Digital Games in Education: Effectiveness and Feasibility Study of *Who is Helena?*

Dušan Žaludko[1], Magda Švecová[2] and Martin Kuruc[3]
[1]Impact Games, Civic Association, Bratislava, Slovakia
[2]University of Ss. Cyril and Methodius in Trnava, Slovakia
[3]Comenius University in Bratislava, Slovakia

dusozaludko@gmail.com
magdalena.svecova@ucm.sk
kuruc@fedu.uniba.sk

Abstract: Past research finds digital game-based learning (DGBL) to have positive effects on student learning outcomes and motivation to learn. However, effectiveness studies are often criticized for lacking scientific rigor or ecological validity, and although best research practices have been proposed, they have not been widely adopted. Compounding this issue, these studies rarely discuss feasibility of recommended study designs or practical issues they encountered during study execution. Given the limited resources typically available for DGBL validation, feasibility studies offer a comparatively affordable yet methodologically rigorous alternative to full-scale effectiveness trials. In light of these critiques, we conducted an ecologically valid longitudinal feasibility study following best practices for DGBL effectiveness research. We investigated the effectiveness of an educational mobile game, *Who is Helena?*, in which players uncover a fictional but factually plausible family history rooted in the Soviet history of Czechoslovakia. Through snowball cluster sampling, we recruited 209 elementary and high school students from 10 pre-existing classes across Slovakia, for whom this topic was part of the current curriculum. Classes were assigned to control (lecture only) and experimental (lecture + DGBL) conditions by matching on variables collected at baseline. Both conditions received 12 standardized lectures. Experimental-group students played the game at home between lectures 9 and 12. To assess effectiveness, we tracked changes in subject-matter knowledge, motivation to learn History, and interest in the topic at baseline, post-intervention, and at follow-up (2-4 weeks later). At follow-up, students in the experimental condition also reported on their experience with the game and teachers provided feedback on the study's materials and procedures. Mixed-effects modeling was used to examine change over time and the between-condition differences in the three variables of interest. We found no evidence that the DGBL intervention was more effective than traditional teaching alone. While some students appreciated the game, only 21% finished it and a majority disliked the forced wait times during gameplay. We discuss the implications of our findings for DGBL design and reflect on the feasibility of best-practice recommendations in view of the practical and methodological challenges encountered during the study.

Keywords: Game-Based Learning, Motivation, Educational Outcomes, History Education, Serious Games, Effectiveness Assessment

1. Introduction

Digital game-based learning (DGBL) refers to some form of use of a digital game (element of fun/entertainment) with an educational purpose/component, balancing both aspects for optimal outcomes (Prensky, 2001). Interest in DGBL has been on a rise during the last two decades. Technological advances and global societal challenges (e.g. the Covid-19 pandemic) have led to the proliferation of DGBL, with formal educational settings arguably being among the most promising areas for its application (see Mikrouli, Tzafilkou and Protogeros, 2024). Against the backdrop of stagnating or decreasing student performance (Jakubowski, Gajderowicz and Patrinos, 2025) and the generally low intrinsic motivation to learn (e.g. OECD, 2024), DGBL has a great potential to address both of these challenges (e.g. Zhang and Yu, 2022; Barz et al., 2024; Ren, Xu, and Liu, 2024).

While DGBL effectiveness research findings generally support the positive impacts of DGBL interventions in education, they also point to issues in the field. Among these issues are low quality of evidence (All, Castellar and Van Looy, 2014; 2016; Aeschbach, de Castro Hüsler and Opwis, 2024) and the lack of published feasibility studies reflecting on methodological and pragmatic aspects of executing this research to a high methodological standard (All, Castellar and Van Looy, 2021). We elaborate on these issues and how this study contributes to this discussion below.

