammar, and culture.

4. Development of the Optimization Project

The optimization is focused on both curriculum and platform layout.

Table 1: revisions that have made:

Design Element	Revision Made	Purpose
Course structure	Expanded from 3 modules to 5 modules	Address different proficiency levels and broaden learning pathways
Content organization	Shorter sections with clear subheadings	Improve readability and support self-paced learning
Language support	Bilingual written content and trilingual videos	Accommodate diverse student language backgrounds
Navigation	Indexed layout and side navigation buttons	Help learners locate relevant content efficiently
Learning activities	Added reflection and repetition exercises	Promote active engagement rather than passive reading
Media design	Added 30 short, subtitled videos	Support comprehension and flexible independent learning
Evaluation	Embedded 10-question feedback form	Collect formative and summative feedback on the redesign

The existing three modules— "Learning Chinese Characters by Yourself," "Learning Cantonese by Yourself," and "Learning Putonghua by Yourself"—were updated with new information and new angles. The content is prepared in both Chinese and English; users can switch languages by clicking the "Switch language" button on the webpage. Furthermore, the written content of all learning modules is divided into small sections, each with a clear subheading. Users can use the buttons on the left side of the webpage to jump directly to the content they wish to view. Please see Figure 2 for an example of how the module has been presented.



Figure 1

In response to students' feedback that the existing modules were too basic, two new modules—"Modern Chinese Grammar" and "Introduction to Chinese Literature and Culture"—were added to broaden the curriculum and better accommodate differences in learner proficiency, motivation, and depth of interest. The grammar module introduces key features of modern Chinese grammar, such as quantifiers, word order, tenses, and auxiliaries, for intermediate-level learners, while the culture module has the broadest scope of the five modules, covering traditional Chinese religions (*Please see Figure 3) in Hong Kong as well as major literary genres, including poetry, lyrics, novels, and prose, alongside an introduction to Hong Kong culture and literature.



Figure 2

Additional written explanations, reflective questions, and exercises were integrated into the revised platform so that the site would support guided self-learning rather than function only as a repository of links and notes. Please see Figure 4 for an example.

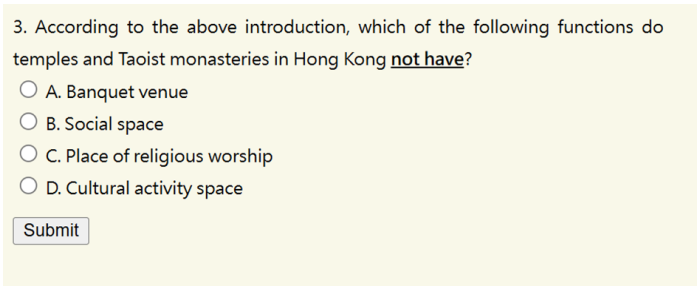


Figure 3

To enhance independent learning, the project added instructional videos to the online platform. Instead of the original plan for 5 English videos, the team produced 30 subtitled videos in English, Mandarin, and Cantonese to better suit students' diverse language backgrounds. The videos are usually under four minutes, and Figure 5 shows a screen capture of the video on Chinese kung fu culture.



Figure 4

An evaluation form was developed and embedded into the online platform to collect users' opinions on the optimization. Please see Figure 6 for the first part of the feedback system.



Figure 5

5. Implementation of the Revised Course

The project was originally scheduled from February 2024 to April 2025, but the need to develop multilingual videos extended the work until the end of 2025. The revised course was then implemented in the second semester of the 2025–2026 academic year, from January to April 2026. During this stage, students engaged with the optimized platform as part of CHLT 1104 and were invited to complete an embedded evaluation form after using the revised materials.

6. Evaluation and Findings

This paper adopted a mixed-method research approach to examine the optimization of CHLT 1104 Self-learning Elementary Chinese. Burke et al. (2014) contend, mixed-method research can take several forms, and this study adopted a quantitatively driven design in which quantitative analysis was given priority while qualitative comments served to complement and enrich the interpretation of the survey results. The revised course was implemented in the 2025–2026 academic year and evaluated through an embedded project-specific questionnaire. The questionnaire comprised 10 items: nine closed-ended questions rated on a five-point agreement scale with a non-applicable option— "Strongly Agree" (6), "Agree" (5), "Slightly Agree" (4), "Slightly

Disagree" (3), "Strongly Disagree" (2), and "Non-Applicable" (N/A) (1)—and one open-ended question for additional comments. The evaluation questionnaire was project-specific, as it was designed explicitly for this course optimization project to gather feedback on the usability, students' perceptions, and overall effectiveness of the revised CHLT 1104 course platform. It was therefore tailored to the aims of the study rather than adopted as a generic institutional survey. Between January and April 2026, 156 students completed the form. Quantitative data were analyzed descriptively, and the open-ended responses were analyzed thematically to identify recurring patterns in student feedback. Overall, the findings show very positive student perceptions of the redesigned course.

6.1 Quantitative Findings

For Question One: The interface of the platform is user-friendly, we can see from Figure 7 that 3 students have indicated dissatisfaction (*slightly disagree and below) with the system's interface. The combined ratings of "Slightly Agree" and above have collected a total of 153 students, or almost 98% of the total participants indicated that most of the students found the interface user-friendly. Moreover, 121 students, or over 77% of the total participants, have selected "Agree" or a higher rating for this question.

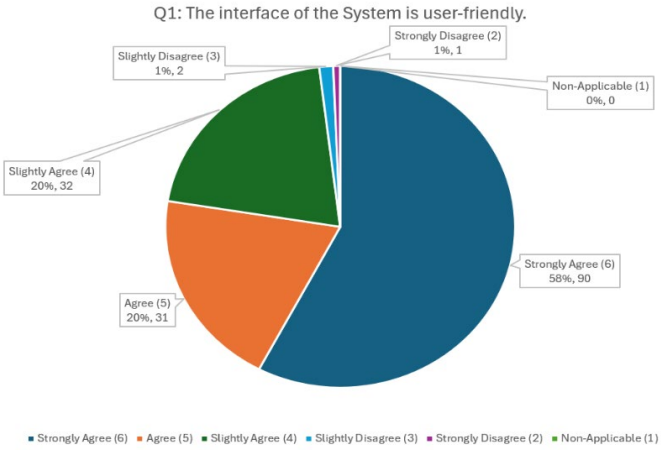


Figure 6

For Question Two: The difficulty level of each learning module is appropriately balanced, a total of 152 students from the combined ratings of "Slightly Agree" and above, or almost 97% of the participants, found the difficulty level of each learning module is appropriately balanced (based on Figure 8). A total of only 4 students expressed dissatisfaction. Furthermore, 114 students, or over 73%, picked the "Agree" or higher rating option for this question, indicating that one of the main objectives of this optimization—addressing the needs of students with varying proficiency levels in Chinese—should be considered quite successful.

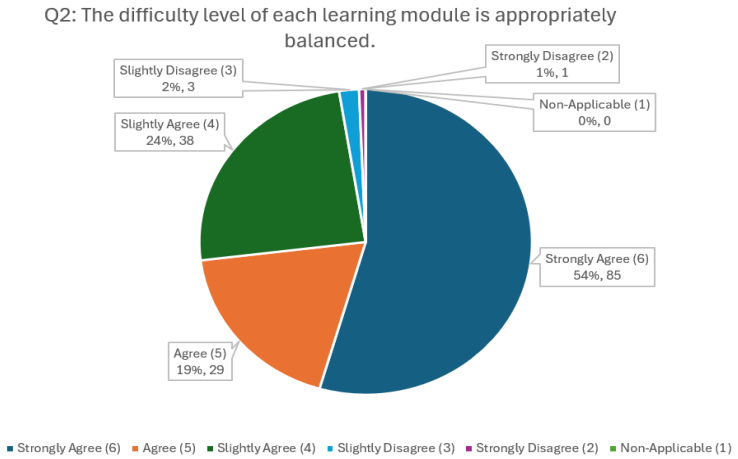


Figure 7

For Question Three: The content of the learning modules facilitates my understanding of the evolution and

characteristics of Chinese characters, only 1 student indicated dissatisfaction with the content of the learning module on Chinese characters, while 3 students selected "Non-Applicable" (N/A) (based on Figure 9). A total of 152 students, covering almost over 97% of participants, rated the learning modules as "Slightly Agree" or higher, indicating that these modules facilitate the understanding of Chinese characters. Furthermore, 126 students, or almost 81% of the total participants, selected "Agree" or a higher rating. Among the 5 learning modules, this is the highest number that combines these two ratings.

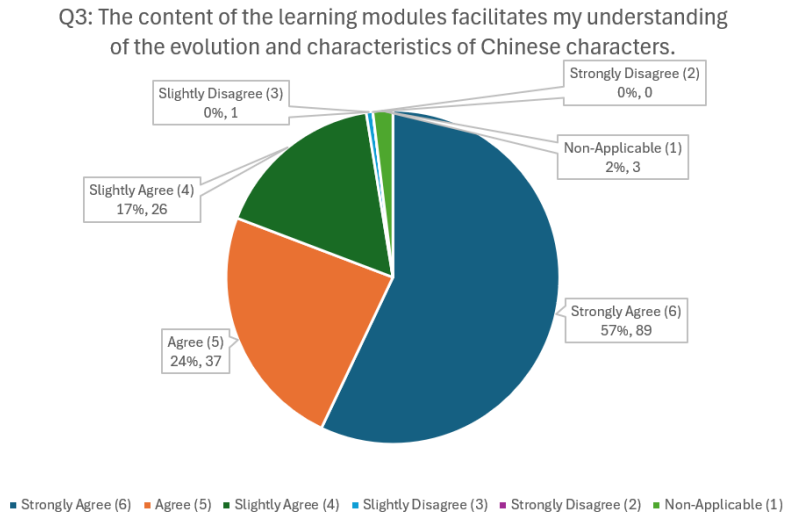


Figure 8

For Question Four: The content of the learning modules enhances my awareness of the fundamental features of Cantonese, although there are up to 3 students indicated dissatisfaction with the content of the learning module on Cantonese, we can see from Figure 10 that the combined ratings of “Slightly Agree” and above is still very high, there were 152 students or almost over 97% of the total participant. It would indicate that the student still found the learning module very good for learning Cantonese. Furthermore, the number of students who chose “Agree” or a higher rating is also very high; there are a total of 116 students, or over 74% of the total participants.

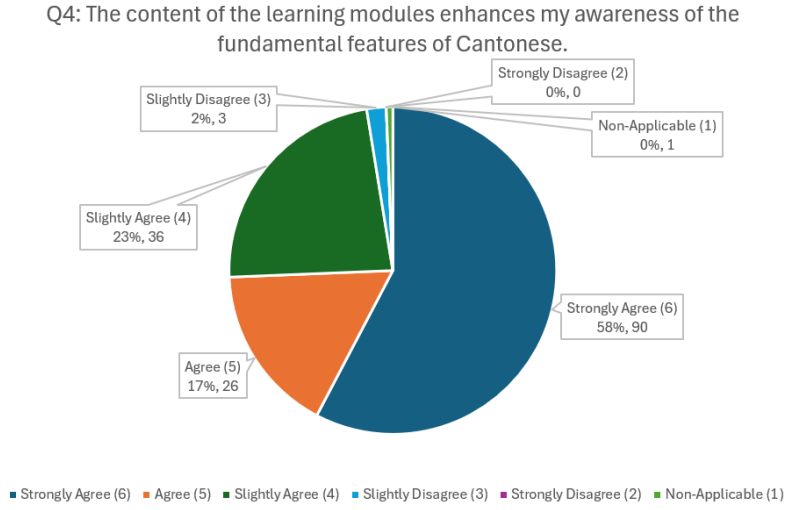


Figure 9

For Question Five: The content of the learning modules enhances my awareness of the fundamental features of Putonghua, 3 students indicated dissatisfaction with the content of the Putonghua learning module, while 3 others selected "Non-Applicable" (N/A), resulting in a total of 150 students from the combined ratings of “Slightly Agree” and above (based on Figure 11). Although 150 is still very high (covering over 96% of the participants), it is the lowest number among the 5 learning modules (Chinese characters, Cantonese, Putonghua, Modern Chinese grammar & Chinese culture). Furthermore, 114 students, or more than 73% of the total

participants, have chosen an "Agree" or higher rating. This number is also a bit low when compared with other learning modules.

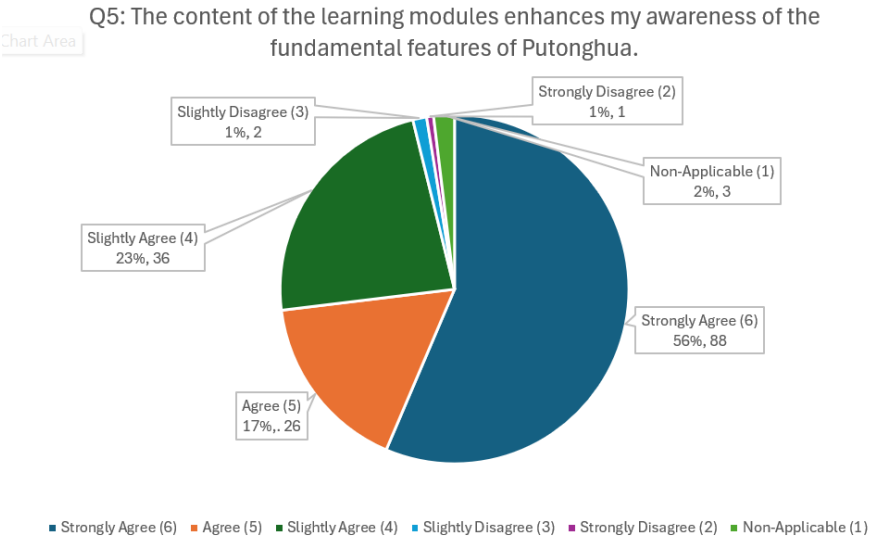


Figure 10

For Question Six: The content of the learning modules helps me understand the learning difficulties of modern Chinese grammar, this is the only question for which no students indicated dissatisfaction, while 2 students selected “Non-Applicable” (N/A). Based on Figure 12, a total of 154 students, or nearly 99% of the total participants, selected “Slightly Agree” or above, indicating that the learning module supports their comprehension of modern Chinese grammar; this is the highest total across the combined positive ratings. Furthermore, 119 students, or over 76% of the total participants, chose “Agree” or above.

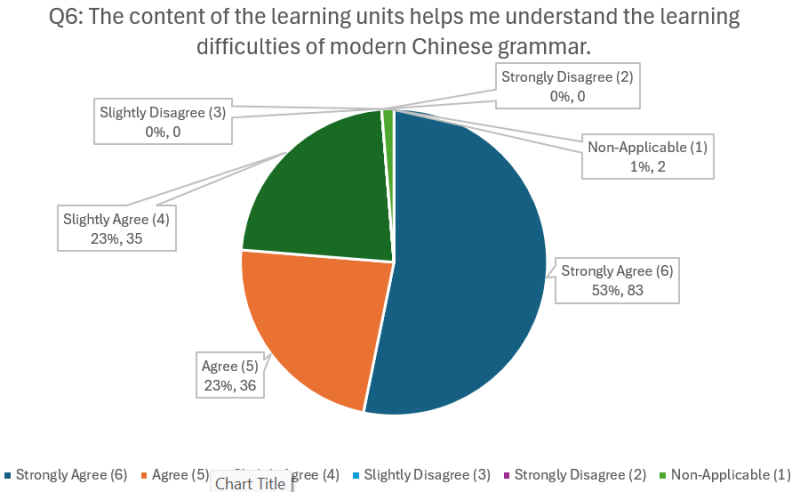


Figure 11

For Question Seven: The content of the learning modules broadens my understanding of the diverse aspects of Chinese culture, only 1 student has indicated dissatisfaction with the content of the Chinese culture learning module, while 3 students selected “Non-Applicable” (N/A) (based on Figure 13). A total of 152 students from the combined rating of “Slightly Agree” and above, or over 97% of the total participants, indicated that the learning modules support their comprehension of Chinese culture. Furthermore, 122 students, or over 78% of the total participants, have chosen an "Agree" or higher rating. Among the 5 learning modules, this is the second highest number that combines these two ratings.

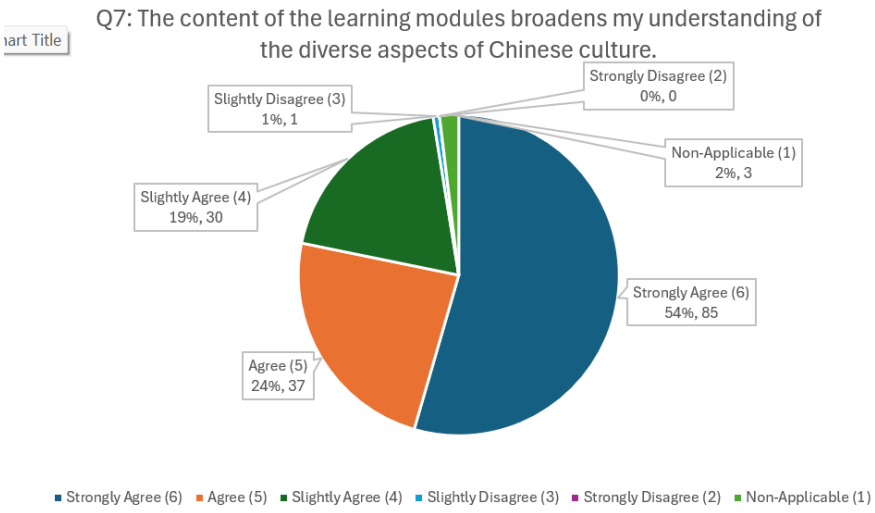


Figure 12

For Question Eight: The suggested learning resources are useful for my own learning of Chinese, we can see from Figure 14, almost all students, or nearly 99% of the total participants from the combined ratings of “Slightly Agree” and above, indicated that learning resources are useful for the learning of Chinese. Only 1 student indicated dissatisfaction, while 1 student selected “Non-Applicable” (N/A). Furthermore, 124 students, or over 79% of the total participants, have chosen an "Agree" or higher rating.

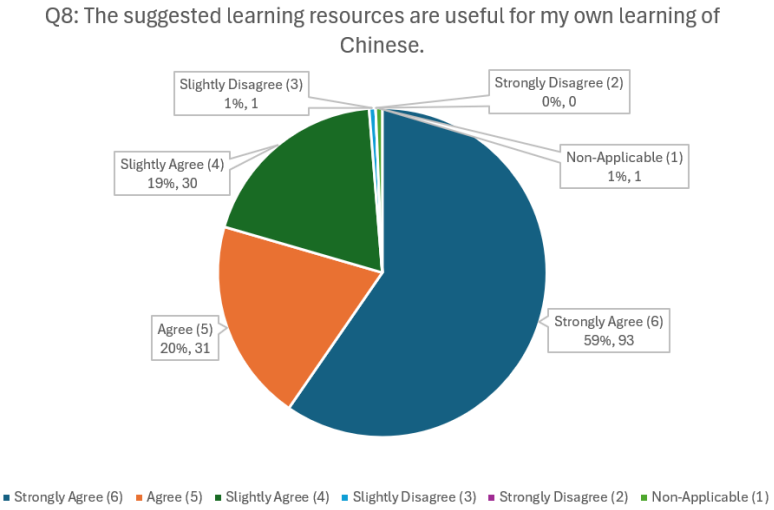


Figure 13

For Question Nine: I will recommend the platform to my fellow CUHK students, 5 students indicated that they would not recommend the platform to other CUHK students, while two students selected "Non-Applicable" (based on Figure 15). This is the greatest number of negative responses among all feedback questions. Nonetheless, there are still 149 students from the rating of "Slightly Agree" and above, or over 95% of the total participants, who indicated that they would recommend the platform. Moreover, 115 students, or nearly 74% of the total participants, selected a "Agree" rating or above.

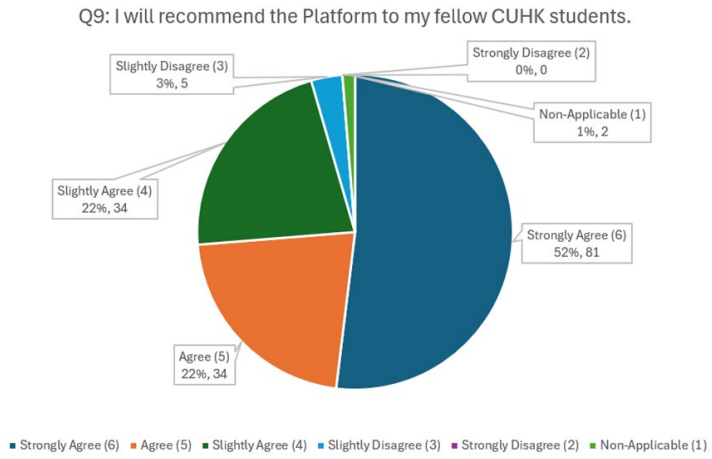


Figure 14

6.2 Qualitative Findings

For Question Ten, the open-ended responses provided the qualitative component of the evaluation. These responses were first reviewed and then coded into recurring thematic categories to identify the main patterns in students' feedback. Seven themes were derived inductively from the content of the 32 comments : (1) user experience, (2) suggestions for improvement, (3) language learning modules, (4) platform design, (5) platform content, (6) learning and usage experience, and (7) no comments. This qualitative analysis complements the quantitative survey findings by providing a deeper understanding of how students perceived the revised CHLT 1104 platform.

As shown in Figure 16, the overall feedback was overwhelmingly positive. Only two comments were negative: "The duo push is not available for those who change devices and do not delete them" (Informant A) and "Slow website" (Informant B), and four entries contained no comment. In addition, six comments offered positive remarks on the overall learning experience, such as "Very good" (Informant C) and "Very efficient" (Informant D). The content and design of the platform also received ten positive comments, including "It is clear and well organized" (Informant E) and "A fairly well designed and made webpage with significant resources that are well selected and well prepared. Would rate it 8 out of 10" (Informant F), suggesting that students valued the enhancements in these areas. Furthermore, six comments provided constructive suggestions for further development, indicating that students were engaged thoughtfully with the platform and remained committed to their learning.

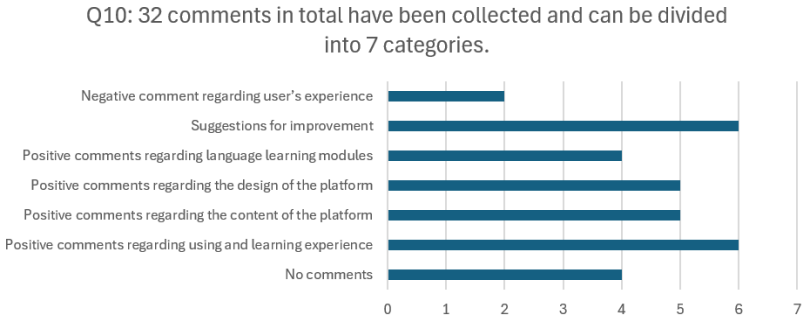


Figure 15

7. Discussion

7.1 Platform Design and Independent Learning Support

Feedback from the evaluation form suggests that the platform's design effectively supports independent learning. Question One received strong support: 153 students (about 98%) described the interface as user-friendly, and more than three-quarters selected "Agree" or "Strongly Agree." (over 77%) This indicates that the platform lowers usability barriers, helping learners focus on learning tasks rather than navigation or technical

issues. Comments in Question Ten reinforce this view, for example: “This site is very helpful and well-organized for students of any level of learning Mandarin and Cantonese.” (Informant G) and “It is clear and well organized.” (Informant H). One student also noted, “There are many resources to help with self-learning; it’s very convenient.” (Informant I) (*translated from Chinese). These comments align with the platform’s aim of selecting practical resources for self-learning. Hua, Wang, and Li (2024) suggest that student capacity to monitor their learning process and choose resources and strategies that align with their requirements is essential for effective self-directed learning. The responses to Question Eight have shown a nearly universal agreement (“Slightly Agree” or above) that the recommended resources are beneficial, indicating high satisfaction and perceived learning value. This should be considered an indication that students are confident in selecting the appropriate tools for self-learning. Therefore, putting together the responses from Questions One and Eight with selected comments from Question Ten, we can consider the platform meets student expectations for a webpage design that facilitates independent learning.