1.1 Quality of Evidence in DGBL Effectiveness Research

DGBL intervention development can be costly, limiting the quantity of DGBLs and leaving few resources for impact validation (Yaman et al., 2024). To avoid resource waste, DGBL development should be informed by high-quality evidence on what makes them effective. Some existing evidence is of questionable quality (All, Castellar and Van Looy, 2016; 2021): quality of study design has been negatively associated with reported effect sizes

(Tamim et al., 2021) while the large heterogeneity in the types and quality of study research designs makes cross-study comparisons difficult (All, Castellar and Van Looy, 2014; 2016).

Best research practices have been proposed to address these issues – guidelines by All, Castellar and Van Looy (2016) are arguably one of the most comprehensive ones, covering all aspects of study design. Even though it has been 10 years since their publication, these (and similar) guidelines have not been widely adopted and the issues in the field persist (All, Castellar and Van Looy, 2016; 2021). Additionally, most of these recommendations come at a cost to ecological validity, which raises questions about the extent to which their implementation in studies of realistic school settings is feasible.

1.2 Feasibility of Study Design Recommendations

Feasibility of best research practices is rarely empirically tested. In general, feasibility studies evaluating study research design and research procedures in education research are rare (Green et al., 2019; All, Castellar and Van Looy, 2021), often unpublished and dismissed by their authors as mere pilots with no added value (Green et al., 2019). However, disseminating such information can help refine best research practices, inform better study designs, and ultimately support the development of more effective DGBL interventions.

Preliminary research – such as pilot, feasibility, or field tests/studies – can help address these issues, as its function is to test the viability of a given project, usually before full-scale implementation (Green et al., 2019; Pauline-Graf & Mandel, 2019). Given that limited resources often prevent the execution of randomized controlled trials (RCTs), the gold standard of intervention research (Rajasekar & Kumar, 2019), field studies/tests are a type of a preliminary studies especially suited for these situations. They are highly ecologically valid (at a cost to internal validity), and relatively highly methodologically rigorous which makes them suitable to test both intervention efficacy and the feasibility of the research methodology itself, at a fraction of a cost of RCTs (Pauline-Graf & Mandel, 2019).

1.3 Current Study

The current paper presents an ecologically valid feasibility field study which: a) tests the effectiveness of a DGBL intervention while following the best research guidelines in order to produce high-quality evidence, and b) reflects on the feasibility and practical aspects of implementing best research practices in real school environments. The DGBL intervention investigated in this study is an educational mobile game *Who is Helena?,* which covers the Soviet history of Czechoslovakia in the second half of the 20th century. The study aims to answer three research questions. **RQ1**: Does the DGBL intervention have an effect on subject-matter knowledge, motivation to learn History, and interest in the subject and does this effect vary across time? **RQ2**: What perceptions and experiences did students report regarding the game intervention, and what implications do these have for designing future DGBL interventions? **RQ3**: What practical and methodological challenges arise when implementing best-practice DGBL research design recommendations within a real, ecologically valid classroom setting?

2. Methods

2.1 Population and Participants

The target population was Slovak 9th grade elementary and 3rd grade high school students covering the Soviet history of Czechoslovakia during the 2025/2026 academic year. Participants were recruited via snowball cluster sampling through a call for teachers in Impact Games' newsletter (almost 3000 recipients). Nine teachers and 209 students were recruited. Mean age was almost 15 years (*min*=14, *max*=19). The sample was 47 % female, 45% male, with 8% other or undisclosed. Students came from varied socioeconomic backgrounds (*M*=7.29, *SD*=1.69, min=1, max=10, 10-point MacArthur Scale of Subjective Social Status, Adler et al., 2000;) across Slovakia.

2.2 Materials

2.2.1 Educational Game Who is Helena?

Who is Helena? is a free mobile messenger detective adventure game (Android/iOS) developed by Impact Games. Players, through the character of 16-year-old Alica, uncover the secret story of her great-aunt Helena,

who emigrated from Czechoslovakia to the United States during the 1968 Prague Spring and never returned. The game uses a smartphone interface - messaging, email, browser, and encyclopedia apps - through which players piece together family history via conversation with relatives and other research. Discovered clues are organized in an in-game "whiteboard" app, and new sections of the encyclopedia become available as the players touch upon subject matter as it arises through gameplay. Players spend most of their time exchanging messages/email, choosing between two response options at each step — these choices are cosmetic, only seldom allowing players to influence the course of the game. The game aims to build curiosity about Czechoslovak history through a family-based narrative, and to develop critical and moral thinking by having players consider the consequences of Helena's emigration — reflected in two possible endings (meeting or not meeting Helena). Gameplay spans roughly four hours across four days.

2.2.2 Lecture Materials

Twelve standardized lectures covering the subject matter were developed with two participating teachers, incorporating feedback from other teachers. All participating teachers were instructed to use them identically. Each followed the same structure: brief revision, new content (dictated student notes), a textbook exercise, and discussion (time-permitting).

2.3 Design and Procedure

The main study used a longitudinal quasi-experimental 2(BS, control vs experimental conditions) x 3(WS: measurement occasion baseline = T_1, post = T_2, follow up = T_3) design. Participants were recruited in December 2025. Testing materials were piloted and revised in January 2026 (see section 2.4.1). The main study took place during January-June 2026. After providing written consent (both students and legal guardians) and being given a randomly generated six-digit identification code (pseudonymization measure), students completed baseline measures (T_1). These included knowledge via paper test; motivation to study History and interest in the subject matter via online questionnaire – the three main dependent variables assessed at each measurement occasion. Control/demographic variables were also collected at this time (2.5.2). As random assignment was not possible given pre-existing class groupings, classes were cluster-matched to conditions using T_1 data, balancing group sizes. All classes then received the same 12 lectures. After the ninth lecture, students in the experimental group were instructed to play *Who is Helena?* at home over the following two weeks due to a) the game's built-in tutorial requiring no additional instructor support and b) its episodic format not being suitable for play during classes. After the twelfth lecture, all students underwent the announced second post measurement (T_2), the only measurement announced in advance. The third follow-up measurement (T_3) took place two to four weeks afterwards, followed by debriefing. At this time, experimental-group students also completed a game-experience questionnaire. To limit practice effects, each student filled a different but equivalent test version at each measurement occasion and graded tests/answers were returned to students after the last measurement. Question order was randomized to reduce answer copying.

2.4 Repeated Measures

Internal consistency (Cronbach's α) of all repeated measures ranged from acceptable to excellent (.71-.93), except Test version A at T_1 (α=.65).

2.4.1 Žaludko1 Learning Outcome

Three equivalent tests (A/B/C) were developed with two History teachers, each comprising 12 open-ended questions evenly covering all lecture content. Six questions required one- or two-word answers assessing basic recall (one point each). The remaining six questions required students to describe, explain, or identify causes and consequences of historical events, three questions worth two points each (two-sentence answers) and three worth three points each (three- to four-sentence answers). Each test version shared four anchor items used to evaluate test equivalence. Test items targeted lower-order thinking (remembering/understanding; Anderson and Krathwohl, 2001) to support grading consistency across instructors. Tests were piloted with 50–57 students per version across nine classes (six schools) who had not yet covered the material; mean scores did not differ significantly (ANOVA), while test reliabilities (Cronbach αs .73-.84) and item difficulties were comparable, supporting equivalence.

2.4.2 Motivation to Learn History

Motivation was measured using an adapted Slovak version (Kuruc, 2017) of the Academic Self-Regulation Questionnaire (Ryan and Connell, 1989), assessing external, introjected, identified, and intrinsic regulation per Self-Determination Theory. We used Relative autonomous indices (RAI) – a composite score based on the scores for each of the motivation regulation styles - for the analyses. Item wording was adapted to reference History class specifically, as done in similar research (e.g. Wang et al., 2017). Students were asked to express their degree of agreement with these statements (v*ery true* (1) to *totally untrue* (4)).