7.2 Curriculum Development and Learner Diversity

The revised curriculum for CHLT1104 has been expanded to cover a wider range of Chinese language components, including spoken Chinese (Putonghua and Cantonese), written Chinese (Chinese characters), Chinese grammar, and Chinese culture. A core objective of the optimization is to strengthen and modernize the online platform to accommodate students with different proficiency levels and interests in Chinese learning. We can consider this objective met, because across all nine closed-ended questions on the evaluation form, more than 95% of the 156 respondents selected “Slightly Agree” or above, suggesting consistently positive perceptions of the platform. These high satisfaction levels indicates that the current platform already provides an accessible, supportive environment that satisfies the learning needs of an increasingly diversified cohort of local, mainland, and international students. Further evidence can be found in the comments provided by students. One student praised the platform for providing comprehensive explanations for Chinese learning: “This website provides an accessible explanation of Chinese culture and Chinese characters, offering comprehensive and diverse information. It's suitable for beginners, those interested in Chinese culture and language who wish to further improve their language skills, and even native speakers seeking advanced learning opportunities. Highly recommended.” (Informant J) (*Translated from Chinese). Meanwhile, another student appreciated the site’s support for different proficiency levels, noting: “This site is very helpful and well-organized for students of any level of learning Mandarin and Cantonese.” (Informant K). These comments suggest that students can choose what they need based on their own requirements, and the online system can support effective self-disciplined learning. A focus study conducted by ILC in 2023–2024 on the perception of independent learning of CUHK students revealed that self-discipline is considered a critical component of effective independent learning, particularly in an online environment, as it is perceived as self-determination and a result of self-motivation. (Chao, 2025) The optimization has enhanced the online platform with more diverse and broader content, offering more entry points and directions for exploration and making it easier for students, especially beginners, to maintain a self-disciplined learning attitude.

7.3 Module-Level Learning Outcomes and Self-Learning

Basith et al (2019) argued that anxiety can significantly impede Chinese language comprehension, especially for foreign learners who struggle with tones, characters, and performance pressure. Learning difficulties may also stem from the design of the materials themselves. Chen (2021) found that international students often have difficulty engaging with online materials related to Chinese cultural elements and therefore rely heavily on teachers for learning and problem solving. In this regard, Informant L’s comment— “it's a very well-balanced read that meets the middle-ground between depth and simplicity. Most of the text cuts to the chase with minimal filler, which is something I can appreciate.”—suggests that the platform’s written content was well received, consistent with the high ratings for all five learning modules. Guo, Li, and Guo (2024) further note that student satisfaction and continued engagement in online learning can be improved through stronger content, delivery, and instructional design. The five modules showed only minor differences in evaluation. The Chinese characters module (Question Three) performed particularly strongly, with about 81% of students selecting “Agree” or above, the highest among the five modules for these two positive categories. This suggests that it was especially effective in supporting students’ understanding of the evolution and characteristics of Chinese characters. The Chinese culture module (Question Seven) also performed strongly, with more than 78% selecting “Agree” or higher. By contrast, although the Putonghua module (Question Five) maintained a high approval rating, it recorded the lowest percentages for both “Slightly Agree” or above and “Agree” or above. Although over 96% of students still rated it at least “Slightly Agree,” and more than 73% selected “Agree” or higher, this

relative decline suggests that students may have found the Putonghua content slightly less beneficial or less aligned with their needs. These differences are likely related to variations in the writing approach. Additional elements such as historical facts, distinctive angles, and unique characteristics enriched the Chinese characters and Chinese culture modules, and this approach appears to have been well received, as shown in the comment "very rewarding and informative" (Informant N). Taken together, the evaluation results and student comments suggest that the updated content across all five modules successfully increased students' interest in Chinese learning and supported effective self-learning.

7.4 Feedback on Putonghua and Cantonese Modules

The independent learning survey conducted by ILC in 2022–2023, indicated that most students tended to associate self-learning with academic writing and language learning (Chao et al 2023). The optimization intends to accommodate student's language learning needs, by further develop their interest and proficiency in Chinese languages through revising and updating the content and design of the two Chinese language modules. Fundamental introductions, informative explanations and interesting facts have been provided to Putonghua (the official language of China) and Cantonese (the local dialect of Hong Kong) modules. In result, both modules received positive feedback, including comments such as "This site is very helpful and well-organized for students of any level of learning Mandarin and Cantonese." (Informant O), indicating that students appreciate the content and design. However, there are small differences between the Putonghua and Cantonese modules. The Cantonese module shows a slightly higher acceptance rate than the Putonghua module, exceeding it by about 2% in the "Strongly Agree "categories. In addition, written comments about Cantonese are more detailed and informative. One student remarked: "Thank you ILC teachers who created this course for foreigners and other students. I have learnt a lot more regarding the background and history of my mother tongue Cantonese. And I would love to preserve this language to future generations and help other people discover this amazing and complex language." (Informant P). This comment is notable because it reflects not only knowledge gains but also reflection on how the platform may support self-learning of local language and culture. Another student even makes a deeper reflection on how to further develop the online platform by proposing: "It would be even better to launch a self-study platform for 'Hakka' and 'Weitou' dialects in collaboration with the Association for Conservation of Hong Kong Indigenous Languages. Both 'Hakka' and 'Weitou' dialects are local languages, yet they are neglected and lack sufficient learning resources. Therefore, it would be beneficial if the self-learning platform could help preserve these endangered local dialects." (Informant S) (*translated from Chinese). These comments suggest that the language modules—especially the Cantonese module—can successfully promote effective independent learning and support deep learning by motivate students to evaluate and reflect deeply on the significance and potential application of what they have learned.

7.5 Multimedia Learning and Video Effectiveness

The 30 videos produced for the online platform are highly appreciated by students. Comments such as "Really like the content of the videos, helped me gain a better understanding of Chinese" (Informant T) and "Videos are very helpful and easy to understand." (Informant W) suggest that the videos are an effective instructional method. Chan and Chao (2022) have studied the effect of adding sign language and subtitles to instructional videos and found that well-produced videos can encourage not just SEN students but also average students to engage in independent learning. The newly created videos are designed to foster an environment in which students can practice and strengthen independent learning skills; based on the positive comments received, this component should be considered successful. Li et al (2025) suggested that improved or additional academic guidance from teachers is crucial for achieving better learning results in online educational environments. However, teacher guidance can also be provided through alternative formats, such as an online platform with comprehensive instructional videos and curated learning resources. The high evaluation ratings for the interface (Question One) and the recommended learning resources (Question Eight), together with the strongly positive written comments about the videos, suggest that students can rely on an effective online platform, like the one provided by CHLT1104 to support successful online learning.

8. Conclusion

Despite project delays and minor adjustments, the evaluation findings indicate that the optimization of CHLT1104 Self-learning Elementary Chinese was effective. The revised curriculum and online platform appear to meet the needs of a more diverse student cohort, with particularly strong responses to the Chinese characters, Cantonese, and Chinese culture modules. Overall, the updated materials were perceived as useful,

accessible, and motivating, and the positive feedback on the interface, learning resources, and instructional videos suggests that the platform effectively supports independent learning. The results further indicate that the Cantonese module may be particularly beneficial for reflection and deeper learning. In addition, the positive responses and comments indicate that a well-designed online platform, equipped with sufficient instructional videos and suitable learning resources, can serve as an appropriate alternative to traditional educational settings. Taken together, the optimized CHLT 1104 Self-learning Elementary Chinese course demonstrated an effective blended self-learning approach for tertiary level Chinese language learning by engaging students with a more independent and deep learning-ready content and design.

Ethics Declaration

This study was approved under the project "Enriching the Self-Learning Resources of CHLT 1104 Elementary Self-Learning Chinese for International Students" funded by Teaching Development and Language Enhancement Grant (TDLEG) (reference no. #4171102) and was conducted in accordance with the guidelines of the Survey and Behavioural Research Ethics Committee (SBRE) of The Chinese University of Hong Kong (CUHK).

AI Declaration

Generative AI tools have been applied solely to verify the grammar and refine the language of this paper. The content of this paper, which includes all data, concepts, research methods, research outcomes, research findings, citations and references was entirely developed or composed by the author.

References

Basith, A. et al. (2019) "Chinese Learning Anxiety on Foreign", *Students European Journal of Educational Research*, Vol 8, Issue 4, pp1193-1200. https://doi.org/10.12973/eu-jer.8.4.1193

Burke, R. J. et al. (2014) "Conducting Mixed Methods Research: Using Dialectical Pluralism and Social Psychological Strategies", in Patricia Leavy (ed.), *The Oxford Handbook of Qualitative Research*, Oxford Library of Psychology (online ed, Oxford Academic), pp557-569 https://doi.org/10.1093/oxfordhb/9780199811755.013.022

Chan, P., & Chao, F. (2022) "Improving e-Learning Videos by Adding Sign Language Interpretation and Subtitles", In the *Conference Proceedings and Abstracts of the 21st European Conference on e-Learning (ECEL2022)*, pp42-50. https://papers.academic-conferences.org/index.php/ecel/article/view/576/768

Chao, F. (2015) "The Design and Application of the CUHK Online Learning Platform for University Chinese Grammar in Use" , *Conference Proceedings and Abstracts of ILC Symposium 2015*. https://www.ilc.cuhk.edu.hk/ILCSymposium2015/CH/paper.aspx

Chao, F. et al. (2023) "Preliminary Findings of the Needs Perceived and Expectations of Users of an Independent Learning Centre Survey in Hong Kong: What do They Tell Us?" ,*Conference Proceedings and Abstracts of the 22nd European Conference on e-Learning (ECEL2023)*, pp39-50. https://papers.academic-conferences.org/index.php/ecel/issue/view/27/29

Chao, F. (2025) "The Findings of Focus Group Study on Independent Learning in A Digital Age", *Conference Proceedings and Abstracts of the 24th European Conference on e-Learning (ECEL2025)*, pp49-59. https://papers.academic-conferences.org/index.php/ecel/issue/view/54

Chen, C. (2021) "Using scaffolding materials to facilitate autonomous online Chinese as a foreign language learning: A study during the COVID-19 pandemic", *SAGE Open*, Vol11, No 3, pp1-12. https://doi.org/10.1177/21582440211040131

Guo, X., Li, X., & Guo, Y. (2024) "Factors Influencing the Satisfaction of Second Language Learners of Chinese in Online Courses", *Behavioral Sciences*, Vol 14, No 5: 387. https://doi.org/10.3390/bs14050387

Hua, M., Wang, L., & Li., J (2024) "The impact of self-directed learning experience and course experience on learning satisfaction of university students in blended learning environments: the mediating role of deep and surface learning approach", *Front Psychol*, Vol 14:1278827. doi: 10.3389/fpsyg.2023.127882

Li, L. et al. (2025) "Unlocking Deep Learning: How Self-Regulated Learning Shapes Chinese Students' Online Course Experiences". *SAGE Open*, Vol15, No. 3, pp1-14. https://doi.org/10.1177/21582440251365390

Lin, L., Gong, Y., & Xu, N. (2021) "Online Self-Regulated Learning Profiles: A Study of Chinese as a Foreign Language Learners", *Front. Psychol,* Vol 12:797786. doi: 10.3389/fpsyg.2021.797786

Stiller, H. S. & Chao F. (2021) "OPSCL: A Comprehensive System of Reviews for Chinese Learning Materials", *Conference Proceedings and Abstracts of the 20th European Conference on e-Learning (ECEL2021)*, pp435-444. https://www.academic-conferences.org/wp-content/uploads/2023/01/ECEL21-Proceedings_embedded.pdf

PhD Research Papers

Shifting the Time of Practice: A Digital Environment for Student-Athletes and Remediation

Paul Ronald Klensch and Yves Kreis
University of Luxembourg, Esch-sur-Alzette, Luxembourg

paul.klensch.001@student.uni.lu
yves.kreis@uni.lu

Abstract: A student who trains and competes during school time misses mathematics lessons that are never given again. At the High School for Sports in Luxembourg, these absences are authorised, but authorisation does not return the missed lesson. Students must catch up, often when they are least available, and difficulties in doing so may have consequences for end-of-year promotion. The formal end-of-year remediation process arrives late: a summer assignment is issued in July, and the promotion decision is made in September. In 2023/24, 6.9 per cent of students in classical secondary education received a summer assignment, and 51.4 per cent of these students were not promoted after completing it (MENJE, 2025a). Autonomous study hours were introduced at the High School for Sports in 2025/26 across several subjects. Based on the school's implementation experience, the arrangement was judged not to have achieved its intended purpose and was redesigned for 2026/27. In mathematics, students in grades 7 and 8 move from four taught hours a week to three, the fourth becoming an autonomous study hour supported by the school's newly implemented learning management system. EMERALD (Enhancing Mathematics Engagement through Responsive and Autonomous Learning Design) is a curriculum-aligned digital practice environment for secondary mathematics, built and studied by the first author and integrated into the school's learning management system. Two studies follow. The first accompanies one cohort through the autonomous study hours for three school years. Its primary outcome is the temporal pattern of practice, tracking when students work, for how long, and how that reorganises around training and competition. The second study asks whether a digital environment can identify learning gaps early enough to support earlier remediation. This paper reports the design phase only and presents no empirical results.

Keywords: Autonomous Learning, Learning Time, Student-Athletes, Remediation, Design-Based Research

1. Introduction

Research on digital learning environments in mathematics has often focused on their effects on students' learning outcomes, attitudes, or engagement. Systematic reviews show that many studies use technology to enhance existing teaching rather than change the organisation of learning (Bray and Tangney, 2017; Darmanova, Abylkassymova and Nurmukhamedova, 2025). Less attention has been given to how digital environments may change the conditions under which learning takes place, particularly when, where and with what support students can practise.

This paper reports a doctoral project conducted at the High School for Sports in Luxembourg, a public secondary school for student-athletes where training and competition make attendance structurally irregular. The first author teaches the participating classes, supervises autonomous study hours, and is a member of the school development unit.

The paper reports the design phase of the project. It describes the setting, the theoretical framework, the environment, and the two study designs that have been submitted for ethical review. The central question is whether a digital environment can help student-athletes by allowing them to practise at times compatible with their training and competition schedules and giving them greater control over when and what they practise.

Research on distributed practice suggests that the temporal distribution of practice can affect later retention (Cepeda et al., 2006). This makes the timing of practice an important study dimension, rather than merely a consequence of students' schedules. In this project, however, students have autonomous study hours outside their taught hours and decide which topics to practise. The decision about when, where and what to practise therefore moves to the student.

2. Problem Setting: Missed Lessons and Delayed Remediation

Student-athletes train and compete during school time. Their absences are authorised, but authorisation does not replace the missed learning time. Students are expected to catch up afterwards, often in their own time, without guidance, and when they are least available.

Support is available at several points during the school year, but each comes with limitations. A teacher can differentiate, but the support is limited to the time available during class. Optional support courses run during the school year, but they are voluntary and therefore depend on students choosing to attend. A summer assignment follows the end-of-year results, and repeating the year comes last. Luxembourg also offers a national Summerschool programme during the summer holidays (MENJE, 2025b). These measures provide additional learning opportunities, but not continuous diagnosis of learning gaps.

The summer assignment follows a regulated national procedure (Règlement grand-ducal, 2005), while its specific content is determined locally within the school. The process leaves considerable room for differences in how teachers identify and address difficulties. Some teachers differentiate assignments according to specific learning gaps, whereas others provide a more general set of exercises. This variability was observed in two conference presentations on mathematics in Luxembourg (Klensch et al., 2025a; 2025b). At the national level, there is currently no digital environment that continuously estimates topic-level mastery to support curriculum-aligned remediation.

In classical secondary education, 6.9 per cent of students were given a summer assignment based on their end-of-year results in 2023/24, and 51.4 per cent of these students were not promoted after completing it. **Table 1** places the 2023/24 figures in a ten-year series (MENJE, 2020; 2025a). The share receiving summer assignments has ranged from 5.7 to 9.1 per cent, while non-promotion has generally remained close to one half.

Missed learning and delayed remediation therefore share the same timing problem: support arrives at a moment fixed by the school rather than when the learner needs it or is available. The project asks whether a digital environment can shift practice and diagnosis closer to when they are needed.

Table 1: The summer assignment in classical secondary education over ten years (MENJE, 2020; 2025a).

School Year	14/15	15/16	16/17	17/18	18/19	19/20	20/21	21/22	22/23	23/24
Passed in July (per cent of students)	87.3	88.4	89.0	89.2	88.6	92.3	87.7	88.4	90.3	89.9
Students assigned a summer assignment (per cent of students)	9.1	8.5	8.3	7.6	8.1	5.7	8.4	8.4	7.0	6.9
Not promoted after summer assignment (per cent of those assigned)	52.6	48.7	46.4	50.6	51.7	50.4	55.8	53.2	54.5	51.4

3. Theoretical Framework

The main theoretical framework is cognitive apprenticeship (Collins, Brown and Newman, 1989). It describes how learners develop expertise by making expert processes visible and gradually transferring responsibility to the learner. The framework is influenced by sociocultural perspectives associated with Vygotsky's (1978) zone of proximal development and situated learning (Lave and Wenger, 1991). The digital practice environment acts as a mediating artefact. The topic map, worked solutions and feedback carry part of the guidance otherwise provided by the teacher.

Collins, Brown and Newman (1989) organise cognitive apprenticeship around four dimensions: content, methods, sequencing, and sociology. EMERALD addresses the first three, while the sociology dimension is treated as a design boundary. The content dimension is addressed through domain knowledge, problem-solving strategies, and control strategies embedded in lessons and exercises. The sequencing dimension is addressed through increasing complexity and diversity. The six methods of cognitive apprenticeship, which are modelling, coaching, scaffolding, articulation, reflection, and exploration, provide the main link between the framework and the digital environment. **Figure 1** maps these methods onto the design of EMERALD.

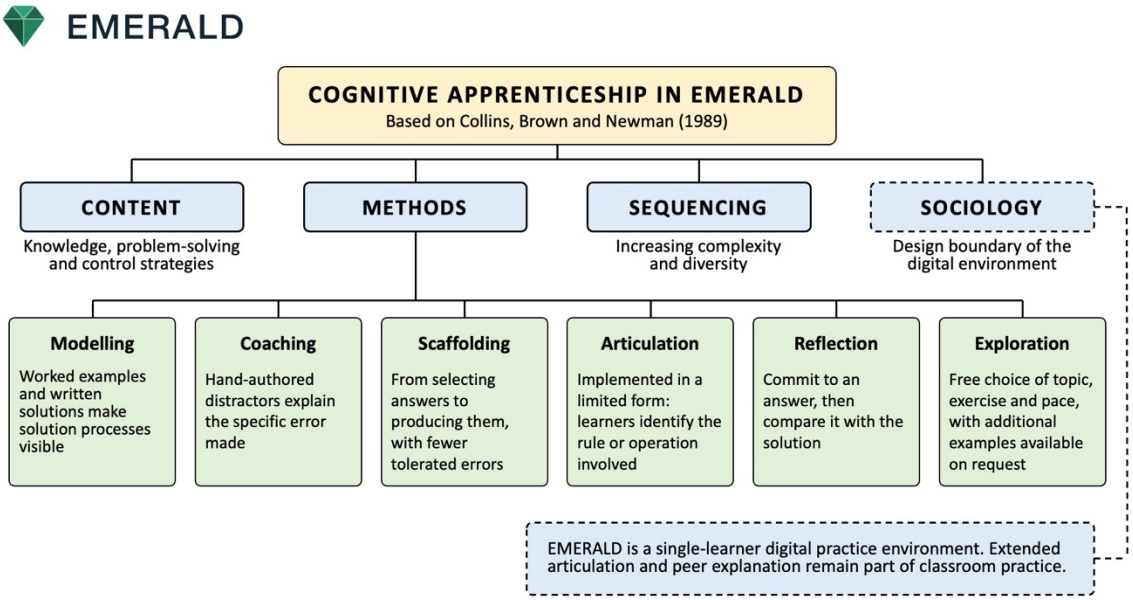


Figure 1: The mapping of the cognitive apprenticeship methods onto the design of EMERALD.

- Modelling is provided through worked examples and written solutions, allowing learners to observe complete solution processes. This also draws on research showing that worked examples can support the learning of problem-solving procedures in algebra (Sweller and Cooper, 1985).
- Coaching is provided through immediate and diagnostic feedback. Where possible, distractors are linked to specific misconceptions and trigger an explanation of the corresponding error. This is consistent with the formative-assessment view that feedback should provide information that supports learning (Black and Wiliam, 1998). EMERALD combines the two granularities distinguished by VanLehn (2011): single-step items are marked on the final answer, while multi-line calculations are marked line by line, without the expected value being revealed.
- Scaffolding is implemented mainly through the response format. As learners progress, they move from selecting answers to producing them. The number of tolerated errors also decreases as difficulty increases. Structurally, this exercise flow resembles programmed instruction: small steps, active responding, immediate feedback and self-pacing (Skinner, 1958). The wider design differs, however: the learner model is open and transparent, feedback is explanatory rather than only corrective, and learners choose their topics and their stopping points.
- Articulation is implemented in a limited form. Learners are asked to make part of their reasoning explicit by identifying an operation or rule, completing a structured representation, or naming the property involved.
- Reflection is implemented through comparison between the learner's response and the worked solution in the upper-year version of the environment. Learners first commit to an answer and then compare it with the solution.
- Exploration is supported through learner choice. Students can choose topics, start and stop sessions freely, and, in some workshops, request additional examples. Some workshops also allow learners to enter their own values and observe how the system responds.

The sociology dimension is treated as a design boundary. EMERALD does not implement peer comparison, ranking, or a community of practice, and learner-produced work is not shared between learners. It nevertheless supports classroom use through projection and teacher-controlled presentation features. Extended articulation and peer explanation remain part of classroom practice.

Cognitive apprenticeship provides the main pedagogical framework, while self-regulated learning and self-determination theory provide complementary perspectives on learner autonomy and engagement. The autonomous study hours give students greater control over their practice, but this freedom also creates demands for planning, monitoring, and adjustment. These processes are central to self-regulated learning (Zimmerman, 2002). Self-regulation may influence how much students benefit from the environment and is also

a condition for using autonomous study hours effectively. Self-determination theory provides a second perspective on autonomy and competence (Ryan and Deci, 2000). EMERALD gives learners choices about what to practise while providing information about their progress and estimated mastery level. The design therefore aims to support autonomy and competence.

Gamification provides a further, more limited design perspective. Meta-analytic evidence suggests positive average effects of gamification on cognitive, motivational, and behavioural outcomes (Sailer and Homner, 2020). However, rewards can have different effects depending on whether they are experienced as informational or controlling (Deci, Koestner and Ryan, 1999). EMERALD therefore uses playful workshops, feedback and progress information, and it deliberately excludes streaks, daily-login rewards, leagues and public rankings. The design aims to make rewards informational rather than controlling. The two lower medals are awarded when the estimate crosses its threshold; the two highest are not, because an online estimate can cross a threshold by chance, and they are confirmed by a short block of previously unseen items at the highest difficulty level. These choices aim to support learning rather than prolonged use and follow European guidance on engagement-oriented mechanisms for minors (European Commission, 2025). Whether students experience medals in this way is examined rather than assumed in Study 1.

4. Digital Practice Environment EMERALD

The present project does not seek to replace national digital mathematics practice environments such as MathemaTIC, launched in 2016 (MENJE, 2016). EMERALD is a digital practice environment developed and maintained by the first author for a more specific purpose: to align digital practice with teaching sequences, provide an open learner model, retain control over the practice environment, and integrate practice directly into the school's learning management system.

Its design draws on the concept of open learner models (Bull and Kay, 2010): learners can see the system's estimates and use them to guide their practice. They see a map of topics, each with an estimated mastery level, and can choose which topics to practise. Each topic has a pool of at least 1000 distinct items for each difficulty level. This reduces repetition for learners who practise a topic extensively and supports the study of practice over time. Each estimate is accompanied by a stability indicator based on the number, consistency and recency of the underlying responses. The indicator shows how stable the estimate currently is.

After each answer, the learner's rating is updated using an Elo-type rule (Elo, 1978), an approach used in adaptive educational systems (Pelánek, 2016) and in school mathematics (Klinkenberg, Straatemeier and van der Maas, 2011). Item selection then adapts to the learner's current rating. The design targets a modelled success probability of around 85 per cent on produced answers, and higher on selected-response items because of guessing. This target is a design decision, not a mastery criterion. Wilson et al. (2019) reported an accuracy associated with fastest learning in a class of algorithms, not a mastery threshold for human learners. Difficulty levels are therefore mixed during practice, most items falling close to the current estimate, rather than changing abruptly. Within a topic, successive items are also drawn from several distinct question forms rather than one repeated form, which is consistent with research showing that mixing problem types improves later performance in mathematics (Rohrer and Taylor, 2007). The current rating is displayed alongside the highest medal earned, providing both a current estimate and a record of achievement. Medals remain visible once earned, even if the learner's rating later falls, distinguishing current performance from accumulated achievement.

Practice time and pace are also visible to learners. A timer records elapsed practice time without imposing a time limit, while the system displays the learner's rate of practice in seconds per question. These indicators are presented as descriptive information and allow learners to monitor and reflect on their practice, consistent with the monitoring component of self-regulated learning (Zimmerman, 2002).

EMERALD is delivered as self-contained modules that run in the browser. The environment is therefore reached where students already go for their other subjects, and a teacher sees a learner's progress in the same place as the rest of the course material. The school's learning management system and EMERALD together report progress, timing, duration, pace, topic selection and item-level activity of each practice session. These traces allow Study 1 to examine when, for how long, and at what pace students practise, and how practice changes with training and competition. For the analysis, a practice session is a run of item-level events separated by no more than fifteen minutes, and its duration sums the intervals between consecutive events rather than the time the module remained open in the browser.

5. Research Questions

Three questions organise the design phase. Together, they shift the focus from attainment alone to the timing, organisation and conditions of mathematics practice.

> RQ1: When and for how long do students practise mathematics when practice is no longer tied to taught lessons, and how does practice reorganise around training and competition?
>
> RQ2: How do students differ in their use of the allocated autonomous study hours, and how is this use related to prior attainment, training and competition load, and self-regulation?
>
> RQ3: Can continuous diagnosis identify learning gaps early enough to allow targeted remediation before the end-of-year remediation process?

6. Two Studies

The project consists of two connected studies. Study 1 examines how students use the digital environment when mathematics practice is no longer tied entirely to taught lessons. Study 2 examines whether the same environment can support earlier and more targeted remediation.

6.1 Study 1: Three Years of Flexible Mathematics Practice

Study 1 follows eight lower-cycle classes, with EMERALD used during the autonomous study hours in grades 7 and 8 and follow-up data collected in grade 9. Two upper-cycle classes without autonomous study hours also participate, with the teacher allocating approximately one weekly hour per class to the digital environment. Across these ten classes, the number of students potentially accessible to the study ranges from approximately 40 to 220 per year. The primary longitudinal analysis focuses on the grade-7 cohort, while the other classes provide additional information on patterns of use across grades.

Within mathematics, the autonomous study hours are designed to allow selected topics to be practised independently of the taught lesson. A working group of three mathematics teachers, including the first author, selects suitable topics. Topics are particularly suitable when their content is later revisited or required as a foundation for subsequent teaching. Gaps identified during autonomous study hours can therefore be revisited through ordinary classroom teaching.

Each study wave combines two short questionnaires, an adapted Intrinsic Motivation Inventory and the Abbreviated Math Anxiety Scale (Hopko et al., 2003), a curriculum-aligned pre-test, a teaching sequence using the environment, a post-test, and a delayed test. Self-regulation is not measured by a questionnaire scale: it is approached through the practice traces, the confidence rating students give before each test, and the end-of-session reflections. Pre-test and post-test are parallel forms, and no test item is drawn from the practice pool of the environment. Sporting load is obtained from the school's records of training and competition, which are already held for attendance purposes. End-of-sequence questionnaires and voluntary student interviews provide additional information about students' experience of the environment. Each year, interviews are also conducted with three to six collaborating teachers to examine implementation and perceived changes in practice.

Analysis follows the primary outcome. The traces characterise each student's temporal pattern of practice: the number and length of sessions, the time of day, the spacing between sessions, and how these patterns change in weeks containing training camps or competitions, with each student serving as their own comparison. Differences between students are then described in relation to the three variables named in RQ2. Attainment is examined from pre-test to post-test and delayed test, descriptively, since Study 1 has no randomised comparison group. Questionnaire subscales are summarised by wave, while the self-regulation indicators are derived from traces and the reflections. Interviews are analysed thematically. The design assumption is supported if practice shifts to periods compatible with training and competition, and questioned if greater flexibility reduces overall practice.

The primary focus of Study 1 is therefore not whether the environment replaces classroom teaching, but how it changes when, for how long, and under what conditions students practise mathematics.

6.2 Study 2: Remediation and the Moment of Diagnosis

Study 2 follows approximately 10 to 25 students per year across three existing forms of support: mandatory or voluntary remediation with the first author, optional support courses during the school year, and the national summer programme in which the first author teaches. The study is primarily concerned with the feasibility and timing of diagnosis rather than with estimating a causal effect of one remediation format over another.

The two studies are connected through diagnosis. In the current system, formal remediation decisions are based on end-of-year results. Although semester grades are available, they do not trigger mandatory support, and a failed first semester cannot be retaken. A learning gap that emerges during the year may therefore remain unaddressed until the end-of-year process.

EMERALD provides a different possibility. Because it maintains an evolving map of topics and learner estimates during ordinary practice, learning gaps can become visible during the school year rather than only at its end. Study 1 therefore provides the continuous diagnostic information, while Study 2 examines how that information can be used to provide earlier and more targeted support.

Across these routes, the study documents the use of two remediation formats: digital remediation sequences tailored to identified gaps, with immediate feedback, and traditional paper-based assignments. Generative AI makes completion of a paper assignment a less reliable indicator of independent practice, an issue raised nationally in a parliamentary question and the ministerial response to it (Chambre des Députés, 2026).

The formal remediation decision remains tied to end-of-year results and cannot be changed by the first author. Study 2 therefore does not attempt to change the institutional decision process. Instead, it examines whether earlier diagnosis and targeted support are feasible and sufficiently promising to address gaps before the end-of-year process begins and therefore justify considering earlier remediation at the institutional level.

7. Discussion

In a setting where attendance is structurally irregular, the contribution of a digital environment is not necessarily that it teaches mathematics better, but that it can decouple practice from the timetable. This decoupling is only useful if students receive enough support to use the additional autonomy effectively. In EMERALD, this support includes a map of topics and prerequisites, targeted practice sequences, and immediate feedback. Taught mathematics lessons remain the safety net alongside the autonomous study hours.

The decision to introduce autonomous study hours was made by the school development unit and applies to all students in grades 7 and 8. The project therefore studies an institutional change in which the first author is directly involved. The autonomous study hours themselves are a second attempt: they ran in 2025/26 without a digital practice environment and, on the school's own assessment, did not achieve their intended purpose. The 2026/27 redesign links them to the school's learning management system and to EMERALD.

The design also recognises that greater autonomy may not benefit all students equally. Students who have difficulty regulating their own learning may make less effective use of the autonomous study hours and, like all students in these year groups, have one fewer taught mathematics lesson each week. Ineffective use of the autonomous study hours is therefore not neutral.

8. Limitations

This paper reports the framework, study designs, and design decisions underlying the project. It reports no empirical results. The most important limitation concerns attribution. Study 1 has no randomised comparison group, so changes in attainment or motivation cannot be attributed causally to the environment. The primary outcome is therefore the within-student pattern of practice. Study 2 compares routes that students enter through existing institutional mechanisms, so differences between routes may reflect the students who enter them.

The participants are student-athletes in a single school. The setting makes the study particularly relevant to schools where attendance is affected by training and competition. However, this limits the generalisability of the findings. This mismatch is not confined to sport: illness, caring responsibilities and family mobility can produce the same mismatch between when a lesson is given and when a student is available. The first author is also the designer of the environment and the teacher of the participating classes, so the account of the design decisions documents the reasoning behind them rather than verifying it independently.

This dual role supports access and iterative design but may influence implementation and interpretation.

9. Next Steps

The ethics application covering both studies was submitted in August 2026 and is under review for the panel's September meeting. The national procedure governing research in Luxembourg schools also applies. Data collection will begin after the required authorisations have been obtained and will provide the first empirical test of the design assumptions presented in this paper. The successive waves follow the iterative logic of design-based research, in which design, implementation, analysis and refinement alternate in authentic educational settings (Brown, 1992; Anderson and Shattuck, 2012).

10. In Conclusion

This paper asks what educational technology can contribute when irregular attendance is a structural constraint rather than a reflection of mathematical ability. The proposed contribution is to shift not only where mathematics practice takes place, but when it takes place. EMERALD moves practice into autonomous study hours in which students have greater control over what and when they practise.

The two studies examine what happens after this shift. Study 1 follows when, for how long, and how students practise. Study 2 examines whether the same environment can make learning gaps visible early enough for remediation to begin before the end of the school year.

The contribution of the project is therefore not a claim that digital learning improves attainment. It is a design proposal and an empirical programme for studying when mathematics is practised, how students use autonomous study hours, and whether earlier diagnosis can shift when targeted support becomes possible.

Acknowledgements

The first author thanks his family; Prof. Dr Yves Kreis, his doctoral supervisor; Christian Meyers, for his support in the early stages; Robert A.P. Reuter, for critical feedback on the theoretical framework; the school headmaster, Pascal Schaul, for supporting conference participation; and Anne Achten and Clara Popescu, who contribute videos and printed formula summaries to the autonomous study hours.

Ethics Declaration

This paper reports no empirical data from human participants. It presents the theoretical framework, study protocols, and design decisions developed before data collection begins.

AI Declaration

AI tools were used to help structure material from the first author's existing French-language documents and design register, to discuss and refine the presentation of design choices, to translate, edit, and refine passages into English, and to check reference formatting. The research questions, theoretical framework, study designs, design decisions, and underlying reasoning reported in this paper were developed by the first author. The author reviewed and revised all text and takes full responsibility for its content.

References

Anderson, T. and Shattuck, J. (2012) *Design-Based Research: A Decade of Progress in Education Research?*, Educational Researcher, Vol 41, No 1, pp 16-25, [online], https://doi.org/10.3102/0013189X11428813.

Black, P. and Wiliam, D. (1998) *Assessment and Classroom Learning*, Assessment in Education: Principles, Policy and Practice, Vol 5, No 1, pp 7-74, [online], https://doi.org/10.1080/0969595980050102.

Bray, A. and Tangney, B. (2017) *Technology Usage in Mathematics Education Research: A Systematic Review of Recent Trends*, Computers and Education, Vol 114, pp 255-273, [online], https://doi.org/10.1016/j.compedu.2017.07.004.

Brown, A. L. (1992) *Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings*, Journal of the Learning Sciences, Vol 2, No 2, pp 141-178, [online], https://doi.org/10.1207/s15327809jls0202_2.

Bull, S. and Kay, J. (2010) Open Learner Models, in Nkambou, R., Bourdeau, J. and Mizoguchi, R. (eds) *Advances in Intelligent Tutoring Systems*, Springer, Berlin, pp 301-322, [online], https://doi.org/10.1007/978-3-642-14363-2_15.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. and Rohrer, D. (2006) *Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis*, Psychological Bulletin, Vol 132, No 3, pp 354-380, [online], https://doi.org/10.1037/0033-2909.132.3.354.

Chambre des Députés (2026) *Question parlementaire n° 4110 : Évaluation des travaux de vacances et intelligence artificielle, et réponse du Ministre de l'Éducation nationale, de l'Enfance et de la Jeunesse*, Luxembourg, 3 July 2026, [online], https://men.public.lu/content/dam/men/fr/actualites/articles/questions-parlementaires/2026/07/qp-4110.pdf.

Collins, A., Brown, J. S. and Newman, S. E. (1989) *Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics*, in Resnick, L. B. (ed) Knowing, Learning, and Instruction, Erlbaum, Hillsdale NJ, pp 453-494.

Darmanova, Z., Abylkassymova, A. and Nurmukhamedova, Z. (2025) *A Systematic Review of Technology Use in Middle and High School Mathematics Education: Insights from Contextual, Methodological, and Evaluation Characteristics*, Frontiers in Education, Vol 10, 1644284, [online], https://doi.org/10.3389/feduc.2025.1644284.

Deci, E. L., Koestner, R. and Ryan, R. M. (1999) *A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation*, Psychological Bulletin, Vol 125, No 6, pp 627-668, [online], https://doi.org/10.1037/0033-2909.125.6.627.

Elo, A. E. (1978) *The Rating of Chessplayers, Past and Present*, Arco, New York.

European Commission (2025) *Communication from the Commission, Guidelines on Measures to Ensure a High Level of Privacy, Safety and Security for Minors Online, Pursuant to Article 28(4) of Regulation (EU) 2022/2065 (Digital Services Act)*, Official Journal of the European Union, C/2025/5519, 10 October 2025, [online], https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C_202505519.

Hopko, D. R., Mahadevan, R., Bare, R. L. and Hunt, M. K. (2003) *The Abbreviated Math Anxiety Scale (AMAS): Construction, Validity, and Reliability,* Assessment, Vol 10, No 2, pp 178-182, [online], https://doi.org/10.1177/1073191103010002008.

Klensch, P. R., Meyers, C., Reuter, R. A. P., Busana, G. and Kreis, Y. (2025a) *Reimagining Remediation in Secondary Mathematics through MathemaTIC*, presentation at Computer Algebra and Dynamic Geometry in Mathematics Education conference, Esch-sur-Alzette, Luxembourg, 9-11 September 2025, [online], https://hdl.handle.net/10993/66714.

Klensch, P. R., Meyers, C., Reuter, R. A. P., Busana, G. and Kreis, Y. (2025b) *Re-Enchanting Mathematics: Motivation and Digital Learning in Luxembourg's Secondary Education*, presentation at Luxembourg Educational Research Association conference, Esch-sur-Alzette, Luxembourg, 5 November 2025, [online], https://hdl.handle.net/10993/66712.

Klinkenberg, S., Straatemeier, M. and van der Maas, H. L. J. (2011) *Computer Adaptive Practice of Maths Ability Using a New Item Response Model for on the Fly Ability and Difficulty Estimation*, Computers and Education, Vol 57, No 2, pp 1813-1824, [online], https://doi.org/10.1016/j.compedu.2011.02.003.

Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge.

MENJE (2016) *Digital(4)education : lancement de MathemaTIC*, communiqué de presse, Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse, [online], https://men.public.lu/fr/actualites/communiques-conference-presse/2016/05/10-mathemaTIC.html.

MENJE (2020) *Enseignement secondaire classique. Statistiques globales et analyse des résultats scolaires, année scolaire 2018/2019*, Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse, Luxembourg, [online], https://script.lu/sites/default/files/publications/2020-10/script_A4_ESC_2020_finalWEB.pdf.

MENJE (2025a) *Enseignement secondaire classique. Statistiques globales et analyse des résultats scolaires, année scolaire 2023/2024*, Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse, [online], https://men.public.lu/dam-assets/catalogue-publications/statistiques-etudes/secondaire/2023-2024-esc-statistiques-globales.pdf.

MENJE (2025b) *Rapport d'activité 2024*, Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse, [online], https://men.public.lu/dam-assets/catalogue-publications/rapports-dactivite-du-minist%C3%A8re/rapports-du-minist%C3%A8re/menje-rapport-activites-2024.pdf.

Pelánek, R. (2016) *Applications of the Elo Rating System in Adaptive Educational Systems*, Computers and Education, Vol 98, pp 169-179, [online], https://doi.org/10.1016/j.compedu.2016.03.017.

Règlement grand-ducal du 14 juillet 2005 déterminant l'évaluation et la promotion des élèves de l'enseignement secondaire et secondaire technique (2005) Luxembourg, [online], https://legilux.public.lu/eli/etat/leg/rgd/2005/07/14/n6/jo.

Rohrer, D. and Taylor, K. (2007) *The Shuffling of Mathematics Problems Improves Learning*, Instructional Science, Vol 35, No 6, pp 481-498, [online], https://doi.org/10.1007/s11251-007-9015-8.

Ryan, R. M. and Deci, E. L. (2000) *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being*, American Psychologist, Vol 55, No 1, pp 68-78, [online], https://doi.org/10.1037/0003-066X.55.1.68.

Sailer, M. and Homner, L. (2020) *The Gamification of Learning: A Meta-Analysis*, Educational Psychology Review, Vol 32, No 1, pp 77-112, [online], https://doi.org/10.1007/s10648-019-09498-w.

Skinner, B. F. (1958) *Teaching Machines*, Science, Vol 128, No 3330, pp 969-977, [online], https://doi.org/10.1126/science.128.3330.969.

Sweller, J. and Cooper, G. A. (1985) *The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra*, Cognition and Instruction, Vol 2, No 1, pp 59-89, [online], https://doi.org/10.1207/s1532690xci0201_3.
VanLehn, K. (2011) *The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems*, Educational Psychologist, Vol 46, No 4, pp 197-221, [online], https://doi.org/10.1080/00461520.2011.611369.
Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge MA.
Wilson, R. C., Shenhav, A., Straccia, M. and Cohen, J. D. (2019) *The Eighty Five Percent Rule for Optimal Learning*, Nature Communications, Vol 10, 4646, [online], https://doi.org/10.1038/s41467-019-12552-4.
Zimmerman, B. J. (2002) *Becoming a Self-Regulated Learner: An Overview*, Theory Into Practice, Vol 41, No 2, pp 64-70, [online], https://doi.org/10.1207/s15430421tip4102_2.

Beyond Engagement: A Human Flourishing Framework for Emotion-Aware AI in Education

Spyridon Kontis and Sofia Anastasiadou
Department of Midwifery, University of Western Macedonia, Kozani, Greece

dmw00034@uowm.gr
sanastasiadou@uowm.gr

Abstract: Emotion-aware artificial intelligence (AI) is increasingly used to infer learners' affective states, such as frustration, confusion, or engagement, and to adapt instruction in response. These capabilities may make learning environments more responsive, but they also raise questions about privacy, surveillance, bias, and the treatment of emotional experience as data. Evaluation has therefore moved beyond accuracy alone. Trustworthy AI stresses transparency and oversight; AI Ethics adds fairness and harm prevention; Human-Centred AI foregrounds learner agency. These perspectives are necessary, yet they do not fully settle a further educational question: against what longer-term human purpose should emotion-aware AI be judged? This conceptual paper proposes the Human Flourishing Framework (HFF) as a normative evaluative perspective for addressing that question. It is not presented as a computational layer or a causal model. Instead, it examines the conditions created around AI-mediated learning within a wider educational ecosystem. The framework organises five dimensions, Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, Fairness and Inclusion, and Human Flourishing, across Foundation, Empowerment, and Outcome layers. Its purpose is to examine whether the design and use of emotion-aware AI support or constrain learners' longer-term development, without attributing flourishing to the technology itself. The paper clarifies the framework's epistemological and theoretical positioning, distinguishes flourishing from related concepts such as capability, agency, and self-regulation, and translates the HFF dimensions into indicative system requirements. The contribution is conceptual rather than empirical. The framework has not yet been validated or operationalised as a measurement model; future work will therefore require longitudinal, multi-level, and cross-cultural research capable of studying AI-mediated learning within its wider educational context.

Keywords: Emotion-Aware Artificial Intelligence; Human Flourishing Framework; Trustworthy AI; AI Ethics; Human-Centred AI; Educational Technology Evaluation

1. Introduction

Artificial intelligence increasingly attempts to interpret how learners feel, not only what they know. Affect-aware systems use facial, vocal, textual, physiological, and interaction cues to infer states that may matter for learning, drawing on a long tradition of affective computing and research on achievement emotions (Picard 1997; Calvo and D'Mello 2010; Pekrun 2006).

In educational settings, such inferences can inform the timing of feedback, pacing, or support during frustration (Grawemeyer et al. 2017). The same capability is also difficult to treat as neutral. Emotional expression is context-sensitive and can be interpreted incorrectly, while the collection and inference of affective data may create surveillance and datafication concerns (Barrett et al. 2019; McStay 2020; Williamson and Eynon 2020).

As these concerns have become more visible, evaluation has widened from technical performance to questions of trust, ethics, and human agency. Trustworthy AI emphasises reliability, transparency, privacy, and oversight; AI Ethics adds questions of fairness, dignity, and justified use; Human-Centred AI asks whether people remain meaningful participants in AI-mediated decisions (European Commission High-Level Expert Group on Artificial Intelligence 2019; UNESCO 2021; Shneiderman 2020). These perspectives overlap rather than forming a simple sequence, and each addresses an important part of responsible educational AI.

A further question nevertheless remains. A system may be accurate, transparent, and ethically governed while leaving uncertain whether its sustained use is educationally worthwhile for the learner over time. This is especially important for emotion-aware AI, because repeated interpretation of affect can shape how learners understand their own responses, when they seek help, and how much authority they give to automated feedback (McStay 2020; Kontis and Anastasiadou 2026a).

Against this background, the paper asks: How can human flourishing provide a normative evaluative lens for emotion-aware AI in education beyond technical performance, ethical compliance, and short-term engagement? The Human Flourishing Framework (HFF) is proposed as a conceptual response. It builds on, rather than replaces, existing approaches and shifts attention toward the educational conditions surrounding AI use and their relevance to learners' longer-term development. Figure 1 summarises the conceptual move that motivates the framework.

2. Conceptual Foundations

2.1 Emotion-Aware AI in Education

Emotion-aware AI rests on a simple but consequential premise: learning is not purely cognitive. Curiosity, frustration, anxiety, and other affective states influence attention, persistence, and the way learners engage with difficult material (Pekrun 2006; D'Mello and Graesser 2012). Systems may infer such states from facial, vocal, textual, behavioural, or physiological signals and adapt their responses accordingly (Calvo and D'Mello 2010).

The inference itself, however, is probabilistic and context-dependent. Facial movement, for example, cannot be mapped mechanically onto discrete emotions across individuals and cultures (Barrett et al. 2019). An adaptive response based on a confident but mistaken interpretation may therefore be pedagogically inappropriate even when the underlying model performs well. Educational value depends not only on detection accuracy, but on how affective interpretations are situated, communicated, and used.