2.4.3 Interest in the Topic of Soviet History of Czechoslovakia

Interest was assessed with a purpose-built measure of individual interest (Hidi and Renninger, 2006), informed by Knekta et al.'s (2020) biology-interest scale. The scale consists of three facets: positive affect (four items; e.g. "Soviet history of Czechoslovakia fascinates me."), perceived value (two items; e.g. "Knowledge of this history is important for my understanding of current day Slovakia."), and intention to re-engage (three items; e.g. This topic interests me even outside the school."). The scale utilized a five-point scale (*absolutely disagree* (1) to *absolutely agree* (5)).

2.5 Other Measures

2.5.1 Game-related Questionnaire

Experimental-group students completed a game-experience questionnaire at T_3, covering game completion, perceived effect on interest and understanding, encyclopedia use, enjoyment (open-ended).

2.5.2 Demographics/Control Variables

In terms of demographics, we measured age, gender, and subjective socioeconomic status (MacArthur Scale; Adler et al., 2000). Other variables included History grades (last three report cards, 1 signifies highest performance, 5 is a fail grade), and game skill (items: 1. When I start playing a new game, I quickly figure out what to do; 2. I usually don't have any trouble with game controls; 3. When it comes to games, I consider myself an experienced player, α=.75; 4-point scale: *absolutely disagree* (1) to *absolutely agree* (4)). Lastly, we asked teachers to give us feedback about the study procedures and materials in a separate survey.

3. Results

3.1 Descriptives and Data Missingness

Table 1 shows means and standard deviations for the main dependent variables for both conditions across three measurement occasions, and for the control variables at baseline. Conditions did not differ on these variables at T_1, suggesting successful matching. Table 1 also shows conditions were unbalanced, as one class was reassigned to control due to failed intervention implementation. Valid responses declined over time with missingness concentrated at T_2 and T_3. For test data, number of valid responses fell by 28% between T_1 and T_3, largely due to missing data from a whole class at T_3 (26 responses). Other missing data reflects natural absenteeism present in education settings. Survey data missingness was a bigger concern. 50% of the responses in the experimental condition (31% at T_1) versus 39.9% in the control (19.3% at T_1) had missing data. Missingness varied between classes, ranging from 30.2% to 60.4%.

Table 1: Means and standard deviations for the three main dependent variables for both conditions across three measurement occasions (T_1, T_2, T_3) groups

	Condition			
	Control		Experimental	
Variables	*n*	*M (SD)*	*n*	*M (SD)*
Dependent				
Test scores (T_1)	123	2.07 (2.81)	74	2.68 (2.95)
Test scores (T_2)	107	11.20 (5.65)	67	8.52 (5.92)

	Condition			
	Control		Experimental	
Test scores (T_3)	99	13.10 (6.03)	43	12.80 (5.73)
RAI (T_1)	106	-.45 (1.85)	59	-.30 (1.90)
RAI (T_2)	62	-.28 (2.18)	51	-.53 (1.59)
RAI (T_3)	53	-.15 (1.97)	29	-.55 (1.86)
Interest (T_1)	106	3.00 (.96)	59	3.36 (.86)
Interest (T_2)	62	3.19 (.84)	51	3.26 (.90)
Interest (T_3)	53	3.11 (.86)	29	2.95 (.99)
Independent				
SES (T_1)	108	7.55 (1.72)	59	6.88 (1.87)
Average grade (T_1)	108	1.47 (.66)	59	1.67 (.78)
Game skills (T_1)	108	3.22 (.70)	59	3.20 (.71)

3.2 Learning Outcome

We fitted a mixed model to analyze how knowledge of the subject matter, in the form of test scores, developed across time (T_2 vs T_3) and between conditions, accounting for test version and baseline levels of: test scores, RAI, interest in the topic, average History grade, SES, and game skill (Table 2). Test version did not influence average test scores at baseline (all *ps*>.05), while anchor-item composite scores did not differ between test versions either (one-way ANOVA $F(2, 194)=1.45$, $p=.237$). Both results support test version equivalence, allowing score comparisons across test versions.