2.2 Trustworthy AI

Trustworthy AI provides a necessary baseline for this discussion. Robustness, transparency, privacy, accountability, and human oversight matter whenever automated inferences influence educational decisions (European Commission High-Level Expert Group on Artificial Intelligence 2019; Floridi and Cowls 2019; Anastasiadou 2026a). In emotion-aware learning, these requirements are particularly important because learners may not know that an affective inference has been made or how it has shaped the system's response. Trustworthiness helps establish whether reliance on a system is justified and governed, but it does not by itself specify the broader educational end that such reliance should serve.

2.3 AI Ethics and Human-Centered AI

AI Ethics extends the discussion to fairness, dignity, harm, and the justification of deployment. These issues become sharper when the data concern emotion, because inferred affect can be intimate, uncertain, and culturally variable (McStay 2020; Barrett et al. 2019; UNESCO 2021). Human-Centred AI adds another important concern: whether learners and educators can understand, question, and influence AI-mediated decisions instead of becoming passive recipients of them (Amershi et al. 2019; Shneiderman 2020).

Taken together, these perspectives establish strong protective and participatory conditions. Yet preserving agency within an interaction does not automatically answer what that interaction should contribute to over months or years of learning. The question of educational purpose remains, which is where human flourishing becomes relevant.

2.4 Human Flourishing

Human flourishing is used here as a broader developmental idea, not as a synonym for wellbeing, capability, autonomy, or self-regulation. The Capability Approach asks what people are genuinely able to do and become, rather than judging development only by resources or outcomes (Sen 1999; Nussbaum 2011). Self-Determination Theory identifies autonomy, competence, and relatedness as psychological conditions that support motivated development (Ryan and Deci 2020). Agency refers here to a learner's capacity to act, choose, question, and influence an educational interaction, whereas self-regulation concerns the more specific management of cognition, behaviour, and emotion. These concepts overlap, but they operate at different levels.

Flourishing provides the wider educational horizon within which these capacities acquire significance. Positive Education and recent policy work on education for human flourishing place meaning, development, relationships, and participation alongside academic achievement (Seligman et al. 2009; OECD 2025). In this paper, flourishing refers to the sustained development of learners as capable, autonomous, reflective, and purposeful participants in education and society. A system might improve short-term engagement or help regulate frustration without contributing to that broader trajectory. This distinction is central to the HFF: flourishing is an evaluative horizon, not a score or state produced by an AI system.

2.5 Current Educational AI Evaluation

Educational AI is commonly evaluated through measures such as performance, engagement, usability, trust, and fairness, all of which remain useful (Vistorte et al. 2024). For emotion-aware AI, however, short-term indicators

can leave important questions open. Does support become dependence? Can an affective label be challenged? Do repeated adaptations preserve the learner's own judgement? A flourishing-oriented perspective adds these questions of purpose and time without displacing established technical and ethical criteria.

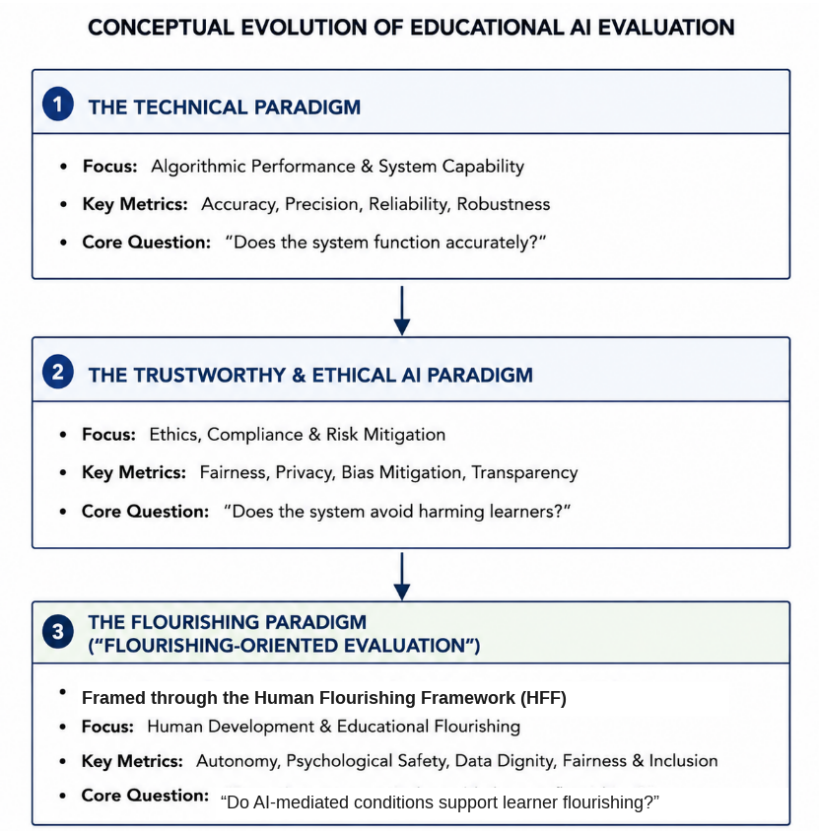


Figure 1: Conceptual evolution of educational AI evaluation from technical performance toward the Human Flourishing Framework

3. The Human Flourishing Framework

3.1 Why a Human Flourishing Framework?

The rationale for the HFF is not that existing frameworks have failed. They answer different questions. Trustworthy and ethical approaches ask whether systems are reliable, governed, and acceptable; human-centred approaches ask whether people retain meaningful agency. Human flourishing adds a longer educational horizon, asking whether the conditions associated with AI use are compatible with learners becoming more capable, autonomous, and able to pursue meaningful learning over time (Nussbaum 2011; Ryan and Deci 2020; OECD 2025).

That question cannot be located in an isolated learner-AI relationship. Educational development takes place within a wider socio-technical ecosystem involving learners, educators, pedagogical choices, institutional arrangements, governance, and cultural context. Emotion-aware AI is one element of that ecosystem, not an independent developmental agent. The HFF therefore appraises AI-mediated educational conditions within this broader setting and does not assume that learner flourishing can be causally attributed to the technology alone.

3.2 Conceptual Architecture of the Human Flourishing Framework

The HFF is not intended as a checklist or causal pathway. It is a layered architecture for organising questions about the conditions under which emotion-aware AI is designed, used, and evaluated. Flourishing is treated as a developmental orientation that may be supported or constrained within AI-mediated educational contexts, rather than as a property that a system possesses.

Three layers structure the framework. The Foundation Layer brings together the trustworthy and ethical conditions required for responsible use, including transparency, privacy, accountability, oversight, and

governance (European Commission High-Level Expert Group on Artificial Intelligence 2019; UNESCO 2021). The Empowerment Layer examines the learner's lived interaction through four HFF dimensions: Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, and Fairness and Inclusion. The Outcome Layer places Human Flourishing as the longer-term educational horizon against which those conditions are considered.

The layers are related but should not be read as a deterministic sequence. A trustworthy foundation does not guarantee empowerment, and empowerment does not guarantee flourishing. Their value lies in making different evaluative questions visible and in showing where technical, ethical, pedagogical, and developmental concerns meet.

Figure 2 presents this architecture and locates the HFF dimensions within the three layers through which conditions relevant to flourishing are examined.

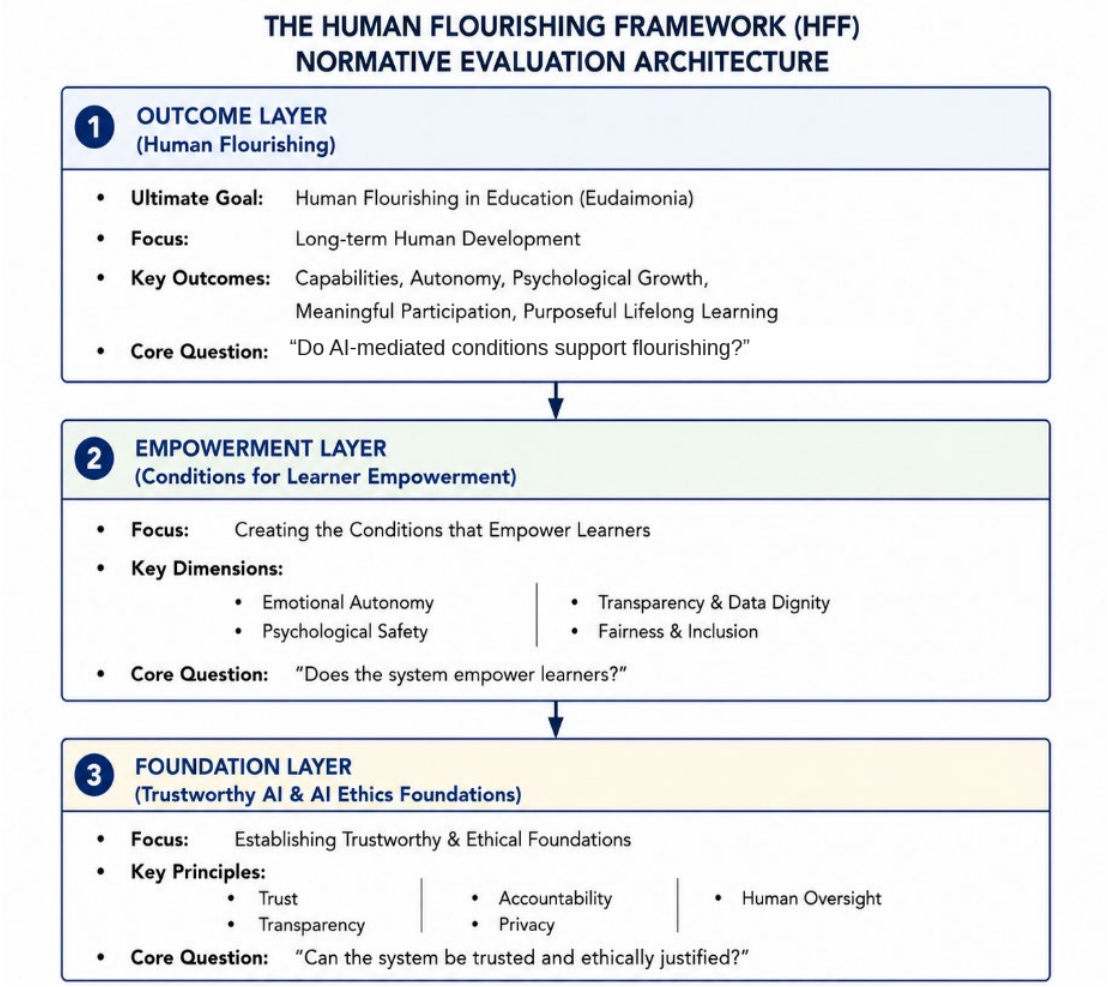


Figure 2: Conceptual architecture of the Human Flourishing Framework (HFF), showing how foundational principles support learner empowerment and flourishing-oreinted evaluation.

3.3 HFF Dimensions

3.3.1 Emotional Autonomy

Emotional Autonomy concerns the learner's capacity to retain meaningful authority over the interpretation and regulation of their own emotional experience during AI-mediated learning. It is distinct from self-regulation, which concerns the management of emotion, and from agency, which refers more broadly to the capacity to act, question, choose, and influence the educational interaction. Emotion-aware AI may support reflection, but an inference should not become a verdict about how a learner feels. Learners should be able to question, override, or disregard affective recommendations when they do not fit their experience (Ryan and Deci 2020; Shneiderman 2020).

3.3.2 Psychological Safety

Psychological Safety asks whether learners can experience uncertainty, difficulty, and error without feeling judged or penalised by affective monitoring. The concept originated in research on interpersonal learning climates, but its underlying concern is relevant when educational systems continuously observe behaviour and infer internal states (Edmondson 1999; McStay 2020). In this setting, safe design means avoiding punitive or high-stakes uses of inferred emotion, limiting unnecessary monitoring, and preserving space for learners to make mistakes without adjusting their behaviour simply to satisfy perceived algorithmic expectations.

3.3.3 Transparency and Data Dignity

Transparency and Data Dignity concern what learners know and can influence about the collection and use of emotional information. Here, data dignity means treating affective data and inferences as closely connected to personhood and vulnerability, rather than as ordinary behavioural features (McStay 2020; UNESCO 2021). Learners should know what is collected or inferred, why it is used, how long it is retained, and how it influences system responses. Transparency is therefore meaningful only when learners can also question or contest an interpretation that affects them.

3.3.4 Fairness and Inclusion

Fairness and Inclusion concern whether emotion-aware AI works equitably across learners whose emotional expression differs by culture, language, context, or neurodiversity. Research on facial expression cautions against assuming that outward behaviour provides a universal reading of internal emotion (Barrett et al. 2019). Fair design should therefore test affective models across diverse populations and provide ways to correct or contest misclassification, rather than treating difference as error.

3.3.5 Human Flourishing

Human Flourishing functions as the framework's evaluative horizon rather than as an additional system output. It concerns whether the wider pattern of AI-mediated learning is compatible with learners' longer-term development as autonomous, capable, reflective, and purposeful participants. The preceding dimensions identify conditions that may support such development, but none is sufficient on its own and no direct causal claim is made. Evidence about flourishing would require attention to time, context, pedagogy, and the wider educational ecosystem (Nussbaum 2011; OECD 2025).

3.4 Epistemological and Theoretical Positioning of the Human Flourishing Framework

The HFF adopts an interpretive and humanistic position toward emotion and educational development. Learners' emotions are not treated as fixed internal states that can be transparently read from data. AI outputs are probabilistic interpretations shaped by training data, modelling choices, behaviour, and context; the learner's own account and the educator's contextual judgement therefore retain epistemic weight (Barrett et al. 2019; McStay 2020). Human flourishing is approached in the same spirit: as a normative educational orientation, not an objective variable generated by an algorithm.

The normative dimension of the HFF is therefore external to the algorithm. It is a perspective used by designers, educators, institutions, and researchers when deciding which objectives should be pursued, which interventions are educationally justified, and which consequences should count as desirable. Normativity resides in these human criteria for design, deployment, interpretation, and evaluation, not in the data or algorithm as such. In this sense, the HFF complements Trustworthy AI, AI Ethics, and Human-Centred AI rather than adding a computational layer to the system (Floridi and Cowls 2019; Shneiderman 2020).

4. Implications

4.1 Implications for Educational AI Design

The practical value of the HFF depends on whether its normative commitments can inform concrete design decisions. It does not provide engineering specifications or universal rules. Instead, each dimension can be translated into an indicative system requirement that helps designers and institutions ask what follows from an affective inference, who can challenge it, and what kind of learner-system relationship repeated use may create. Table 1 summarises these requirements.

Table 1: Indicative system requirements derived from the HFF.

HFF Dimension	Indicative System Requirement
Emotional Autonomy	Allow learners to question, override, or ignore affective interpretations; use AI to support reflection rather than replace learner judgement.
Psychological Safety	Avoid punitive or high-stakes action based on inferred emotion; minimize unnecessary monitoring and conditions that encourage self-censorship.
Transparency and Data Dignity	Explain what affective data are collected or inferred, why they are used, how long they are retained, and how they influence system responses; provide meaningful controls.
Fairness and Inclusion	Test affective interpretation across culturally, linguistically, and neurodiverse populations and provide mechanisms to correct or contest misclassification.
Human Flourishing	Evaluate success beyond engagement or immediate performance, including autonomy, capability development, meaningful participation, and dependence on AI support over time.

For designers, the key question is therefore not only whether an affective inference is accurate, but what educational action follows from it, whether the learner can contest it, and whether repeated use strengthens rather than displaces human judgement (Sadek et al. 2024).

4.2 Implications for Educational Practice

For educational practice, emotion-aware AI is better treated as a source of provisional pedagogical insight than as a substitute for judgement. An affective inference may prompt an educator to look more closely, but it cannot provide the relational and contextual knowledge needed to understand a learner's experience on its own (Holstein and Aleven 2022; Celik 2023). The practical question is therefore not simply whether the system detects frustration, but how a teacher or learner responds to that suggestion and whether support gradually strengthens independent judgement rather than dependence on automated interpretation.

4.3 Implications for Research and Educational AI Governance

For research, the HFF offers a vocabulary for examining developmental questions alongside established measures of performance, engagement, trust, and usability. Its dimensions are not yet validated constructs, and future studies will need to examine them within the wider educational setting rather than attempting to isolate AI as a single cause. Governance can use the framework in a similar way. Existing requirements for safety, privacy, transparency, and accountability remain essential; the HFF adds an external educational-purpose test concerned with whether deployment choices are compatible with learner agency and longer-term development (UNESCO 2021; Miao and Holmes 2023).

5. Contribution, Limitations and Future Research

5.1 Theoretical Contribution

The contribution of this paper is integrative rather than the invention of a new set of ethical principles. The HFF brings Trustworthy AI, AI Ethics, Human-Centred AI, and Human Flourishing into a single educational evaluative architecture and makes two issues more explicit: educational purpose and the time horizon over which AI use is judged. It reframes evaluation from system performance and ethical acceptability alone toward the conditions that AI-mediated learning may create for human development. As a conceptual framework, its claim is one of coherence and usefulness, not empirical superiority over existing approaches.

5.2 Limitations

Several limitations follow from that position. The HFF has not been empirically validated, and its dimensions remain conceptual rather than operationalised measurement constructs. Flourishing is also culturally and philosophically contested; the framework draws substantially on the Capability Approach, Self-Determination Theory, Positive Education, and OECD work, traditions that may not transfer unchanged across educational settings. Nor does the framework imply that flourishing can be causally attributed to an AI system in isolation.

Learner development is shaped by interacting pedagogical, relational, institutional, and cultural factors. Finally, concepts such as Emotional Autonomy and Data Dignity require further operational clarification before they can support comparative measurement.

5.3 Future Research

The immediate research task is to operationalise the HFF dimensions using established measures where appropriate and new items where emotion-aware AI introduces distinct concerns. Longitudinal and multi-level designs are especially important because developmental change unfolds over time and across learner, classroom, and institutional levels. Mixed-method work could combine behavioural or survey measures with learner accounts of how affective interpretations are experienced and contested. Cross-cultural studies are also needed, together with design-based research examining whether the indicative system requirements proposed here remain useful across intelligent tutoring systems, adaptive platforms, and generative AI environments.

6. Conclusion

Emotion-aware AI makes a difficult educational question visible: a system can respond to affect without knowing what that response should ultimately be for. Trustworthy AI, AI Ethics, and Human-Centred AI provide indispensable conditions for responsible use, but educational purpose still needs to be articulated. The Human Flourishing Framework offers one way of doing so while avoiding the claim that technology itself produces flourishing.

Its central shift is from asking only what an intelligent system can do to asking what kinds of educational conditions its design and use help create. Emotional Autonomy, Psychological Safety, Transparency and Data Dignity, and Fairness and Inclusion make that question more concrete, while Human Flourishing keeps the longer developmental horizon in view.

The framework should now be tested, challenged, and refined. Its value will depend less on whether it predicts a single outcome than on whether it helps researchers, educators, designers, and institutions recognise when technically successful systems support, or quietly undermine, the human purposes education is meant to serve.

Ethics Declaration

This conceptual study did not involve human participants, personal data, or animal subjects. Therefore, ethical approval was not required.

AI Declaration

Generative AI tools were used for language editing, structural review, and formatting support. The authors retained responsibility for source selection and verification, conceptual decisions, interpretation of the literature, and the final manuscript.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fournier-Tombs, E., Horvitz, E., Johnston, S., Kikin-Gil, R., McDermott, K., Penrose, C., Morris, M.R. and Olsen, J. (2019) "Guidelines for human-AI interaction", Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, 4-9 May 2019, ACM, New York, pp. 1-13. https://doi.org/10.1145/3290605.3300229

Anastasiadou, S.D. (2026a) "H-IoT Ethics Instrument (H-IoTEI): Development, Validity, and Reliability", in Anastasiadou, S., Seremeti, L. and Papalexandris, S. (eds.), Bioethics and Legal Aspects of AI in Healthcare and Life Sciences, IGI Global Scientific Publishing, Hershey, PA, pp. 225-236. https://doi.org/10.4018/979-8-3373-5646-4.ch010

Barrett, L.F., Adolphs, R., Martinez, A., Ide, D.L. and Russell, J.A. (2019) "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements", Psychological Science in the Public Interest, Vol. 20, No. 1, pp. 1-68. https://doi.org/10.1177/1529100619832930

Calvo, R.A. and D'Mello, S. (2010) "Affect detection and an overview of expressing affective states", IEEE Transactions on Affective Computing, Vol. 1, No. 1, pp. 18-37. https://doi.org/10.1109/TAFFC.2010.2

Celik, I. (2023) "Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education", Computers in Human Behavior, Vol. 138, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

D'Mello, S. and Graesser, A. (2012) "Dynamics of affective states during complex learning", Learning and Instruction, Vol. 22, No. 2, pp. 145-157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Edmondson, A. (1999) "Psychological safety and learning behavior in work teams", Administrative Science Quarterly, Vol. 44, No. 2, pp. 350-383. https://doi.org/10.2307/2666999

European Commission High-Level Expert Group on Artificial Intelligence (2019) Ethics Guidelines for Trustworthy AI, European Commission, Brussels.

Floridi, L. and Cowls, J. (2019) "A unified framework of five core principles for AI in society", Harvard Data Science Review, Vol. 1, No. 1. https://doi.org/10.1162/99608f92.8cd550ed

Grawemeyer, B., Mavrikis, M., Holmes, W., Hansen, A., Loibl, K. and Gutierrez-Santos, S. (2017) "Affective transitions in open-ended learning environments", IEEE Transactions on Learning Technologies, Vol. 10, No. 3, pp. 315-326. https://doi.org/10.1109/TLT.2016.2614810

Holstein, K. and Aleven, V. (2022) "Designing for human-AI complementarity in K-12 classrooms", in International Conference on Artificial Intelligence in Education, Springer, Cham, pp. 2-13. https://doi.org/10.1007/978-3-031-11644-5_1

Kontis, S. and Anastasiadou, S. (2026a) "Do Machines Feel?: The Emergence of Artificial Qualia and the Ethics of Emotion-Aware AI", in Seremeti, L., Anastasiadis, L. and Anastasiadou, S. (eds.), Philosophical Considerations of Computational Consciousness and AI Qualia, IGI Global Scientific Publishing, Hershey, PA, pp. 223-246. https://doi.org/10.4018/979-8-3373-5530-6.ch007

McStay, A. (2020) "Emotional AI and education: assessment of an electronic surveillance ecosystem", Policy Insights from the Behavioral and Brain Sciences, Vol. 7, No. 2, pp. 158-165. https://doi.org/10.1177/2372732220943222

Miao, F. and Holmes, W. (2023) Guidance for Generative AI in Education and Research, UNESCO Publishing, Paris.

Nussbaum, M.C. (2011) Creating Capabilities: The Human Development Approach, Harvard University Press, Cambridge, MA.

OECD (2025) Education for Human Flourishing, OECD Publishing, Paris.

Pekrun, R. (2006) "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice", Educational Psychology Review, Vol. 18, No. 4, pp. 315-341. https://doi.org/10.1007/s10648-006-9029-9

Picard, R.W. (1997) Affective Computing, MIT Press, Cambridge, MA.

Ryan, R.M. and Deci, E.L. (2020) "Intrinsic and extrinsic motivation from a self-determination theory perspective", Contemporary Educational Psychology, Vol. 61, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sadek, M., Kallina, E., Bohné, T., Mougenot, C., Calvo, R.A. and Cave, S. (2024) "Challenges of responsible AI in practice: scoping review and recommended actions", AI & Society, Vol. 40, No. 1, pp. 45-62. https://doi.org/10.1007/s00146-024-01880-9

Seligman, M.E., Ernst, R.M., Gillham, J., Reivich, K. and Linkins, M. (2009) "Positive education: Positive psychology and classroom interventions", Oxford Review of Education, Vol. 35, No. 3, pp. 293-311. https://doi.org/10.1080/03054980902934563

Sen, A. (1999) Development as Freedom, Oxford University Press, Oxford.

Shneiderman, B. (2020) "Human-centered artificial intelligence: Reliable, safe & trustworthy", International Journal of Human-Computer Interaction, Vol. 36, No. 6, pp. 495-504. https://doi.org/10.1080/10447318.2020.1741118

UNESCO (2021) Recommendation on the Ethics of Artificial Intelligence, UNESCO Publishing, Paris.

Vistorte, A.Y.O., Deroncele-Acosta, A., Ayala, M., Barrasa, A., López-Granero, C. and Martí-González, M. (2024) "Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review", Frontiers in Psychology, Vol. 15, Article 1122356. https://doi.org/10.3389/fpsyg.2024.1122356

Williamson, B. and Eynon, R. (2020) "Historical threads, datafication, and the governance of education", Learning, Media and Technology, Vol. 45, No. 3, pp. 223-235. https://doi.org/10.1080/17439884.2020.1772226

Digitally Supported Formative Assessment and Achievement Motivation: Evidence from the REFRESH Project

Veronika Murinová and Dana Vicherková
Department of Pedagogy and Andragogy, Faculty of Education, University of Ostrava, Czech Republic

veronika.murinova@osu.cz
dana.vicherkova@osu.cz

Abstract: This study explores the impact of digitally supported formative assessment on changes in achievement motivation among higher vocational school students preparing for the teaching profession in pre-school and after-school education. The research is conducted within doctoral studies at the Faculty of Education, University of Ostrava, and is linked to the REFRESH – Centre for Research on Energy and Social Change project (CZ.10.03.01/00/22_003/0000048). The research sample consisted of 136 students from four classes. A quasi-experimental design was employed, including three experimental groups using digital formative assessment tools (e.g., Microsoft Forms, Mentimeter, self-assessment and peer assessment tools) and one control group receiving traditional instruction. Due to natural attrition during the study, the data were analysed as non-paired. Achievement motivation was assessed using the School Performance Motivation Questionnaire (Hrabal & Pavelková, 2011), focusing on the need for achievement (NfA) and avoidance of failure (NAF). The results show an increase in the need for achievement, while changes in avoidance of failure differed across groups among students exposed to digitally supported formative assessment. Differences between experimental and control groups indicate a positive association between formative assessment and the development of achievement motivation. Further analysis suggests variation in the intervention effect between students from mining family backgrounds and other students. These findings are consistent with prior research highlighting the role of formative assessment and feedback in enhancing student motivation and engagement (Black & Wiliam, 2009; Dunn & Mulvenon, 2009; Fong et al., 2019; Winstone & Carless, 2020) and extend current knowledge by providing a more detailed analysis of motivational components within a digitally enriched learning environment in initial teacher education. The study offers practical implications for educators, researchers, and policy makers seeking to design effective technology-supported feedback environments in teacher preparation.

Keywords: Formative Assessment, Digital Feedback Tools, Achievement Motivation, Pre-Service Teacher Education, Educational Technology

1. Introduction

The landscape of contemporary education is increasingly shaped by digital technologies. Such tools transform both the accessibility of knowledge and the mechanisms through which student learning is assessed. Within the spheres of vocational and teacher education, these shifts present novel challenges regarding student engagement and self-regulation. Developing performance motivation is particularly critical for pre-service teachers, as it significantly influences their own learning processes as well as their future professional practice.

Formative assessment (FA) is defined as a continuous process of collecting and interpreting information about student progress, with the primary aim of adapting instruction to meet specific learner needs (Black and Wiliam, 2009). Previous research has consistently shown that FA increases student engagement and supports a deeper understanding of the subject matter. When digital tools are integrated into FA, they create enhanced opportunities for immediate feedback and personalised learning trajectories (Winstone and Carless, 2020). Digital platforms, such as Microsoft Forms, Mentimeter, and various self-assessment tools, enable real-time interactivity that was previously difficult to achieve in larger or more diverse learning environments.

Despite the documented benefits of digitally supported formative assessment, there remains a lack of empirical research specifically situated within the context of initial teacher education. This gap is particularly significant in vocational education, where student populations are often heterogeneous. The present study is conducted as part of the REFRESH project (Research Excellence for Region Sustainability and High-tech), which examines social and technological changes in regions undergoing profound economic transformation. In such contexts, students' motivational profiles are frequently shaped by diverse socio-economic backgrounds, including families with a multi-generational history in the mining industry.

The aim of this study is to investigate the impact of digitally supported formative assessment on the performance motivation of students in higher vocational education. The research focuses on two key components of performance motivation: the need for successful performance (NfA) and the need to avoid failure (NAF), as conceptualised by Hrabal and Pavelková (2011). Utilising a quasi-experimental design, the study compares students exposed to systematic digital formative assessment with a control group receiving traditional instruction.

This paper contributes to the ongoing academic discourse on the role of feedback in digitally enriched learning environments. It provides empirical evidence on how technology-supported formative assessment can enhance students' performance motivation within specific socio-economic contexts. Ultimately, the findings offer practical implications for educators and policy makers involved in designing effective feedback practices for modern teacher education programmes.

2. Research Design and Hypotheses

The primary objective of this study is to evaluate the impact of digitally supported formative assessment (FA) on students' performance motivation within a quasi-experimental framework. The research design is situated within the specific socio-economic context of the REFRESH project, which examines educational adaptations in regions undergoing industrial transformation. By comparing experimental groups exposed to systematic digital feedback with a control group receiving traditional instruction, the study seeks to explore potential associations between assessment methods and motivational shifts.

2.1 Research Questions

To achieve the research objectives, the following research questions (RQ) were formulated:

- RQ1: How do changes in performance motivation differ between students exposed to digitally supported formative assessment and those taught using traditional instructional methods?
- RQ2: How do the individual components of performance motivation—specifically the need for successful performance (NfA) and the need to avoid failure (NAF)—develop in response to the systematic application of digital formative assessment?
- RQ3: Does the effect of digitally supported formative assessment on performance motivation differ between students from mining backgrounds and those from other socio-economic backgrounds?

2.2 Hypotheses

Based on the theoretical framework and the research questions, the following hypotheses (H) were formulated as analytical assumptions guiding the descriptive comparison of motivational development across groups.

- H1: Students exposed to digitally supported formative assessment will demonstrate a higher level of need for successful performance (NfA) at the end of the intervention compared to students in the control group.
- H2: Students exposed to digitally supported formative assessment will demonstrate a lower level of need to avoid failure (NAF) compared to students in the control group.
- H3: The overall positive shift in performance motivation (characterised by an increase in NfA and a decrease in NAF) will be more pronounced in the experimental groups than in the control group.
- H4: The impact of digitally supported formative assessment on performance motivation will differ between students from mining backgrounds and students from other backgrounds, reflecting the interaction between the intervention and specific socio-economic factors.

3. Theoretical Background

The theoretical framework of this study integrates three key domains: performance motivation, digital formative assessment, and the socio-economic context of vocational teacher education. In the era of Society 4.0, the preparation of future teachers must transcend traditional pedagogical methods to foster digital agency and self-regulated learning.

3.1 Performance Motivation and the Impact of Feedback Timeliness

Initial teacher education (ITE) represents a specific environment where professional competencies are formed, heavily influenced by students' performance motivation. According to Hrabal and Pavelková (2011), this motivation consists of the need for successful performance (NfA) and the need to avoid failure (NAF). Recent research by Fisher, Brotto and Lim (2025) identifies motivation as a key driver of learning, which can be quickly enhanced or diminished by the nature of feedback.

Crucially, the timeliness of feedback is a significant determinant of student motivation. Fisher et al (2025) found that students express significantly lower levels of motivation if feedback takes longer than 10 days to be delivered. While students do not necessarily distinguish between 'timely' (e.g. 7 days) and 'extremely timely' (e.g. less than 5 days) feedback, exceeding the 10-day threshold leads to a marked motivational decline. Furthermore, Solis Trujillo et al (2025) argue that timely and specific feedback is essential for reducing student anxiety, which directly relates to mitigating the NAF component and preventing avoidance behaviour.

3.2 Digital Formative Assessment and Personalised Feedback Loops

Formative assessment (FA) is defined as a continuous process of collecting evidence of learning to adapt instruction (Black and Wiliam, 2009). Modern digital tools, such as Microsoft Forms and Mentimeter, facilitate this process by enabling immediate interactivity. Solis Trujillo et al (2025) highlight that technological tools empower students to manage their learning more effectively by providing personalised and accessible feedback that tailors the experience to individual needs.

However, the content of feedback remains the most impactful factor. Fisher et al (2025) report that 57% of students prioritise the quality of feedback content, particularly appreciating thorough, personalised, and specific assessments of their strengths and weaknesses. This aligns with the concept of feedback literacy (Winstone and Carless, 2020), where students develop the capacity to use feedback for self-regulation. Digital FA platforms support this by creating transparent feedback loops that allow for feed-forward—providing students with actionable information on what to do next to build on their current performance (Fisher et al, 2025; Palmberg et al, 2024).

3.3 Socio-Economic Context and Educational Equity

The motivational effects of digital FA are deeply intertwined with the students' social environment. Within the REFRESH project, this study focuses on students from mining backgrounds, for whom industrial transformation presents unique career challenges. Solis Trujillo et al (2025) emphasise that feedback strategies that consider individual differences contribute to greater equity and effectiveness in education. By fostering a sense of competence and autonomy (Deci and Ryan, 2000), digitally supported FA can mitigate the performance strain (NAF) often found in students from regions undergoing economic transition.

Effective formative assessment, supported by high-quality teaching and timely digital interventions, provides the necessary emotional and cognitive support to shift students towards a positive performance type (Type 1) as defined by Hrabal and Pavelková (2011). Thus, the integration of technology is not merely a technical enhancement but a pedagogical strategy to strengthen engagement and professional resilience in diverse student populations.

4. Methodology

This study employs a quasi-experimental research design with a non-equivalent control group and two distinct measurement points: an entry measurement (T1) and an exit measurement (T2). This design was selected to reflect the authentic conditions of educational practice, where random assignment of participants is not feasible. Due to anonymous data collection and natural attrition between measurements, individual responses could not be paired; consequently, the data were analysed as independent samples.

4.1 Participants and Setting

The research was conducted at a Czech higher vocational school within a three-year study programme preparing future teachers for pre-school and out-of-school education. The total sample comprised 136 students distributed across four classes: A (n = 37), B (n = 41), C (n = 29), and Year 2 (n = 26). Classes A, B, and the Year 2 group formed the experimental groups, in which digitally supported formative assessment was systematically implemented. Class C served as the control group and followed a traditional instructional approach without targeted formative intervention.

The Year 2 group represented a longitudinal case, as these students had been exposed to formative assessment strategies for two consecutive years. Their results are therefore interpreted separately, with particular emphasis on the effects of long-term exposure to formative practices. Furthermore, the sample included a subgroup of 18 students from mining backgrounds, reflecting the socio-economic focus of the REFRESH project.

4.2 Intervention

The intervention was integrated into regular teaching over the course of one academic year. It was grounded in the core principles of formative assessment: clarifying learning objectives, establishing transparent assessment criteria, providing continuous and actionable feedback, and fostering active student involvement. Digital tools, specifically Microsoft Forms and Mentimeter, were utilised to facilitate immediate feedback and support adaptive instruction. Additional strategies included self-assessment, peer assessment, and structured pedagogical dialogue, all aimed at enhancing students' self-regulation and engagement.

4.3 Instrumentation

Performance motivation was measured using the standardised instrument School Performance Motivation of Students (Hrabal and Pavelková, 2011). This questionnaire distinguishes between two primary dimensions: the need for successful performance (NfA) and the need to avoid failure (NAF). The instrument is well established in the Czech educational context and demonstrates satisfactory psychometric reliability. For the purposes of this study, raw scores for NfA and NAF were used for descriptive comparison. The standardized norms of the instrument informed the interpretation of motivational profiles.

4.4 Data Analysis

Data analysis was conducted descriptively. Mean scores and standard deviations were calculated for both NfA and NAF at T1 and T2. Due to anonymous data collection, individual responses could not be paired across measurements; therefore, the analysis focused on group-level comparisons of independent samples. Changes between T1 and T2 were interpreted using mean differences. No inferential statistical testing was applied, and the results are therefore interpreted cautiously as descriptive trends rather than causal effects.

4.5 Ethics and AI Declaration

Ethical approval for this research was obtained from the Ethics Committee of the Faculty of Education, University of Ostrava (Ref: REFRESH-2023-04). Participants were informed about the purpose of the study and their right to withdraw at any time.

In the preparation of this manuscript, an AI tool (large language model) was used for linguistic refinement and stylistic consistency. All AI-generated suggestions were carefully reviewed and revised by the authors to ensure academic integrity and factual accuracy.

5. Results

Descriptive statistics for all groups are presented in Table 1. Changes between the input measurement (T1) and output measurement (T2) are summarized in Table 2 and illustrated in Figures 1 - 3.

Table 1: Descriptive Statistics of Performance Motivation Across Groups at T1 and T2

Group	N T1	N T2	NfA T1 (M±SD)	NfA T2 (M±SD)	NAF T1 (M±SD)	NAF T2 (M±SD)
Class A	32	25	21.72 ± 2.65	22.75 ± 2.58	20.97 ± 3.67	23.36 ± 3.94
Class B	36	31	21.33 ± 3.02	22.23 ± 2.96	20.78 ± 4.06	21.90 ± 3.87
Class C	23	16	21.35 ± 2.41	22.25 ± 2.86	22.35 ± 3.64	20.94 ± 4.33
Year 2	26	26	21.23 ± 2.63	21.69 ± 2.87	20.73 ± 4.17	20.15 ± 4.35

Source: own elaboration 2026

5.1 Development of the Need for Successful Performance (NfA)

As shown in Table 1 and Figure 1, the mean values of the need for successful performance (NfA) increased across all groups between T1 and T2.

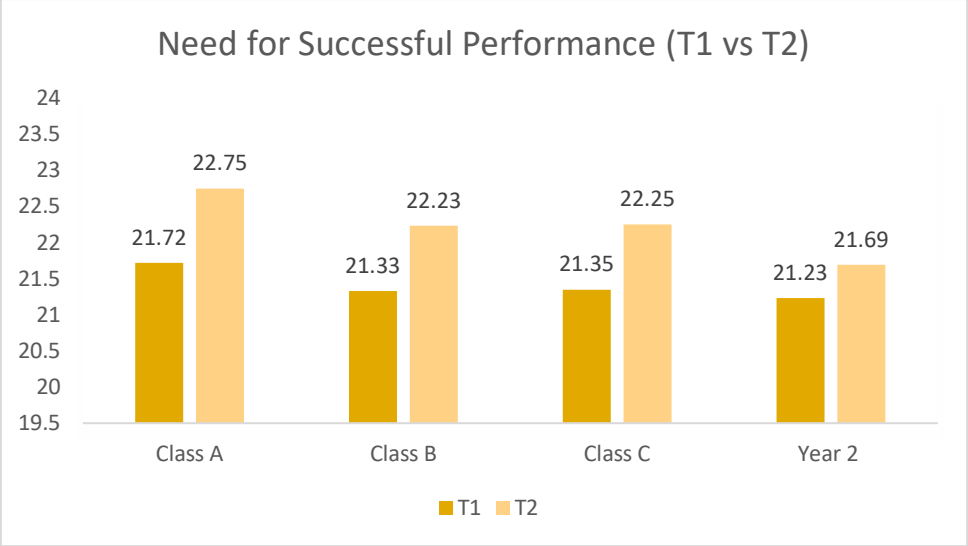


Figure 1: Need for successful performance (NfA) across groups at T1 and T2.

Source: own elaboration 2026

In Class A, the mean score increased from M = 21.72 (SD = 2.65) to M = 22.75 (SD = 2.58). A similar trend was observed in Class B (from M = 21.33, SD = 3.02 to M = 22.23, SD = 2.96) and Class C (from M = 21.35, SD = 2.41 to M = 22.25, SD = 2.86).A smaller increase was found in the second-year cohort, where NfA rose from M = 21.23 (SD = 2.63) to M = 21.69 (SD = 2.87).

Overall, an increase in performance-oriented motivation was observed across all groups.

5.2 Development of the Need to Avoid Failure (NAF)

The development of the need to avoid failure (NAF) showed a more differentiated pattern (Table 1, Figure 2).

In Classes A and B, NAF increased between T1 and T2. In Class A, the mean rose from M = 20.97 (SD = 3.67) to M = 23.36 (SD = 3.94), and in Class B from M = 20.78 (SD = 4.06) to M = 21.90 (SD = 3.87).

In contrast, a decrease in NAF was observed in Class C (from M = 22.35, SD = 3.64 to M = 20.94, SD = 4.33) and in the second-year cohort (from M = 20.73, SD = 4.17 to M = 20.15, SD = 4.35).

These results suggest that while performance motivation increased across all groups, changes in failure avoidance were not uniform.

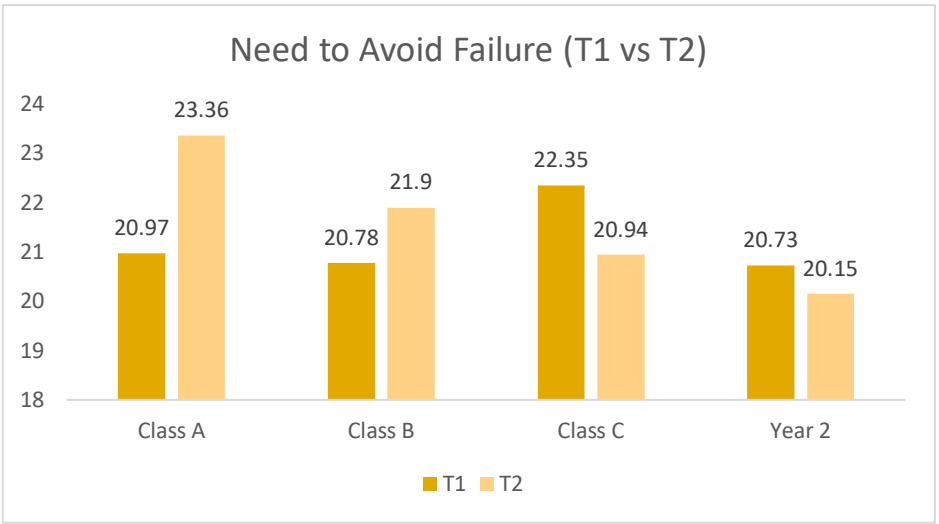


Figure 2: Need to avoid failure (NAF) across groups at T1 and T2.

Source: own elaboration 2026

5.3 Changes in Performance Motivation

The magnitude and direction of changes between T1 and T2 are summarized in Table 2 and illustrated in Figure 3.

Table 2: Changes in Performance Motivation Between T1 and T2

Group	Change in NfA	Change in NAF
Class A	+1.03	+2.39
Class B	+0.90	+1.12
Class C	+0.90	−1.41
Year 2	+0.46	−0.58

Source: own elaboration 2026

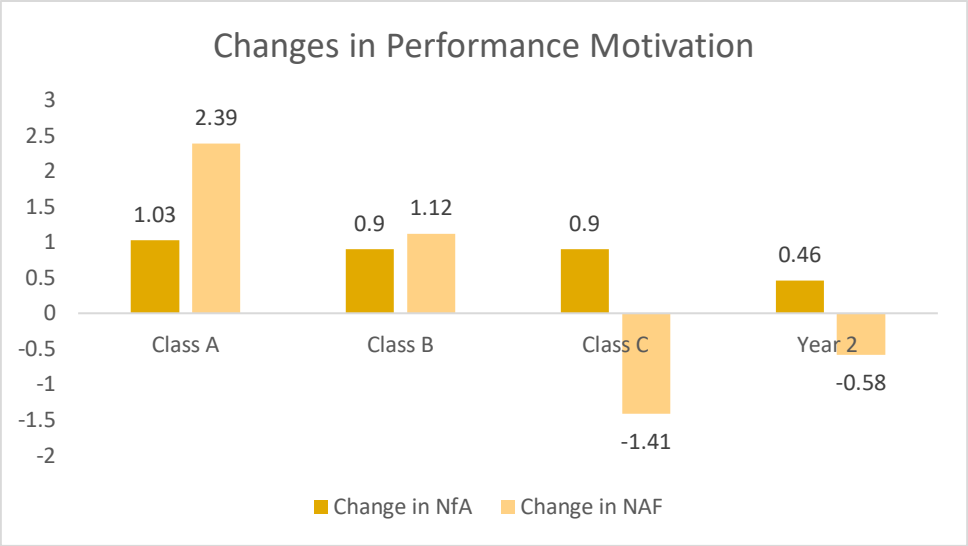


Figure 3: Changes in performance motivation between T1 and T2.

Source: own elaboration 2026

An increase in NfA was observed in all groups, with the largest increase in Class A (Δ = +1.03), followed by Classes B and C (Δ = +0.90), and the smallest increase in the second-year cohort (Δ = +0.46).

Changes in NAF differed across groups. While Classes A (Δ = +2.39) and B (Δ = +1.12) showed an increase, Classes C (Δ = −1.41) and the second-year cohort (Δ = −0.58) exhibited a decrease.

This pattern highlights variability in the development of performance motivation across different groups.

5.4 Mining Background Subgroup

A separate descriptive analysis was conducted for students from a mining background (N = 18), as presented in Table 3.

In this subgroup, the mean NfA increased from M = 21.22 at T1 to M = 22.88 at T2. The mean NAF showed a slight increase from M = 21.16 to M = 21.77.

Given the smaller sample size, these findings should be interpreted with caution.

Table 3: Performance Motivation of Students from Mining Backgrounds

Group	N	NfA T1	NfA T2	NAF T1	NAF T2
Mining background	18	21.22	22.88	21.16	21.77

Source: own elaboration 2026

6. Discussion

The aim of this study was to examine the impact of digitally supported formative assessment on the performance motivation of students in vocational teacher education, with a particular focus on differences between groups and a subgroup of students from a mining background.

6.1 Interpretation of the Main Findings (RQ1, H1)

The results indicate that the implementation of formative assessment supported by digital tools was associated with an increase in the need for successful performance (NfA) across all observed groups. This trend was evident not only in the experimental classes (A, B, and Year 2), but also, to a certain extent, in the control group (Class C).

These findings are consistent with the conceptualization of formative assessment as a process that supports learning through continuous feedback and adaptive instruction (Black & Wiliam, 2009). From the perspective of self-determination theory, the observed increase in performance-oriented motivation may reflect enhanced feelings of competence and autonomy supported by formative feedback (Deci & Ryan, 2000).

This finding is partially consistent with Hypothesis 1, which assumed that students exposed to formative intervention would show higher NfA and lower NAF after the intervention. While an increase in NfA was observed, the development of the need to avoid failure (NAF) was not consistent across groups. In some cases (Classes A and B), NAF increased, whereas in others (Class C and Year 2), it decreased.