Table 2: Results of a parametric mixed model for the dependent variables of test scores, RAI, and Interest.

	Model dependent variable								
	Test scores			RAI			Interest		
Variable	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Fixed effects									
(Intercept)	11.16	3.06	**<.001**	-1.88	0.97	.053	3.68	0.40	**<.001**
test score (T_1)	0.42	0.13	**.001**	0.10	0.04	**.020**	0.04	0.02	.064
RAI (T_1)	0.27	0.21	.185				0.18	0.03	**<.001**
Interest (T_1)	0.08	0.42	.844	0.58	0.14	**<.001**			
SES (T_1)	0.16	0.19	.398	-0.01	0.07	.884	-0.03	0.03	.435
Grade (T_1)	-2.32	0.53	**<.001**	-0.49	0.18	**.008**	0.01	0.09	.915
Game skill (T_1)	-0.21	0.43	.622	0.07	0.16	.656	-0.06	0.08	.427
Test version B[a]	-0.12	0.57	.834						
Test version C[a]	0.38	0.59	.524						
Condition:control[b]	1.92	2.18	.381	0.07	0.43	.877	-0.27	0.19	.151
Measurement T_2[c]				-0.34	0.21	.105	-0.13	0.10	.193
Condition:control × measurement T_2[bc]				0.21	0.28	.452	0.13	0.14	.322
Measurement T_3[c]	1.15	0.88	.191	-0.16	0.26	.528	-0.25	0.13	**.049**
Condition:control × measurement T_3[bc]	0.73	1.03	.481	0.20	0.32	.528	0.31	0.16	.051
Random effects									
$\sigma^2_{(residuals)}$			11.53			1.13			0.27
$\tau^2_{00(class)}$			6.79			1.55			0.39

	Model dependent variable		
	Test scores	**RAI**	**Interest**
$\tau^2_{00(student)}$	9.86	0.26	0.04
N: class/student/observation	10/155/254	10/165/360	10/165/360
Marginal R^2/Conditional R^2	.23/.69	.18/.68	.20/.69

Note. [a]Compared to test version A. [b]Compared to experimental condition. [c]Compared to T_2 in the test score model, but compared to T_1 in the RAI and Interest models.

Given the small number of clusters in our sample (10 classes), REML estimation (more accurate variance estimates) and Kenward-Roger degrees of freedom (more appropriate fixed effect estimates) were used across all analyses (McNeish, 2017).

Empty model's ICC (39.8%) indicated substantial between-class variance. In relation to RQ1, we did not find any effect of condition, time of measurement, or their interaction (all *p*s>.191). We found a significant effect of baseline test score (*b*=0.42, *p*=.001). For every point achieved on the test at baseline, later test scores were 0.42 points higher. Grade was the only other significant predictor (*b*=-2.32, *p*<.001) - students with better grades (decrease of 1) scored higher (increase by 2.32 points). Cluster-robust standard errors sensitivity check, addressing heteroscedasticity of level-1 residuals, corroborated the parametric model results.

3.3 Motivation Outcomes

3.3.1 Motivation to Learn History - RAI

We fitted a mixed model to examine how motivation to study History, in the form of RAI, changed across time (T_1-T_3) and between conditions, accounting for baseline measures of: test score, interest in the topic, average History grade, SES, and game skill (Table 2).

The ICC of the empty model was about half of the test score model - 21.2%. Effects of condition, measurement occasion, nor their interaction were significant (all *p*s>.105). Baseline test score (*b*=0.10, *p* =.020), baseline interest (*b*=0.58, *p*<.001), and average History grade (*b*=-0.49, *p*=.008) all significantly predicted RAI. Students with better grades, higher baseline test scores and interest in the subject matter had a more autonomous and self-regulated motivation to learn History (higher RAI). In light of heavy-tailed level 1 residuals (double-sided), a robust mixed model with t-distributed residuals was fitted as a sensitivity check. Results were consistent with the parametric model, except for an additional significant decrease in RAI between T_1 and T_2 (*b*=-0.34, *p*=.044).