These results suggest that formative assessment may effectively enhance students' performance orientation, but its impact on reducing fear of failure is more complex and context-dependent.

6.2 Differences Between Groups (RQ2, H2)

The comparison between groups with and without formative intervention reveals a differentiated pattern. Although Class C (control group) showed an increase in NfA, this change was not accompanied by a consistent decrease in NAF. This contrasts with some of the experimental groups, where a more balanced motivational shift was observed.

Therefore, the observed pattern is only partially consistent with Hypothesis 2, which assumed no substantial change in the control group. The observed increase in NfA in Class C may reflect external influences such as maturation, teaching style, or classroom climate, rather than the effect of formative assessment alone.

These findings highlight the importance of considering contextual and instructional factors when interpreting changes in student motivation.

6.3 Variability in Motivational Development

A key finding of this study is the variability in the development of performance motivation across groups. While all groups showed an increase in NfA, the direction of change in NAF differed, indicating that the relationship between performance ambition and fear of failure is not linear.

This variability suggests that formative assessment does not operate as a uniform intervention, but rather interacts with classroom-specific dynamics, including teacher practices, peer interactions, and the overall learning environment.

The variability observed across groups may be explained by differences in how students engage with feedback, as effective feedback requires active student interpretation and self-regulation (Nicol & Macfarlane-Dick, 2006). At the same time, the mixed development of NAF corresponds with findings that feedback interventions can have both positive and negative effects depending on how they are perceived and processed by learners (Kluger & DeNisi, 1996).

6.4 Students From a Mining Background (RQ3, H4)

The analysis of students from a mining background provides additional insight into the role of socio-economic context in shaping performance motivation. In the pretest, this subgroup exhibited slightly lower NfA and relatively higher NAF, which may reflect increased performance-related uncertainty.

Following the intervention, an increase in NfA was observed, accompanied by a slight increase in NAF. These findings are partially consistent with Hypothesis 4, suggesting that while formative assessment may enhance performance ambition in this subgroup, it does not necessarily reduce performance-related anxiety.

Given the limited sample size, these results should be interpreted with caution. However, they indicate that formative assessment may have the potential to support motivation even in students from more challenging socio-economic backgrounds, although its effects may differ in magnitude and direction.

6.5 Implications for Practice and Research

The findings of this study have several implications for educational practice. First, they suggest that formative assessment, particularly when supported by digital tools, can play a significant role in enhancing students' performance motivation. The use of tools such as Microsoft Forms and Mentimeter enables immediate feedback, supports student engagement, and facilitates adaptive teaching strategies.

These conclusions are consistent with the work of Hattie (2009), who identifies feedback as one of the most powerful influences on student learning and achievement. At the same time, the variability in NAF development highlights the need for a more nuanced approach to feedback. While increasing performance ambition is desirable, it is equally important to address students' fear of failure and create a psychologically safe learning environment.

These findings also align with the principles of self-determination theory (Deci & Ryan, 2000), which emphasize the importance of supporting students' sense of competence, autonomy, and relatedness in order to foster intrinsic motivation.

6.6 Limitations

Several limitations of the study should be acknowledged. The research design is based on non-equivalent groups, and the composition of respondents differs between T1 and T2, which limits the possibility of causal interpretation.

In addition, the analysis is primarily descriptive and does not include inferential statistical testing. The subgroup of students from a mining background is relatively small, which further limits the generalizability of the findings.

Future research should focus on longitudinal designs, larger samples, and the inclusion of additional variables, such as classroom climate or teacher feedback strategies, to better understand the mechanisms underlying motivational change.

7. Conclusion

This study examined the impact of digitally supported formative assessment on the performance motivation of students in vocational teacher education within the socio-economic context of the REFRESH project. Using a quasi-experimental design with non-equivalent groups, the research focused on the development of two dimensions of performance motivation: the need for successful performance (NfA) and the need to avoid failure (NAF).

The findings indicate that the systematic integration of digitally supported formative assessment was associated with positive changes in students' motivational profiles, particularly in relation to achievement-oriented motivation. Across all observed groups, an increase in NfA was identified between the entry (T1) and exit (T2) measurements, although the magnitude of change varied between groups. These results suggest that formative assessment supported by digital tools may contribute to strengthening students' engagement, perceived competence, and orientation towards successful performance.

However, the development of NAF demonstrated a more differentiated pattern. While some groups showed a decrease in failure avoidance, others exhibited a slight increase. This variability indicates that the relationship between formative feedback and performance-related anxiety is complex and influenced by contextual factors, including classroom climate, feedback interpretation, and students' individual motivational characteristics. Consequently, digitally supported formative assessment should not be viewed as a universally applicable intervention, but rather as a pedagogical process shaped by specific educational environments.

Particular attention was devoted to students from mining backgrounds, representing a subgroup connected to the broader socio-economic focus of the REFRESH project. Although these students demonstrated an increase

in NfA following the intervention, the persistence of elevated NAF values suggests that students from economically transforming regions may require longer-term motivational and emotional support. This finding highlights the importance of considering socio-economic context when implementing formative assessment strategies in vocational teacher education.

The study contributes to the growing body of research on formative assessment and digital feedback environments by extending current knowledge into the field of initial teacher education. It emphasises the pedagogical significance of immediate feedback, active student participation, self-assessment, and transparent assessment criteria in supporting motivational development and self-regulated learning. The use of digital tools such as Microsoft Forms and Mentimeter appears to create conditions that facilitate continuous feedback processes and adaptive instructional practices.

Several limitations of the study should be acknowledged. The research was based on a quasi-experimental design with non-equivalent groups and anonymous independent samples, which limits the possibility of causal interpretation. Furthermore, the subgroup of students from mining backgrounds was relatively small, and the analysis remained primarily descriptive. Future research should therefore focus on longitudinal designs, larger samples, and more detailed analyses of the interaction between formative assessment, classroom climate, and socio-economic variables influencing motivational development.

Despite these limitations, the findings suggest that digitally supported formative assessment represents a promising approach to enhancing performance motivation in contemporary teacher education. The study offers practical implications for educators, researchers, and policy makers seeking to develop inclusive, feedback-oriented, and digitally enriched learning environments that support both academic development and student well-being.

Acknowledgements

This research was supported by the REFRESH – Research Excellence For REgion Sustainability and High-tech project (CZ.10.03.01/00/22_003/0000048).

Ethics Declaration

Participation was voluntary, data were anonymised, and no sensitive personal information was collected.

AI Declaration

AI tools (ChatGPT by OpenAI) were used to support language refinement and structure. All outputs were critically reviewed and edited by the author.

References

Black, P. and Wiliam, D. (2009) "Developing the theory of formative assessment", Educational Assessment, Evaluation and Accountability, Vol. 21, No. 1, pp 5–31.

Boud, D. and Molloy, E. (2013) "Rethinking models of feedback for learning: the challenge of design", Assessment & Evaluation in Higher Education, Vol. 38, No. 6, pp 698–712.

Carless, D. and Boud, D. (2018) "The development of student feedback literacy: enabling uptake of feedback", Assessment and Evaluation in Higher Education, Vol. 43, No. 8, pp 1315–1325.

Deci, E.L. and Ryan, R.M. (2000) "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior", Psychological Inquiry, Vol. 11, No. 4, pp 227–268.

Dunn, K.E. and Mulvenon, S.W. (2009) "A critical review of research on formative assessment: The allegory of Plato's cave", International Journal of Education Policy and Leadership, Vol. 4, No. 8, pp 1–11.

Fisher, D.P. et al (2025) "The impact of timely formative feedback on university student motivation", Assessment & Evaluation in Higher Education, Vol. 50, No. 4, pp 622–631.

Fong, C.J., Patall, E.A. and Vasquez, A.C. (2019) "A meta-analysis of negative feedback on intrinsic motivation", Educational Psychology Review, Vol. 31, No. 1, pp 121–162.

Hattie, J. (2009) Visible learning: A synthesis of over 800 meta-analyses relating to achievement, Routledge, London.

Hrabal, V. and Pavelková, I. (2011) Školní výkonová motivace žáků: Dotazník pro žáky, Národní ústav odborného vzdělávání, Prague.

Kluger, A.N. and DeNisi, A. (1996) "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory", Psychological Bulletin, Vol. 119, No. 2, pp 254–284.

Lichtenberger, A. et al (2025) "Enhanced conceptual understanding through formative assessment: Results of a randomized controlled intervention study in physics classes", Educational Assessment, Evaluation and Accountability, Vol. 37, No. 1, pp 5–33.

Muho, A. and Taraj, G. (2022) "Impact of formative assessment practices on student motivation for learning the English language", International Journal of Education and Practice, Vol. 10, No. 1, pp 25–41.

Näsström, G. et al (2021) "Changes in student motivation and teacher decision making when implementing a formative assessment practice", Frontiers in Education, Vol. 6, Article 616216.

Nicol, D.J. and Macfarlane-Dick, D. (2006) "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice", Studies in Higher Education, Vol. 31, No. 2, pp 199–218.

Palmberg, B. et al (2024) "A multi-approach formative assessment practice and its potential for enhancing student motivation: A case study", Cogent Education, Vol. 11, No. 1, Article 2371169.

Solis Trujillo, B.P. et al (2025) "The current landscape of formative assessment and feedback in graduate studies: A systematic literature review", Frontiers in Education, Vol. 10, 1509983.

Winstone, N. and Carless, D. (2020) Designing effective feedback processes in higher education: A learning-focused approach, Routledge, London.

Work in Progress Papers

Challenges and Opportunities for Applying Artificial Intelligence to Learning in Secondary Schools

Ioannis Panagiotis Lappas
South-West University "Neofit Rilski", Blagoevgrad, Bulgaria

nnslappas@gmail.com

Abstract: Artificial intelligence is gradually becoming part of everyday practice in secondary education. Teachers now encounter AI-based tools in different aspects of school work, from feedback and classroom organisation to adjusting learning activities to fit individual student needs. As the use of such technologies expands, their educational role and practical value in schools require closer examination. This becomes particularly relevant in classrooms with varied learning characteristics, where students differ in learning pace, participation, learning needs, and overall educational readiness. In such settings, AI technologies can support the adaptation of instructional material, provide more individualised feedback, and offer additional forms of participation in the learning process. Their educational value, however, depends on professional judgement and on the way these tools are used within everyday classroom practice. The present paper addresses opportunities and challenges connected with the use of artificial intelligence in secondary education. It focuses on teachers' perspectives and on their experiences with the educational use of AI tools in school practice. This doctoral work adopts a primarily qualitative approach. Data will be collected through semi-structured interviews with secondary education teachers and later analysed through thematic analysis. Through interview responses, the analysis will examine how AI is perceived in relation to lesson preparation and student support, as well as the practical limits of its implementation in schools. Special attention will be given to professional preparedness and to the everyday classroom conditions that affect the responsible use of AI. The analysis will also consider how participants understand the educational value of AI tools and the ways these technologies are applied in secondary school classrooms. The paper therefore approaches AI not as an independent solution, but as a tool whose usefulness depends on teachers' judgement and the pedagogical context in which it is used.

Keywords: Artificial Intelligence, Secondary Education, Teachers' Perspectives, Classroom Practice, Qualitative Research, Responsible AI Use

1. Introduction

Artificial intelligence is becoming more visible in secondary education through applications designed to support feedback, lesson preparation and the adaptation of learning material (Chen et al, 2020; Kasneci et al, 2023). Its presence in schools, however, does not automatically mean that students actually learn better or that AI has pedagogical value. For AI-supported teaching to be meaningful, the tools must be able to serve the purpose of the lesson, fit the classroom context and remain under the educator's professional judgement and control (Selwyn, 2024; OECD, 2026).

Teachers are therefore central to the classroom use of artificial intelligence, with research on their perceptions showing that they may recognise AI as a possible support for teaching while still reporting limited knowledge and concerns about its ethical and practical use (Chounta et al, 2022). This issue becomes particularly relevant in secondary schools, where students may use generative AI for school-related tasks. The OECD (2026) notes that generative AI can improve short-term task performance but may weaken learning when it is used mainly as a shortcut. For this reason, the educational value of AI depends not only on access to the tools, but also on how teachers guide its use and connect it with the purpose of the lesson.

This paper is part of an ongoing doctoral project on the use of artificial intelligence in secondary education. At this stage, it draws on existing literature to explain why the study examines the views of secondary education teachers and to outline the planned qualitative orientation.

2. Literature Background

Artificial Intelligence in Education (AIED) does not refer to a single type of technology. It includes earlier forms of educational AI, such as intelligent tutoring systems and adaptive learning environments, as well as newer tools based on generative AI. Previous research examined structured systems designed for tutoring or adaptation, whereas more recent work has turned toward applications that teachers and students can use directly in learning activities (Chen et al, 2020; OECD, 2026).

This shift becomes clearer in the case of generative AI, where large language models can produce explanations, reformulate texts and support feedback. These capabilities have made such tools more relevant to everyday teaching than many earlier educational AI systems. Easier access, however, does not make AI educationally

meaningful by itself. Its use still requires critical checking and pedagogical planning, especially when teachers use AI-generated material with students (Kasneci et al, 2023).

For this reason, the role of the classroom teacher needs to be examined more closely. Zawacki-Richter et al (2019), in a review of 146 studies, found that AIED research was strongly influenced by Computer Science and STEM perspectives and relied mainly on quantitative methods. Their review also showed weak links with pedagogical theory and limited critical discussion of educational risks. Although the review focused on higher education, it raises a useful question for this paper: AI in education is often examined through the design and effectiveness of systems, while less attention is given to educators' pedagogical judgement in real classroom situations.

Teachers' perspectives are therefore not a secondary issue. In a K-12 sample of 140 participants, limited knowledge of AI coexisted with a generally positive view of its possible value for teaching. This matters for the present paper because uncertainty about AI can be read as a question of preparation and classroom guidance, not only as resistance to technology (Chounta et al, 2022). From a policy perspective, European guidance makes a similar point by emphasising that educators need enough understanding to judge both the possibilities and the risks of AI and data use in teaching (European Commission, 2022).

3. Research Gap and Aim of the Study

The literature reviewed above covers both older forms of AIED, such as adaptive learning, and newer debates around generative AI and responsible use. What remains less clear is how secondary education teachers understand these tools when they think about their own classrooms.

The gap is therefore more practical than technical. Existing studies provide useful evidence about AI applications and teachers' general perceptions, but offer less detail on how secondary education teachers connect these tools with lesson preparation and student support in their own classrooms (Zawacki-Richter et al, 2019; Chounta et al, 2022). The present study addresses this gap by using interview responses to understand the conditions under which AI may become pedagogically useful.

The aim of the doctoral project is to examine the challenges and opportunities connected with the application of artificial intelligence in the learning process in secondary schools. It considers how participants understand the educational value of AI tools and what conditions they see as necessary for responsible implementation. The project examines their reflections on AI in relation to actual teaching practice.

4. Methodology

A qualitative research approach is used, as the purpose is to understand how secondary education teachers make sense of artificial intelligence in the learning process. Rather than measuring only how often AI tools are used, the analysis considers how participants describe their experience with AI and how they explain the conditions that affect classroom use (Creswell and Creswell, 2018).

Data will be collected through semi-structured interviews with secondary education teachers with different levels of familiarity with AI-related tools in school practice. This method provides a common interview structure, while allowing participants to explain their views in more detail. The interviews will address their familiarity with AI tools, as well as possible educational uses and difficulties identified in school practice.

The interview data will be analysed through thematic analysis. The process will involve careful reading of the data, coding of relevant extracts and development of themes in relation to the research questions (Braun and Clarke, 2006). Participation will be voluntary. Participants will be informed about the purpose of the study before taking part, and no individual or school will be identifiable in the presentation of the findings. The methodology is presented as the planned design of an ongoing doctoral project and may be refined as the empirical work develops.

5. Expected Contribution

Since the research is still in progress, this paper does not present empirical findings from interviews. Its contribution lies in defining a teacher-centred framing for examining artificial intelligence in secondary education through participants' interpretations of school practice. This orientation responds to a gap in the literature, where AI is often examined through system capabilities or general implementation issues, while less

is known about how such tools are judged in relation to everyday teaching conditions (Zawacki-Richter et al, 2019; Chounta et al, 2022).

The paper therefore shifts the question from what AI tools can produce to how their use is evaluated in teaching. A generated explanation or feedback response may be useful in one lesson and unsuitable in another. For this reason, teachers' perspectives can help clarify the practical conditions that affect responsible AI use in secondary schools, especially in relation to teacher preparation and classroom judgement (European Commission, 2022).

6. Conclusion

Artificial intelligence is becoming more present in secondary education, but its educational value cannot be assumed from availability alone. AI tools may support teaching when they are connected with a clear lesson purpose and when their outputs are checked by the teacher. Without this kind of pedagogical judgement, the same tools may encourage superficial task completion rather than meaningful learning (OECD, 2026).

The planned qualitative research will examine how secondary education teachers understand the opportunities and limits of AI in their work. By centring these perspectives, the paper links responsible AI use to classroom practice, where teacher preparation and human oversight remain necessary (European Commission, 2022).

Ethics Declaration

No interview data have yet been collected, and this paper presents the planned design of the ongoing doctoral project. Before data collection begins, the ethical procedures required by the doctoral programme and the university will be followed. Participation will be voluntary, and every individual will be informed about the purpose of the research before taking part. Data will be treated confidentially, and no teacher or school will be identifiable in the findings.

AI Declaration

AI-assisted support was used for language support, including translation of draft notes into English and refinement of wording. The author reviewed, edited and approved the final paper, checked the cited literature, and remains fully responsible for the content, argumentation, citations and accuracy of the submitted work.

References

Braun, V. and Clarke, V. (2006) "Using Thematic Analysis in Psychology", *Qualitative Research in Psychology*, Vol 3, No. 2, pp 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, L., Chen, P. and Lin, Z. (2020) "Artificial Intelligence in Education: A Review", *IEEE Access*, Vol 8, pp 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chounta, I.-A. et al (2022) "Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support Their Practice in Estonian K-12 Education", *International Journal of Artificial Intelligence in Education*, Vol 32, pp 725–755. https://doi.org/10.1007/s40593-021-00243-5

Creswell, J.W. and Creswell, J.D. (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th edn, SAGE Publications, Thousand Oaks, CA.

European Commission (2022) *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators*, Publications Office of the European Union, Luxembourg. https://doi.org/10.2766/153756

Kasneci, E. et al (2023) "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education", *Learning and Individual Differences*, Vol 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

OECD (2026) *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*, OECD Publishing, Paris. https://doi.org/10.1787/062a7394-en

Selwyn, N. (2024) "On the Limits of Artificial Intelligence (AI) in Education", *Nordisk Tidsskrift for Pedagogikk og Kritikk*, Vol 10, No. 1, pp 3–14. https://doi.org/10.23865/ntpk.v10.6062

Zawacki-Richter, O. et al (2019) "Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?", *International Journal of Educational Technology in Higher Education*, Vol 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Late Submissions

Beyond the Digital Journal: Epistemic Conversations in Work-Integrated Learning

Ruben Daniels
Cape Peninsula University of Technology, Mowbray Campus, South Africa

DanielsRu@cput.ac.za

Abstract: Work-Integrated Learning (WIL) in teacher education increasingly relies on digital tools, including video-recorded microteaching platforms and online reflective journals, to support and assess pre-service teachers' professional development. This paper examines whether such digital infrastructure produces deeper epistemic engagement or merely reproduces the same compliance-oriented reflection long associated with paper-based journalling. Drawing on a qualitative comparative case study of Post Graduate Certificate in Education (PGCE) programmes at two Western Cape Higher Education Institutions, the study analysed semi-structured interviews and reflective journal entries generated from twenty participants across four stakeholder groups: PGCE students, method lecturers, Work-Integrated Learning coordinators, and school-based mentor teachers. Shulman's Pedagogical Reasoning and Action Model and Bourdieu's concepts of habitus, field, and capital served as theoretical anchors, allowing the analysis to examine both the cognitive work supported by digital tools and the structural conditions determining whether that work was realised in practice. Digital platforms, chiefly private online video channels used to record and revisit microteaching sessions, did create genuine opportunities for sustained self-monitoring and, in a smaller number of cases, collaborative dialogue between students, going beyond what paper-based methods alone could offer. Yet this technological layer sat apart from the reflective journal itself, which kept functioning as an individual, retrospective, compliance-driven artefact whatever format it took. Time pressure, limited prior teaching exposure, and journal designs built around assessment rather than dialogue showed up as barriers whether students worked on paper or online. The paper argues that the missing layer in online reflective practice is not technological capacity but epistemic architecture: digital tools were not designed, and are not currently used, to convert recorded experience into the kind of dialogic, theory-connected conversation that produces professional learning. The single strongest example of such a conversation in the data arose from a lecturer's deliberate pedagogical design choice rather than from any inherent affordance of the technology itself. The paper closes with three recommendations: rebuilding digital reflective infrastructure around structured epistemic conversations, redesigning journals so they hold fewer but more substantial entries, and building teacher educators' capacity to facilitate dialogic reflection. None of this follows automatically from installing a digital repository.

Keywords: Digital Reflective Journals; Epistemic Conversations; Work-Integrated Learning; PGCE; Teacher Education; E-Portfolios; Online Reflective Practice

1. Introduction

WIL matters in teacher preparation for a fairly obvious reason: a university lecture theatre cannot teach what a real classroom does (Billett 2009; Billett 2024). South Africa runs this through the PGCE, a one-year qualification in which WIL takes the shape of supervised teaching practice, all of it governed by the Minimum Requirements for Teacher Education Qualifications (MRTEQ) policy framework (DHET 2015). Somewhere in the middle of how these programmes claim to convert placement experience into professional learning sits reflective practice, and in practice that usually means keeping a journal.

In the last decade or so, PGCE programmes both here and abroad have bolted digital tools onto this arrangement, among them video-recorded microteaching sessions hosted on private online channels, e-portfolio platforms, and reflective journals submitted through learning management systems. The assumption driving this shift is rarely stated outright, but it runs through the literature all the same: digital infrastructure should deepen reflection because experience becomes reviewable and shareable, and more open to feedback than a handwritten or typed entry submitted once and read once. This paper interrogates that assumption directly. Drawing on a completed doctoral study of PGCE students' experiences of reflective practice within WIL at two Western Cape Higher Education Institutions (HEIs), it asks whether digital reflective tools alter the fundamental character of reflection, or whether they merely digitise the same individual, compliance-oriented exercise that has long characterised the paper journal.

The argument advanced here is that the digital layer introduced genuine value, principally through recorded microteaching that enabled students to revisit and self-monitor their teaching over time, but that this value stopped short of producing what this paper terms epistemic conversations: dialogic exchanges in which the knowledge claims underlying a pedagogical decision are made available for joint examination, critique, and revision. The reflective journal, whether rendered on paper or through a learning management system, remained epistemically the same artefact. What was missing was not bandwidth but architecture.

2. Literature Review

2.1 Digital Tools in Work-Integrated Learning

Digital reflective journals and e-portfolio platforms have been widely adopted across teacher education on the assumption that recording and reviewing experience will deepen professional learning (Smart, Sim and Finger 2015). Starkey (2010) extends Shulman's model of pedagogical reasoning into the digital age, arguing that technology reshapes how teachers comprehend, transform, and evaluate their practice, but cautions that the reshaping is not automatic: it depends on how the technology is used pedagogically, not merely on its presence in the curriculum.