3.3.2 Interest in Subject Matter

A parallel model examined how interest in the subject matter changed across T_1–T_3 and between conditions, controlling for baseline test score, RAI, grade, SES, and game skill (Table 2).

The empty model's ICC was 22.3%, similar to the RAI model. We did not find any effect of condition or the condition x time interactions (all *p*s>.051). Interest decreased modestly from baseline to T_3 (*b*=-0.25, *p*=.049). Baseline RAI significantly predicted interest (*b*=0.18, *p*<.001) – for every point in RAI at baseline, interest in the subject matter increased by 0.18.

Due to violations of homoscedasticity and normality of level 1 residuals, we utilized cluster-robust standard errors and robust mixed model with t-distributed residuals as sensitivity checks. These were largely in line with the parametric results with the exception of a small but significant effect of baseline test score (*b*=0.04, *p*=.047) in the robust model.

3.4 Students' Game Evaluation

Of 60 responses about students' experience with the DGBL, only 13 (21%) completed the game, while 40 (67%) got halfway through, the remainder not finishing. Most students (65%) did not use or see a need for the in-game Encyclopedia, while 30% checked new entries as they appeared, and 18% explored beyond provided information. The majority of the written feedback was negative: 21 students disliked long wait times (gameplay disruption), 11 found the game too long, 8 found it boring, and 4 felt they had too little power to influence the story, making it feel non-interactive. Positive feedback centered around the story (16 students), game's

characters (6), and the detective-style gameplay design (6, "…it felt like a mobile version of an Agatha Christie book").

3.5 Teacher Feedback

Eight teachers provided feedback. Lecture materials were seen as too long for a single class, overly extensive in required student notes, and insufficiently engaging for both teachers and students, prompting three teachers to deviate from the provided materials. Testing materials and the test data entry method were well received, although several teachers would have liked a more robust survey-completion check. Three teachers suggested having access to a clear timeline of research activities and deadlines.

4. Discussion

By applying best research practices, this feasibility study investigated the effectiveness of an educational mobile game *Who is Helena?,* contributing to addressing both the often-low quality of study design (and hence generated evidence) and the lack of feasibility studies validating best research practices. Our investigation was centered around three research questions.

While we found evidence for the predictive roles of grade and baseline measures of variables of interest, consistent with suggested best research practices (All, Castellar and Van Looy, 2016), we found no evidence that the DGBL intervention improved subject-matter knowledge, motivation to learn History, or interest in the topic beyond what lecture-based teaching alone achieved (RQ1). Findings related to research questions 2 and 3 suggest possible explanations for the null pattern.

The game-evaluation data offer a plausible explanation for the null effect and insight for future DGBL design (RQ2). Only 21% of students completed *Who is Helena?* One likely explanation is that assigning the game as homework (due to its episodic design) resulted in low engagement. Future DGBL interventions intended for school settings should aim to be easy to integrate into classroom lessons. Negative feedback consistently centered on the occasional long forced breaks between gameplay sessions – friction that likely reduced rather than promoted engagement. Future DGBL interventions should optimize such design features to better sustain engagement.

Turning to feasibility of best research practices and practical challenges encountered during the study (RQ3), several practical and methodological challenges emerged that echo concerns raised in prior literature (Green et al., 2019; All, Castellar and Van Looy, 2021). Notably, several of these challenges also constitute limitations of the present study, illustrating precisely the kind of practical trade-offs feasibility research is intended to surface.

Data were collected from students nearing the end of elementary school who had already secured a high school place. These circumstances may have reduced their motivation to engage fully as study participants, helping explain the substantial data missingness — particularly for survey data (up to 60% in some classes), concentrated at later measurement occasions. Missingness was also driven by insufficient means for teachers to verify survey completion; more robust verification approaches are needed in future research.