This caution is borne out in the broader literature on reflective journalling. Machingambi and Wadesango (2022) treat journals, digital or otherwise, as central to PGCE curricula the world over, and Mohamed, Valcke and De Wever (2022) back this up with a systematic review spanning several continents. Their prevalence outstrips the evidence for their effectiveness, though. The problem is built into the assessment requirements themselves. A journal is meant to support open, exploratory reflection. It is also graded. Those two things pull in opposite directions, and students respond by writing safe and descriptive entries rather than genuinely critical ones (Boud 2010, cited in Suphasri and Chinokul 2021). Moving the journal online changes its medium, not this underlying tension. A compliance document submitted through a portal is still a compliance document.

2.2 Video and Recording Technologies as a Partial Exception

A distinct strand of practice, less examined in the reflective journalling literature, concerns recorded microteaching. Where journals ask students to write about experience after the fact, recorded microteaching allows students to review the experience itself. Farrell's (2015) five-stage model of reflective development suggests recording and reviewing supports several core components of structured reflection, including the capacity to revisit a specific pedagogical decision rather than relying on memory. Nguyen (2015) makes much the same point, treating revisitability as a condition that strengthens the link between thought and action during reflection.

Whether that revisitability actually leads anywhere near dialogic, theory-connected reflection is a different question, and not an obvious one. It might simply carry the same individual, self-directed way of reflecting into a new format. Quinn and Crotty (2024) would argue the former is unlikely on its own: in their account, it is critical collaborative conversation, not individual review of any kind, that most reliably supports teacher reflection and professional development. This raises the central question this paper addresses: does recorded, digitally mediated experience create the conditions for collaborative, theory-connected reflection, or does it remain locked within the same individually authored, individually assessed architecture as the paper journal it was meant to improve upon?

3. Theoretical Framework

Two theoretical anchors underpin this study. Shulman's (1987) Pedagogical Reasoning and Action (PR&A) Model helps explain how teachers move through cycles of comprehension, transformation, instruction, evaluation, and reflection, and, more specifically, how digital tools enter into the evaluation and reflection phases of that cycle. Bourdieu's (1977; 1986) concepts of habitus, field, and capital do different work: they treat reflection not as a purely cognitive act but as a practice shaped by structural conditions, among them the assessment regime governing the journal, the technological capital available at each institution, and the dispositions students bring with them into the WIL placement.

Taken together, these two frameworks let the analysis hold two questions at once: what cognitive work a given reflective tool actually supports, digital or otherwise (Shulman), and what structural and dispositional conditions decide whether that cognitive work gets realised in practice (Bourdieu). This matters for evaluating digital tools specifically, because the promise of technology tends to get framed in purely cognitive terms, better recall, richer review, while the structural conditions that govern its use, chiefly assessment design, go unexamined.

4. Methodology

The empirical study behind this paper used a qualitative, dual-site case study design, working within an interpretivist paradigm to explore PGCE students' experiences of reflective practice during WIL at two Western Cape HEIs. To preserve institutional anonymity these are designated Site 1 and Site 2 throughout; both operate under the same MRTEQ policy framework and offer one-year PGCE qualifications. Semi-structured interviews and analysis of document sources, principally students' reflective journals, generated the data, which then went through thematic analysis following Braun and Clarke's (2006) six-phase framework.

Twenty participants took part in all, drawn from both sites and spanning four stakeholder groups: PGCE students, method lecturers and university-based supervisors, WIL coordinators, and school-based mentor teachers. This paper identifies participants using the coding convention established in the original study, so S1-s1 refers to a Site 1 student and S1-ML2 to a Site 1 method lecturer, for example. Before any data collection began, the study met the ethical requirements of the University of the Western Cape's Humanities and Social Science Research Ethics Committee, and the Faculty of Education at both participating institutions granted permission.

5. Findings

5.1 Recorded Microteaching as a Genuine Site of Self-Monitoring

Method lecturers at Site 1 described using private online video channels to record microteaching sessions, allowing students to revisit their own teaching over an extended period:

> *"And what I also do I record the micro teaching sessions, the before and the after... And it's on YouTube, not public. It's a private YouTube channel. So, they can go back every single time in five years' time to go... it's a way of them being able to constantly see and monitor whether they've improved." (S1-ML2)*

Lecturers were explicit that this was a deliberate pedagogical choice, not an administrative convenience, and tied it to a broader disposition towards preparing students for a profession now saturated with digital tools:

> *"Well, I use quite a lot of digital learning in my lessons so I would focus on the using of video through... reflection through videos... because the world as we go into the fourth industrial revolution I think it's important that our students are kept abreast of the developments." (S1-ML2)*

A related but distinct use of recorded video supported group-based discussion rather than individual review. One lecturer described using recorded classroom footage as a stimulus for structured group dialogue that moved towards what he termed cognitive harmonisation, a process of collective sense-making:

> *"I would give a video of a classroom setting. And then at the end of the day, we would now have groups that give a report or reflect on the practice of what was seen in the video... come to cognitive harmonisation. Cognitive harmonisation is where you either agree to disagree, or you disagree to agree." (S1-ML3)*

This last example is the closest the data comes to what this paper defines as an epistemic conversation: a structured, group-based examination of a recorded pedagogical decision in which multiple knowledge claims are placed alongside one another and worked through collectively, rather than reviewed and written up alone.

5.2 The Journal Itself Remained Unchanged by its Medium

Despite the presence of digital recording infrastructure at Site 1, the reflective journal, whether submitted digitally or on paper, continued to function as an individually authored, retrospective, assessment-driven artefact. The WIL coordinator at Site 1 observed that students entered the programme associating reflective practice narrowly with journal writing itself, regardless of format:

> *"And I think they come in with just a general thing about reflective practice, and that is journaling. When they hear the word reflective practice, they think, oh, I have to write a journal, a journal... Yes, we do, it is part of, but there are so many other things we as teachers need." (S1-WIL)*

Students' own journal entries, analysed as document sources, were overwhelmingly descriptive rather than analytical, recording what happened during a lesson without connecting it to the theoretical frameworks introduced in coursework. This pattern held regardless of whether the journal was maintained on paper or

submitted through a digital learning platform, indicating that the format of the journal was not the operative variable. What determined the quality of reflection was whether the entry was written for an audience capable of dialogic response, and in practice, journal entries were read once, by one assessor, after the reflective moment had passed.

5.3 Structural and Dispositional Constraints on Digital Reflection

Consistent with the Bourdieusian dimension of the analysis, the extent to which digital tools supported richer reflection was shaped by institutional resourcing and time. Method lecturers across both sites identified time constraints as the single largest barrier to developing reflective capacity, digital or otherwise:

> *"The biggest obstacle for the PGCE is time constraints, you know. There is not enough time... I don't think that we've got enough time for them to be able to be skilled on how to best reflect, especially on their teaching." (S1-ML1)*

Students themselves had limited prior exposure to any structured reflective or teaching practice before entering the PGCE, meaning the cognitive load of learning to teach competed directly with the cognitive load of learning to reflect on that teaching through any medium:

> *"They've never been in a platform as a teacher. Never... only one so far had teaching experience as an informal teacher." (S1-ML2)*

It made little difference, in the end, whether the reflective tool in question was digital or paper-based. These constraints showed up either way, which points to a barrier that is structural and dispositional rather than technological.

6. Discussion: The Missing Layer

A specific, previously underexamined gap emerges from these findings. Where digital infrastructure was actually used to record and revisit practice, it genuinely extended what students and lecturers could do. Sessions could be reviewed months or even years later, and in at least one documented case, recorded footage became the basis for structured group dialogue. Digital tools, in other words, are not inert. They can support genuine pedagogical adaptation, which fits Starkey's (2010) argument that technology reshapes pedagogical reasoning when it is used deliberately.

That technological capacity, though, was never connected to the reflective journal itself, the instrument formally designated as the vehicle for reflective practice within the programme. Paper document or digital submission, it made no difference: the journal stayed an individually authored, retrospectively written, once-read artefact either way. This disconnection, not any deficiency in digital capability, is this paper's central claim about the missing layer in online reflective practice. Institutions took up digital tools for recording and reviewing experience on one hand, and kept maintaining digital or paper journals for documenting and assessing reflection on the other, without ever connecting the two into a single dialogic process.

Why this matters becomes clear in the one instance in the data that came closest to deep, theory-connected reflection: the group-based "cognitive harmonisation" activity. What distinguished it was not its use of technology but its dialogic structure. Multiple students' interpretations of the same recorded event were positioned against one another, and the group had to work towards a shared understanding, or at least explicitly acknowledge where they disagreed. This paper calls that an epistemic conversation, a conversational event in which the knowledge claims underlying a pedagogical decision become jointly available for examination, critique, and development, rather than remaining the private property of whoever wrote the journal entry.

Practically, this means institutions investing in digital reflective infrastructure should stop treating recording capability and the journalling requirement as separate, unconnected obligations. Recorded microteaching footage could instead become the shared object around which structured, facilitated group conversation happens, with the journal recast as a record of that conversation rather than standing in for it. None of this calls for technology beyond what Site 1 already had in place; what it calls for is redesigning how that existing technology gets used pedagogically.

7. Conclusion and Recommendations

This paper has argued that adopting digital tools within WIL reflective practice can extend genuine pedagogical value, chiefly through recorded, revisitable microteaching, but that this has not on its own resolved the long-standing limitations of the reflective journal as an individually authored, compliance-oriented artefact. What is missing sits at the level of epistemic design, not technology: digital infrastructure has rarely been built, or used, to turn recorded experience into structured, dialogic, theory-connected conversation.

Three recommendations follow. First, institutions should reposition recorded microteaching footage as a shared object for structured group reflection rather than a purely individual review resource, building on practice already observed at Site 1. Second, digital reflective journals should be redesigned around fewer, more substantial entries structured by critical incidents and explicit prompts connecting experience to theoretical frameworks, rather than daily or weekly logging. Third, teacher educators require professional development in facilitating epistemic conversations, since the technology alone does not generate dialogic engagement; the deliberate pedagogical structuring observed in the single strongest example in this data was the product of a lecturer's design choice, not the platform's affordances.

Future research should examine whether deliberately structuring digital platforms around collaborative, theory-connected dialogue, rather than individual submission, measurably improves the quality of student reflection, and whether the concept of epistemic conversations offers a transferable design principle for WIL programmes beyond teacher education.

Ethics Declaration

The data reported here were collected as part of a completed doctoral study. Ethical clearance was granted by the University of the Western Cape's Humanities and Social Science Research Ethics Committee prior to data collection (HSSREC Reference Number: HS22/9/34, approval period 27 February 2023 to 26 February 2024). The Faculty of Education at both participating Higher Education Institutions gave permission for the research to proceed, as did the Western Cape Education Department. All participants gave informed consent, and institutional and participant identifying information has been anonymised throughout in line with that clearance.

AI Declaration

Generative AI tools assisted in preparing this manuscript, chiefly with drafting and language refinement, working throughout from the author's own doctoral research data and analysis. The theoretical argumentation, interpretation of findings, and conclusions belong to the author alone. Before submission, the author reviewed and edited the manuscript in full, and it will also be checked using institutional plagiarism and AI-detection tools.

References

Billett, S. (2009) "Realising the educational worth of integrating work experiences in higher education", Studies in Higher Education, Vol 34, No. 7, pp 827-843.

Billett, S. (2024) "Constituting integration in work-integrated education and learning", Studies in Continuing Education, Vol 46, No. 3, pp 513-530.

Boud, D. (2010) "Relocating reflection in the context of practice", in Bradbury, H., Frost, N. and Kilminster, S. (eds.) Beyond Reflective Practice, Routledge, London, pp 25-36.

Bourdieu, P. (1977) Outline of a Theory of Practice, Cambridge University Press, Cambridge.

Bourdieu, P. (1986) "The forms of capital", in Richardson, J. (ed.) Handbook of Theory and Research for the Sociology of Education, Greenwood, New York, pp 241-258.

Braun, V. and Clarke, V. (2006) "Using thematic analysis in psychology", Qualitative Research in Psychology, Vol 3, No. 2, pp 77-101.

Department of Higher Education and Training (DHET) (2015) Minimum Requirements for Teacher Education Qualifications, Government Printers, Pretoria.

Farrell, T.S.C. (2015) Promoting Teacher Reflection in Second Language Education: A Framework for TESOL Professionals, Routledge, New York.

Machingambi, S. and Wadesango, N. (2022) "Reflective journals as tools for developing reflective practice in teacher education", South African Journal of Education, Vol 42, No. 1, pp 1-8.

Mohamed, Z., Valcke, M. and De Wever, B. (2022) "Reflective journaling in teacher education: A systematic review", Teaching and Teacher Education, Vol 117, 103794.

Nguyen, Q.D. (2015) What is Reflection? A Conceptual Analysis of Major Definitions and a Proposal of a Five-Component Definition and Model, Unpublished Master's thesis, University of Montreal.

Quinn, F. and Crotty, Y. (2024) "Critical collaborative conversations: Supporting teacher reflection and professional development", Professional Development in Education, Vol 50, No. 1, pp 12-25.
Shulman, L.S. (1987) "Knowledge and teaching: Foundations of the new reform", Harvard Educational Review, Vol 57, No. 1, pp 1-23.
Smart, V., Sim, C. and Finger, G. (2015) Teaching with Digital Technologies: A Guide for Tertiary Teachers, Griffith University, Brisbane.
Starkey, L. (2010) "Teachers' pedagogical reasoning and action in the digital age", Teachers and Teaching: Theory and Practice, Vol 16, No. 2, pp 233-244.
Suphasri, P. and Chinokul, S. (2021) "Reflective teaching and its implementation: A review of literature", LEARN Journal, Vol 14, No. 1, pp 240-261.

PGCE Students' Perceptions of Generative Artificial Intelligence within Learning Ecosystems at a South African Teacher Education Institution

Tunjera Nyarai and Chigona Agnes
Cape Peninsula University of Technology

nztunjera@gmail.com
Chigonaa@cput.ac.za

Abstract: The Fourth Industrial Revolution has brought artificial intelligence (AI) to the forefront of pedagogy in higher education institutions. However, limited empirical literature outlines students ' perceptions of AI-powered learning in resource-poor environments of the Global South. Methods This study adopted an interpretive, mixed-methods, single-case design to investigate undergraduate students ' perceptions of AI in their learning ecosystems. A purposive sampling approach was used to purposefully select 88 students taking the ICTs in Education class at a South African higher education institution. All the data collected in this study were based on one comprehensive research questionnaire that included both fixed-response Likert items and open-ended questions. The survey instrument elicited students ' general perceptions about AI and their experience in a course activity where the task was to prompt AI to develop lesson plans, produce media products (posters, games, or videos) and create digital assessment activities. Results show that AI emerged as a powerful digital mediational artefact within the integrated conceptual framework that combines the principles of Vygotskian sociocultural theory and the ConTIS Model. 75% of respondents said that AI-based tasks made them feel more personalised and confident in the process of learning in their Zone of Proximal Development. However, there are several obstacles to AI-driven transformation. Socio-material limitations in the region, particularly data poverty and load shedding, constantly hinder the continuity of AI interactions, in addition to teachers' lack of competence in using AI tools (68%) and data privacy issues (61%).

Keywords: Artificial Intelligence, Fourth Industrial Revolution, Teacher Education, Personalised Learning, Student Perceptions, Digital Equity

1. Introduction

The Fourth Industrial Revolution (4IR) involves the fusion of digital, physical, and biological aspects and is creating revolutionary changes in society, the economy, and education across the globe (Schwab, 2016). As far as teacher education is concerned, AI has evolved from being merely an aspect of computing technology into a revolution in the way that institutions think, design, and teach (Zawacki-Richter et al., 2019). The current AI includes intelligent tutoring systems, adaptive learning, natural language processing, and automatic feedback (Holmes et al., 2019). Being central to the process of learning, AI holds importance for teaching, learning and institutions.

1.1 Rationale: Epistemic Gaps and Localized Fractures

Despite the incorporation of AI in universities at an unprecedented pace, gaps in the literature can be observed. The recent literature on educational technology after 2022 addresses AI generators or uses in institutions but does not mention students' experiences (Selwyn, 2019). Students are mostly passive consumers of the learning enabled by AI, leading to the marginalization of the epistemology of Global South higher education institutions (HEIs), which are characterized by diverse socio-economic conditions (Pedro et al., 2019). For optimizing Generation AI, the socio-economic conditions of these countries should be considered.

1.2 Problem Statement and Study Questions

The study addresses the empirical and contextual gaps through an exploration of Postgraduate Certificate in Education (PGCE) students ' perceptions of AI-mediated learning in a South African teacher education institution. Through an emphasis on the subjective experience of students, the research aims to unravel PGCE students ' perceptions of the role of Generative Artificial Intelligence (GenAI) in their learning ecosystems and to investigate the factors influencing their experiences of AI-mediated learning.

With a view to formulating a clear direction for the inquiry and answering the specific research problem, the study is informed by the following research questions: What perceptions do PGCE students hold regarding AI's role within their learning ecosystems? What are the complex factors that influence these experiences? Answering the research questions will provide much-needed empirical information for the development of curricula and policy frameworks for sustainability.

2. Literature Review

2.1 The Post-ChatGPT Landscape of AI in Teacher Education

The rapid diffusion of GenAI in the public domain has prompted a worldwide scholarly reaction. Kasneci et al. (2023) posit that GenAI represents a paradigmatic change from rule-based educational technology, allowing dynamic and situational interaction and transforming teacher-learner dynamics. From the perspectives of cognitive and computer science, Kasneci et al. (2023) provide a taxonomy of the capabilities and limitations of Large Language Models (LLMs), with an emphasis on dynamic content creation, automatic formative feedback, and tutoring intelligence. Crompton and Burke (2023) conduct a Preferred Reporting Items for Systematic Reviews and Meta-Analyses **(**PRISMA)-compliant review of 138 articles (2016-2022) and find that AI in teacher education has been growing exponentially before ChatGPT, with volumes doubling or tripling in 2021-2022. The baseline has been established for GenAI-era growth. Traditional AI operates through rules or patterns. In contrast, GenAI synthesizes original text or media.

Kasneci et al. (2023) agree that current LLMs can generate real-time, context-sensitive responses that can mimic dialogic tutoring. Crompton and Burke (2023) reach the same conclusion and add that AI has moved from predictive analytics to pedagogy. However, there exists uneven optimism. Kasneci et al. (2023) appeal for moderation and caution with risks such as misinformation or "hallucinations", overreliance, cognitive atrophy, and pedagogical supervision needed. Again, Crompton and Burke (2023) highlight the scope of the research (pre-2023). Addressing the research gaps, the current paper collects qualitative data from 88 PGCE students in South Africa.

2.2 GenAI as an Enabler of Personalised Learning and Engagement

The use of GenAI in teaching involves personalization. Kasneci et al. (2023) posit that AI can give real-time, high-quality, personalized feedback that is adapted to the learner's ability to scaffold and use appropriate language. In all cases, timely and non-judgmental feedback results in learner autonomy and empowerment. Since feedback by AI is objective and iterative, it is perceived as stress-relieving and allows for trial and error (Miao & Holmes, 2023). However, opinions vary. While Miao & Holmes (2023) believe that personalization requires national infrastructures and teacher training, Kasneci et al. (2023) think that personalization is a matter of the individual, whose success will depend on previous knowledge and may create gaps.

Many studies assume a universal digital infrastructure in terms of the availability of the Internet and electricity at all times, but this is not true for the Global South. Miao & Holmes (2023) ignore the socio-material context, such as the cost of data and load shedding, making continuous dialogic scaffolding impossible.

2.3 Ethical, Institutional, and Equity Barriers to AI Integration

Implementation challenges in responsible GenAI in teacher education encompass several major integrity and policy issues. Cotton et al. (2023) name ChatGPT writing as one of the primary and hidden integrity concerns. Farrokhnia et al. (2023) develop a macro-Strengths–Weaknesses–Opportunities–Threats **(**SWOT) analysis that highlights issues with dishonesty, poor critical thinking, algorithmic bias, privacy risks, and decreased human autonomy. Chan (2023) develops an AI Ecological Education Policy grounded in 457 students and 180 institutional representatives from Hong Kong. Yan et al. (2024) confirm these obstacles through a systematic scoping review of 118 articles.

Inadequate policies are a recurring theme in the articles discussed. Farrokhnia et al. (2023) associate the integrity problem with the current status of assessment and call for policy changes and teachers' training, in addition to restrictions on GenAI. Holmes (2023) emphasises data ownership, transparency of algorithms, and human rights-orientated governance of learning.

Opinions about GenAI disruption differ. Cotton et al. (2023) perceive the issue as a governance problem; Chan (2023) perceives it as the future of education and supports AI-based assessment to promote higher-order thinking skills. Yan et al. (2024) recognise system barriers, such as model opacity and bias, as an ethical issue.

Context and geographical limitations exist. The above sources discuss highly resourced areas (the United Kingdom, Hong Kong, and Australia) and do not address such aspects as multilingualism, digital illiteracy, or socio-economic gaps typical for South Africa. Although Farrokhnia et al. (2023) provide a macro-SWOT approach, application to marginalised students is not provided.

3. Theoretical Framework

The research employs Vygotsky's theory of sociocultural constructivism and Constructivist Technology Integration Strategies (ConTIS) (Tunjera, 2019). AI is regarded as a situational and relational artefact in teacher education. AI is the mediational artefact which provides scaffolding in the process of cognitive and social activities in the Zone of Proximal Development (ZPD). For avoiding northern perspective, the research fits in with ConTIS and considers the digital transformation of TPACK. Constraints associated with data poverty may limit AI-based pedagogy and make it hard to move forward through the ZPD.

Figure 1 shows an integrated framework of this research placed at the intersection of Vygotskian Sociocultural Constructivism and ConTIS, and allowing socio-material analysis of AI-based learning in teacher education in relation to environmental factors impacting on learning effectiveness.

The socio-material core of the framework views AI as an artefact with its pedagogical value emerging from relations between learners, pedagogy, institutions and technologies. Furthermore, the environmental factors (material, socio-cultural and epistemic) that impact AI-based scaffolding and access are foregrounded. The material factors include devices, connectivity, data, and power; socio-cultural factors: language, norms, institutional support, and digital literacy; and epistemic factors: disciplinary background, curricula, and assessment. Digitally inequitable environments can hamper progress along the ConTIS trajectory towards Modification and Redefinition due to data poverty and connectivity.

Therefore, this socio-material framework is the basis of two-level analysis: (1) AI as scaffolding for learning and skill development and (2) structural and environmental factors facilitating or preventing equal access.

4. Research Design

This study employed an interpretative mixed-methods design with a single-case study framework to investigate how PGCE students make meaning of AI-mediated learning in a context-dependent manner (Creswell & Creswell, 2018). The scope of the study was limited to one group of PGCE students at a single higher education institution in South Africa, making the single-case study approach very appropriate for researching a current phenomenon in context (Yin, 2018). This study focused on collecting data using an open-ended and closed-ended questionnaire. This methodology aligns with interpretive paradigm guidelines, which prioritize understanding the subjective, lived experiences of participants within their natural settings (Creswell and Creswell, 2018; Merriam & Tisdell, 2016).