For ethical reasons, teachers were not restricted in their teaching between post- and follow-up measurements, limiting our control over this period and likely biasing follow-up results — suggesting that, contrary to best practices, unrestricted follow-up periods may not be a feasible design element in school settings.

A further issue not fully addressed by best research practices (All, Castellar and Van Looy, 2016) concerns condition assignment and its effect on statistical power. While random participant-level condition assignment is difficult in classroom environments, classroom level assignment may introduce considerable statistical power limitations. In our case, the small number of clusters (10) limited our ability to detect condition and condition x time effects, potentially contributing to the null findings.

Between-class variance was also substantially higher for test scores than for RAI or interest, possibly reflecting inconsistent grading of open-ended test items across teachers (a bias multiple-choice items could avoid) or instructor effects - three teachers admitted deviating from the standardized lectures despite repeated instructions, likely further reducing internal validity. Better communication and lecture materials better suited to teachers' needs could help address this, for example by providing teachers with an easily accessible clear visual timeline of deadlines and study milestones.

5. Conclusion

The DGBL effectiveness field faces several issues, including low evidence quality and a lack of feasibility studies of best practices meant to improve it. This feasibility study contributes to both: while applying best-practice guidelines, we found no evidence of the DGBL intervention outperforming traditional teaching on any outcome, and we offer a transparent account of the practical and methodological challenges encountered in implementing rigorous DGBL research in real classrooms. We hope this study encourages similar feasibility studies and informs both future DGBL development and subsequent effectiveness research in ecologically valid settings.

Acknowledgements

The study is a partial outcome of two projects. Firstly, the study is a part of the project 2024-1-SK01-KA220-HED-000252451 Learn to Play for the Future financed by Erasmus+ KA220. Secondly, the study is financed from an FPU grant, contract n. 25-372-02208, Ministry of Culture of the Slovak Republic (henceforth Ministry). The support of the European Commission and of the Ministry for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission nor the Ministry cannot be held responsible for any use which may be made of the information contained therein.

Ethics Declaration

This research required ethics clearance, which was granted by the Ethical Committee for Sciences and Research at the Faculty of Mass Media Communication at the University of Ss. Cyril and Methodius in Trnava, Slovakia (reference number: 2025/FMK-UCM/06).

AI Declaration

AI was used to help identify relevant literature, to help with language editing/proofreading, and to check the logical structure of the text written by the authors. Any AI suggested edits were thoroughly reviewed before being implemented.

References

Adler, N.E., Epel, E.S., Castellazzo, G. and Ickovics, J.R. (2000) Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health psychology*, *19*(6), pp 586-592. https://doi.org/10.1037/0278-6133.19.6.586

Aeschbach, L.F., de Castro Hüsler, A.B. and Opwis, K. (2024) Systematic review on the methodological rigor and effectiveness of digital game-based learning in education. *preprint, PsyArXiv, May*, *3*. https://doi.org/10.31234/osf.io/6qgvp

All, A., Castellar, E.N.P. and Van Looy, J., (2021) Digital Game-Based Learning effectiveness assessment: Reflections on study design. *Computers & Education*, *167*, pp 104160. https://doi.org/10.1016/j.compedu.2021.104160

All, A., Castellar, E.P.N. and Van Looy, J. (2014) Measuring effectiveness in digital game-based learning: A methodological review. *International Journal of Serious Games*, *1*(2). pp 3-20. https://doi.org/10.17083/ijsg.v1i2.18

All, A., Castellar, E.P.N. and Van Looy, J., (2016) Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92*, pp.90-103. https://doi.org/10.1016/j.compedu.2015.10.07

Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Barz, N., Benick, M., Dörrenbächer-Ulrich, L. and Perels, F. (2024) The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research*, *94*(2), pp 193-227. https://doi.org/10.3102/00346543231167795

Hidi, S. and Renninger, K.A. (2006) The four-phase model of interest development. *Educational psychologist*, *41*(2), pp 111-127. doi: 10.1207/S15326985EP4102_4

Jakubowski, M., Gajderowicz, T. and Patrinos, H.A. (2025) COVID-19, school closures, and student learning outcomes. New global evidence from PISA. *NPJ Science of Learning*, *10*(1), pp 1-7. https://doi.org/10.1038/s41539-025-00297-3

Knekta, E., Rowland, A.A., Corwin, L.A. and Eddy, S. (2020) Measuring university students' interest in biology: Evaluation of an instrument targeting Hidi and Renninger's individual interest. *International Journal of STEM Education*, *7*(1), 23. https://doi.org/10.1186/s40594-020-00217-4

Kuruc, M. (2017) Príručka pre používanie dotazníka SRQ-Academic v pedagogickej praxi. Bratislava, Univerzita Komenského.

McNeish, D. (2017) 'Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction', *Multivariate Behavioral Research*, 52(5), pp. 661–670. https://doi.org/10.1080/00273171.2017.1344538

Mikrouli, P., Tzafilkou, K. and Protogeros, N. (2024) Applications and learning outcomes of game-based learning in education. *International Educational Review*, *2*(1), pp 25-54. https://doi.org/10.58693/ier.212

OECD (2024) *PISA 2022 Results (Volume V): Learning Strategies and Attitudes for Life*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/c2e44201-en.

Pauline-Graf, D. & Mandel, S. (2019) 'Defining preliminary research for digital game-based learning evaluation: best practices', *International Journal of Educational Methodology*, vol. 5, no. 4, pp 623-635. https://doi.org/10.12973/ijem.5.4.623

Prensky, M. (2003) Digital game-based learning. *Computers in entertainment (CIE)*, *1*(1), pp 21-21. https://doi.org/10.1145/950566.950596

Rajasekar, A., & Kumar, V. (2019) Randomized controlled trials: Gold standard of evidence. Drug Invention Today, 12(6), 1215-1217.

Ren, J., Xu, W. and Liu, Z. (2024) The impact of educational games on learning outcomes: Evidence from a meta-analysis. *International Journal of Game-Based Learning (IJGBL)*, *14*(1), pp 1-25. https://doi.org/10.4018/IJGBL.336478

Ryan, R.M. and Connell, J.P. (1989) Perceived locus of causality and internalization: examining reasons for acting in two domains. *Journal of personality and social psychology*, *57*(5), pp 749-761. https://doi.org/10.1037/0022-3514.57.5.749

Shawn Green, C., Bavelier, D., Kramer, A.F., Vinogradov, S., Ansorge, U., Ball, K.K., Bingel, U., Chein, J.M., Colzato, L.S., Edwards, J.D. and Facoetti, A. (2019) Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*(1), pp 2-29. https://doi.org/10.1007/s41465-018-0115-y

Tamim, R.M., Borokhovski, E., Bernard, R.M., Schmid, R.F., Abrami, P.C. and Pickup, D.I. (2021) A study of meta-analyses reporting quality in the large and expanding literature of educational technology. *Australasian Journal of Educational Technology*, *37*(4), pp 100-115. https://doi.org/10.14742/ajet.6322

Wang, C.K.J., Liu, W.C., Nie, Y., Chye, Y.L.S., Lim, B.S.C., Liem, G.A., Tay, E.G., Hong, Y.Y. and Chiu, C.Y. (2017) Latent profile analysis of students' motivation and outcomes in mathematics: An organismic integration theory perspective. *Heliyon*, *3*(5). https://doi.org/10.1016/j.heliyon.2017.e00308

Yaman, H., Sousa, C., Neves, P.P. and Luz, F. (2024) Implementation of Game-Based Learning in Educational Contexts: Challenges and Intervention Strategies. *Electronic Journal of e-Learning*, *22*(10), pp 19-36. https://doi.org/10.34190/ejel.22.10.3480

Zhang, Q. and Yu, Z., (2022) Meta-analysis on investigating and comparing the effects on learning achievement and motivation for gamification and game-based learning. *Education Research International*, *2022*(1), pp 1-19. https://doi.org/10.1155/2022/1519880