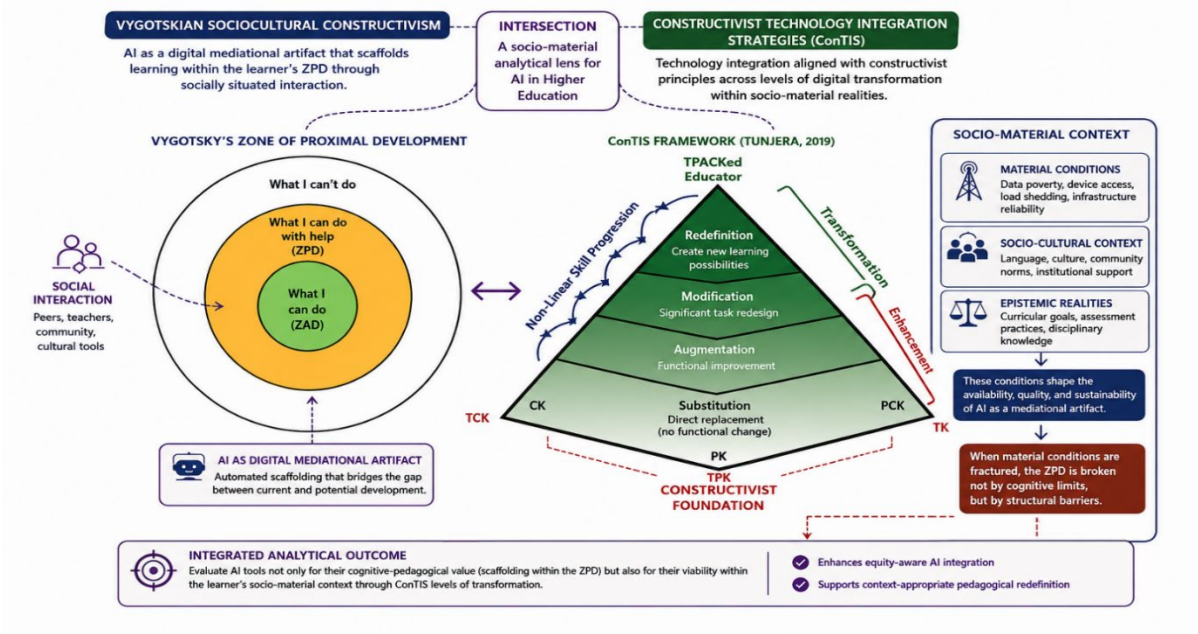


Figure 1: A socio-Material Lens for Evaluating Gen-AI in Higher Education, integrating Vygotskian ZPD and ConTIS

4.1 Participants

The participants comprised undergraduate PGCE students enrolled in a course on ICTs in education at a higher education institute in the Western Cape Province of South Africa. In this regard, a non-probability sampling design was used, which involved the use of both purposive and convenience sampling methods. The sampling design is purposive, as all respondents included in the study were purposely chosen because they were enrolled in the PGCE course and registered for the ICTs in Education module, which was essential for ensuring that participants had adequate baseline digital exposure and a theoretical framework to analyse AI as a mediational device. At the same time, the sampling design can be said to be a convenience one, as it made use of the available pool of individuals in the researcher's own institution.

4.2 Data Collection Instrument

A questionnaire was used to collect the data that would help achieve the objectives set by the study within the socio-material constructivist framework. In doing so, the integration of Likert Scales with broad open-ended questions helped in capturing both statistical and qualitative information from the 88 participants (Creswell & Creswell, 2018). Focus group interviews were conducted with fifteen PGCE students, who volunteered to be interviewed. The Focus Group Interview (FGI) complemented the questionnaire by making follow-up questions on the participants' perceptions of the AI-lesson development exercise.

4.3 Data Analysis

Both quantitative and qualitative methods were employed to analyse the data. Quantitative data obtained from 88 responses to the Likert scale survey questions were analysed using descriptive statistics in terms of frequencies and percentages. The qualitative data collected through open-ended questions and FGI were analysed thematically based on six steps outlined by Braun and Clarke (2006): familiarisation, coding, thematic development, reviewing, defining and reporting. Triangulation was applied sequentially in combining the datasets. Quantitative analysis revealed the overall pattern of perception about GenAI. At the same time, the FGIs were particularly useful in providing more explanations and experiences, as well as the socio-material aspects influencing AI-enabled learning. Triangulation involved the process of interpretation, whereby comparison was done between themes in qualitative data and trends in quantitative data in terms of convergence and divergence.

4.4 Ethical Clearance

Ethical approval for the study was obtained from the research ethics committee of the institution. The researcher shared the research information verbally and in writing. The participants were informed that participating was voluntary, anonymous, and written informed consent was obtained, which was incorporated in the instruments.

5. Findings

5.1 Demographic Composition of the Participant

The study was conducted with 88 PGCE participants enrolled in the ICTs in Education module. Table 1 shows the participants' characteristics: 65 females (73.9%) and 23 males (26.1%). Most of the participants were living at home (70.5%), 19.3% in university residences, and 10.2% in privately rented homes close to the institution. The relevant socio-material features that influence AI adoption have been identified. All the participants had smartphones; 86.4% owned personal laptops, and 13.6% used borrowed computers. Out-of-institution Internet connections depended mostly on mobile internet (77.3%), compared to institution Wi-Fi (13.6%) and personal home connection (9.1%). There was a high degree of language diversity, with the highest percentage of isiXhosa (65.9%) speakers, followed by English (9.1%), Afrikaans (8.8%) (refer Table 1),

Table 1: Demographic profile of participants (N=88)

Variable	Category	n	%
Gender	Female	65	73.9
	Male	23	26.1

Variable	Category	n	%
Living arrangement	Commute from home	62	70.5
	Private accommodation	9	10.2
	University residence	17	19.3
Primary device ownership	Smartphone	88	100.0
	Personal laptop	76	86.4
	Shared computer	12	13.6
Primary off-campus internet access data/Wi-Fi	Mobile	68	77.3
	Campus	12	13.6
	Home	8	9.1
Home language	English	8	9.1
	isiXhosa	58	65.9
	Afrikaans	10	8.8
	Other languages	12	13.6

5.2 AI an Academic Scaffolding Tool

Participants consistently described AI tools as valuable learning supports that assisted them in completing academic tasks, understanding complex concepts, and improving the quality of their work. Many students reported using AI to clarify difficult topics, generate examples, summarise academic readings, and receive immediate feedback on assignments.

On the reflection of GenAI in the lesson preparation activity, 90.1% of the participants indicated that more practice on prompting the AI enhanced their ability to create a more detailed and structurally sound lesson plan than they could have generated manually in the same timeframe. Figure 2.

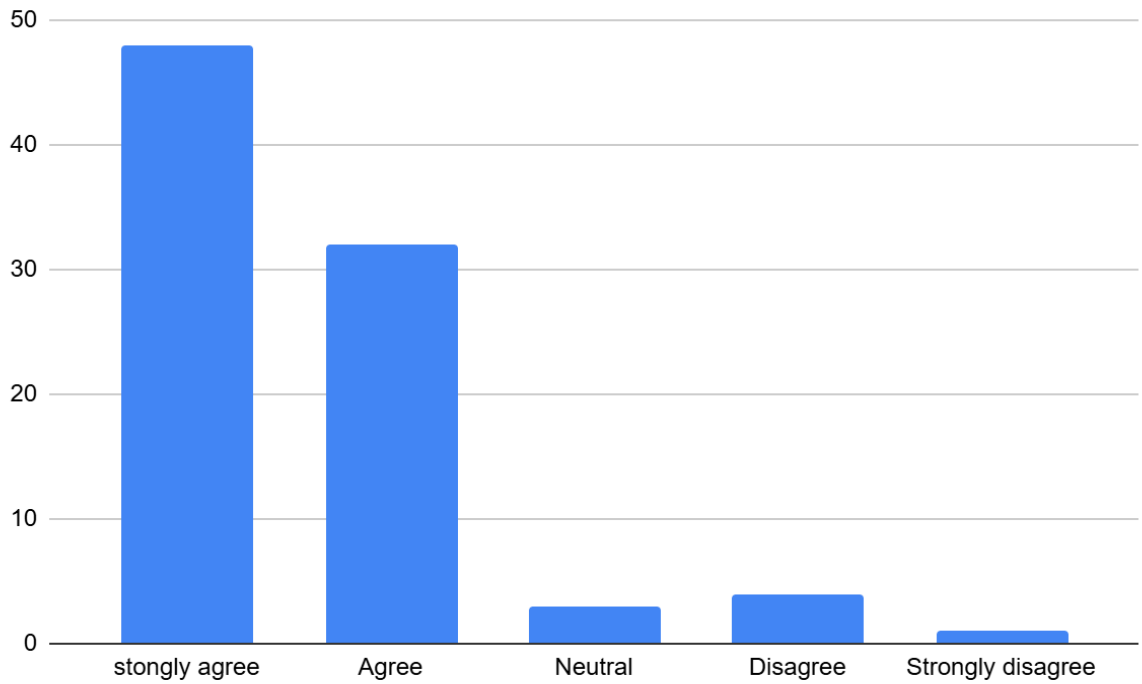


Figure 2: Prompting the AI helped refine their lesson plans.

Similarly, one participant from the FGI mentioned that using the AI prompted questions that helped her improve, and revision by adding areas that needed improvement. Several participants agreed that AI enabled them to work through learning and improving independently before seeking assistance from lecturers or peers. One participant reflected:

"When I get stuck, I ask the AI to explain the concept more simply. It helps me understand things before I go back to my notes."

The students reported that the usage of AI technology enhanced their confidence regarding unfamiliar concepts, appreciating the promptness of responses when no assistance was instantly available. Questionnaires revealed consensus on the notion that AI technology improved knowledge and facilitated self-learning.

5.3 AI-Enhanced Learning Efficiency and Productivity

A second prominent theme concerned the perceived efficiency gains associated with AI use. Participants reported that AI reduced the time required to search for information, generate ideas, structure assignments, and revise academic work. The participants' responses to the AI enhancement of learning efficiency and production were positive, with 89% agreeing to reveal a positive sentiment, with 11% having negative sentiments. Figure 3.

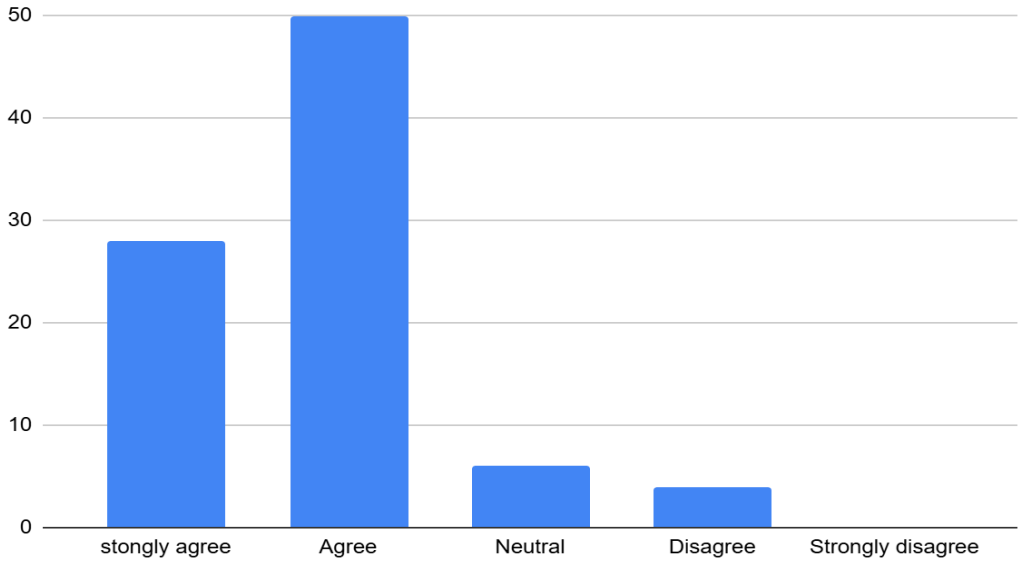


Figure 3: AI enhancement of learning efficiency and production

Many students described AI as helping them organise their thinking and initiate writing tasks that they would otherwise find difficult to begin. Reflective journal entries revealed that AI-assisted brainstorming was particularly beneficial when preparing essays, lesson plans, and research-related activities. One participant stated:

"Sometimes the hardest part is getting started. AI gives me ideas and a structure that I can build on."

The students indicated that the benefits of using AI included assistance in revision through summaries, quizzes, and simple explanations, thereby leaving more time to engage with the content. Even though the efficiency of using AI was recognized, most used their own judgment in evaluating AI products.

5.4 Tensions Between Support and Academic Integrity

Although the purported benefits of artificial intelligence (AI) are many, students voiced apprehensions over the reliance on content generated through AI and its ramifications for learning. The students pointed out that AI can, in some cases, lead to passive learning when not used critically. One remarkable response to the related question was,

“I am distressed about the collection, use, and storage of my personal and academic information by AI learning platforms.”

Every participant strongly agreed that they were uncertain about the lack of training and policies around data protection. Uncertainty was expressed by some students about what institutions expect with respect to AI in terms of the level of help permitted and the distinction between academic help and academic cheating. One such participant expressed this concern:

"Sometimes I am not sure how much AI is collecting from us, using it is too good for free. We use it because it helps, but there are no clear rules … we agree anyway."

While conducting the FGI, the students were concerned about potential problems caused by artificial intelligence, such as false citations, false facts, and oversimplification of scientific terms. Most of the students focused on the importance of verifying the information obtained from GenAI.

5.5 Socio-Material Barriers to AI-Supported Learning

The major constraints raised by the participants are those of accessibility and infrastructure. There were many issues that the students faced that prevented them from using AI technologies regularly. Among these constraints, the cost of data was raised quite often, since the use of AI applications demanded a good internet connection and, therefore, became difficult for poor students. One of the participants said that:

"AI is useful, but if you don't have data, you cannot use it when you need it."

Other barriers included those pertaining to poor internet connectivity, constraints due to hardware, and electricity failures. Such barriers would disrupt their ability to learn and interact with the AI systems. The students who came from deprived backgrounds faced more obstacles in accessing the AI learning experiences consistently. In comparison, those students who had internet access and devices felt that they were using AI systems frequently. Through the survey, participants highlighted their worries regarding access barriers to AI resources for different students. They believed that differences in infrastructure could result in the creation of inequalities in AI-mediated learning experiences and grades.

5.6 Conditions Supporting Sustainable AI Adoption

The following are some of the factors that were highlighted by the respondents as being very vital in increasing the effectiveness and sustainability of AI in teacher education. The factors included institutional guidelines concerning AI, training, low-cost Internet, better technology infrastructure, and policies of assessment.

The students expressed their desire to receive institutionalized support from the university to be able to develop skills concerning AI literacy. Many respondents advocated for the incorporation of AI in teaching and learning instead of considering AI just as another technological tool. According to one respondent:

"…Universities should teach us how to use AI properly because it is already part of how people learn and work."

Equitable access to digital sources was considered equally important in order for AI technologies to benefit all learners regardless of their socioeconomic status. According to the results, students find artificial intelligence useful as an aid in understanding and improving the efficiency of their studies, making learning more appealing. Nonetheless, the benefits associated with the use of AI technologies are subject to issues such as academic dishonesty and, even more importantly, socio-material access to digital technologies.

6. Discussion

In this research, an exploration of PGCE students' perceptions and experiences regarding the use of AI-enhanced learning in the South African higher education system is undertaken. It was found that most students perceive AI as an effective means to support the learning process and increase learning efficiency and independence, while at the same time posing academic integrity risks and socio-material barriers. Through interpretation using the combined Vygotskian-ConTIS model, the complexity of the connection between the learning potential enabled through the use of AI and conditions of its actualization can be seen.

6.1 AI as a Digital Mediational Artifact within the Zone of Proximal Development

Results indicate that learners perceive GenAI as a support system for comprehension, problem solving, and self-regulated learning. The students use AI to understand concepts, receive feedback, and get guided learning without any help from humans (Kasneci et al., 2023; Luckin, 2024). This notion corresponds to Vygotsky's theory of socially mediated learning and tools.

Students regard AI as a digital mediational tool within the ZPD of the learner that can assist with activities and concepts that lie beyond their current capability. Thus, AI assists rather than replaces human learning by offering timely explanations, feedback, and instructions to overcome difficult concepts and challenges (Holmes et al.,

2022; Luckin, 2024). The PGCE students do not see AI as a replacement for learning but rather as an additional tool of knowledge acquisition.

6.2 AI Adoption Across the ConTIS Transformation Continuum

Findings show unequal AI adoption through different levels of ConTIS. Learners initially adopted AI in accomplishing simple tasks such as explanations, summaries, grammatical corrections, and outlining, implying substitution and augmentation (technologies enhancing yet not fully transforming learning). However, more transformative usages came up: using AI for idea generation, perspectives, lesson creation, and iterative learning. Application of redefinition was less common yet noticeable: some learners utilized AI for tailoring engagement, crafting personalized content, and feedback loops beyond traditional means. Most participants continued to be in the level of Substitution-Modification, not Redefinition, in keeping with ConTIS theory since technologies alone do not bring change; pedagogy matters. As pointed out by Tunjera (2019), teachers should allow learners to achieve transformative TPACK during class activities.

6.3 Socio-Material Constraints and the Fragility of AI-Mediated Learning

The socio-material constraints impede students' engagement with AI. The participants encountered barriers such as data cost, connectivity issues, device limitations, and load shedding. AI should be understood in terms of material context and not pedagogical or technological alone (Czerniewicz et al., 2020; Selwyn, 2022).

The combination of Vygotsky's theory with ConTIS emphasizes the constraints of AI. Although AI can function as a mediator in the ZPD, its mediation is based on access. Disruptions caused by connectivity challenges, high data cost, or load shedding lessen the scaffolding.

The findings add to the sociocultural theory by emphasizing that mediation of the learning process via AI is dependent on the materiality of the environment. Access disruptions affect the ZPD of learners, not because of any limitations of their cognition, but material limitations to access AI continuously. It is consistent with the socio-material approach to AI (Selwyn, 2022).

6.4 Digital Equity and AI Adoption in the Global South

These findings impact our comprehension of digital equity in Global South universities. The respondents appreciated the power of GenAI but mentioned that the benefits are unevenly distributed. Those students who had good internet connections, personal devices, and enough money engaged in using AI tools more often than those who were poor, and this is evidence of AI inequity that transcends the traditional digital divide (Selwyn, 2022). AI-powered education involves the use of technical equipment and internet connection, and its quality becomes another equity aspect in this case, since poorer students might face extra obstacles to AI learning.

The present study contributes to the existing literature that warns against making general assumptions about AI usage and success, as most of the research into AI in education is conducted in a technologically privileged environment (Selwyn, 2022). Therefore, digital equity must include equitable participation of students in AI learning processes. To do that, some institutional measures have to be considered.

6.5 Implications for Higher Education Policy and Practice

HEIs should use AI as a tool for transforming the pedagogy. In line with the principles of the ConTIS model, the most beneficial educational impact can be achieved when AI is used for fostering constructivism, collaboration, critical thinking and knowledge construction rather than for automation of some academic work. It can be said that there is a huge potential for AI to improve learning and teaching in higher education. To realize this potential, the intricate interrelation between technology, pedagogy, and socio-material conditions is required.

6.6 Limitation

The paper brings context-specific lessons regarding the integration of artificial intelligence in higher education institutions in the Global South, with certain limitations. Firstly, the paper employs only one case study from South Africa and therefore may not be generalized for other universities in the Global South, which may have different technological infrastructure and demographic characteristics. Secondly, the data used for empirical analysis are obtained via questionnaires, focus group discussions, and participants' diaries, which means that the data are open to the subjectivity of perception since direct observation of the prompting was not possible. Lastly, students were free to use any AI platform to generate content and evaluate it.

7. Conclusion

This research explored the experiences and perceptions of AI-mediated learning amongst PGCE students in a South African university environment. It emerged from the findings that, while AI is viewed as a tool for fostering understanding, self-learning, and academic productivity by the participants, there are certain social material conditions under which AI offers educational value. Based on Vygotskian sociocultural constructivism and ConTIS theory, the current study proposes that AI is a mediating digital artefact that can facilitate learning and assist in the process of transforming the learning experience. However, it is shown that the lack of access to adequate amounts of data, poor connectivity, limited device capacity, and load shedding are some of the factors that prevent students from accessing these benefits. In other words, this study shows that using AI in higher education can significantly enhance teacher preparation programmes. At the same time, the effectiveness of its use cannot be measured solely in terms of technological feasibility but requires an examination of pedagogy and other aspects of student life within the institution.

Ethics Declarations

This research was conducted with full ethical clearance from the institutional research ethics committee at the University of Technology. All participation was voluntary, anonymous, and based on written informed consent. No identifying information has been used in the reporting of findings.

AI Declaration

AI tools were used in the preparation of this paper to assist with structural drafting and language editing. All analytical decisions, theoretical interpretations, and empirical findings are the sole responsibility of the authors. All AI-assisted content was reviewed, verified, and substantially revised prior to submission.

References

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), pp. 77–101. Available at: https://doi.org/10.1191/1478088706qp063oa

Cotton, D.R.E., Cotton, P.A. and Shipway, J.R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), pp. 228–239. Available at: https://doi.org/10.1080/14703297.2023.2190148

Creswell, J.W. and Creswell, J.D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches. 5th. Thousand Oaks, CA: Sage.

Crompton, H. and Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, Article 22. Available at: https://doi.org/10.1186/s41239-023-00392-8

Czerniewicz, L., Agherdien, N., Badenhorst, J., Belluigi, D., Chambers, T., Chili, M. et al. (2020). A virus of inequality: Higher education in the time of social distancing. Postdigital Science and Education, 2(3), pp. 946–967. Available at: https://doi.org/10.1007/s42438-020-00116-x

Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), pp. 460–474. Available at: https://doi.org/10.1080/14703297.2023.2195846

Field, A. (2018). Discovering statistics using IBM SPSS Statistics. 5th. London: Sage.

Hamilton, E.R., Rosenberg, J.M. and Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. TechTrends, 60(5), pp. 433–441. Available at: https://doi.org/10.1007/s11528-016-0091-y

Holmes, W., Bialik, M. and Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Boston, MA: Center for Curriculum Redesign.

Kasneci, E., Sessler, K., Küchmann, S., Bannert, M., Dementieva, D., Fischer, F. et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. Available at: https://doi.org/10.1016/j.lindif.2023.102274

Lincoln, Y.S. and Guba, E.G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Luckin, R. (2024). Generative AI and the future of education: A socio-technical perspective on teacher agency. Computers and Education: Artificial Intelligence, 6, Article 100192. Available at: https://doi.org/10.1016/j.caeai.2024.100192

Merriam, S.B. and Tisdell, E.J. (2016). Qualitative research: A guide to design and implementation. 4th. San Francisco, CA: Jossey-Bass.

Miao, F. and Holmes, W. (2023). Guidance for generative AI in education and research. Paris: UNESCO.

Mishra, P. and Koehler, M.J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), pp. 1017–1054. Available at: https://doi.org/10.1111/j.1467-9620.2006.00684.x

Ng, D.T.K., Leung, J.K.L., Chu, S.K.W. and Qiao, M.S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. Available at: https://doi.org/10.1016/j.caeai.2021.100041

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. Paris: UNESCO.

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Cambridge: Polity Press.

Tunjera, N. (2019). Technology integration in teacher education: The ConTIS framework. Doctoral thesis. Cape Town: Cape Peninsula University of Technology.

UNESCO (2023). Guidance for generative AI in education and research. Paris: UNESCO. https://unesdoc.unesco.org/

van Dijk, J. (2020). The digital divide. Cambridge: Polity Press.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Yan, L., Sha, L., Zhao, L., Li, Y., Martínez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), pp. 90–112. Available at: https://doi.org/10.1111/bjet.13370

Yin, R.K. (2018). Case study research and applications: Design and methods. 6th ed. Thousand Oaks, CA: Sage.