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.

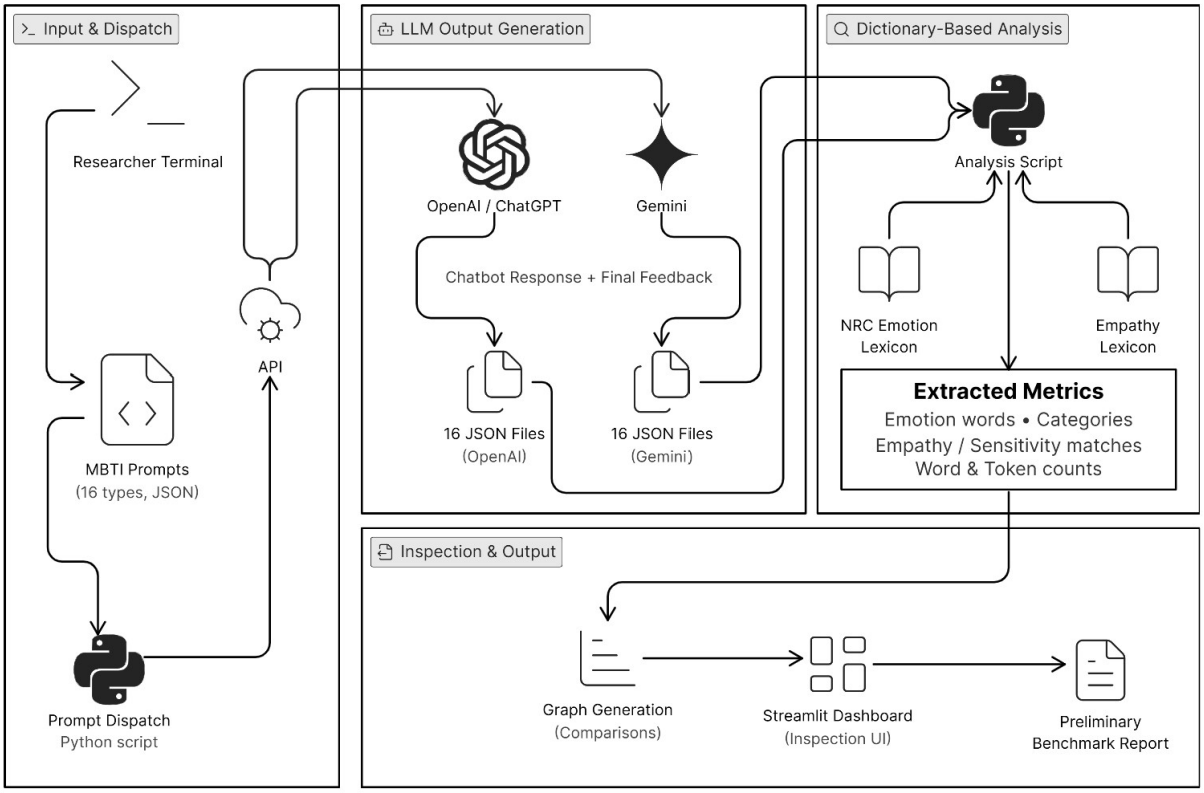


Figure 1: SMAP-ST system architecture and feedback-generation workflow.

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.

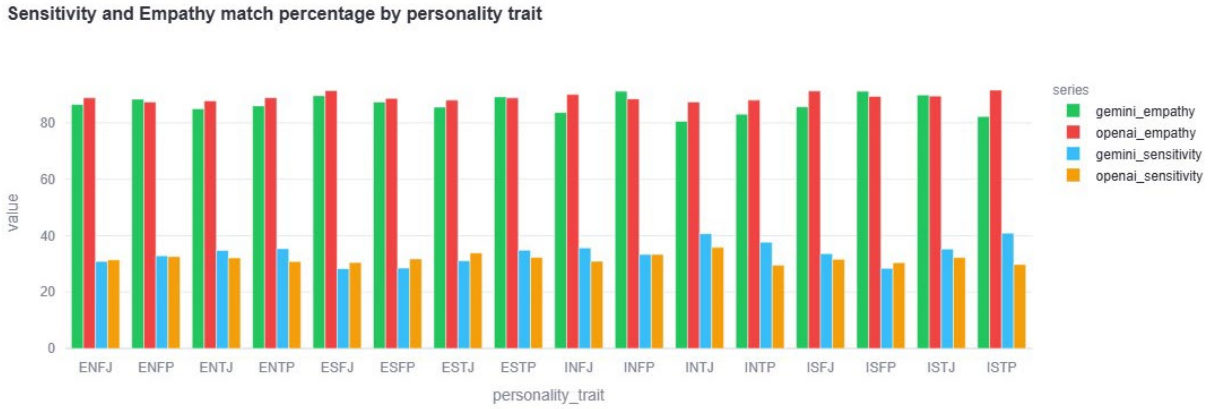


Figure 2: Sensitivity and empathy match percentage by personality.

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.

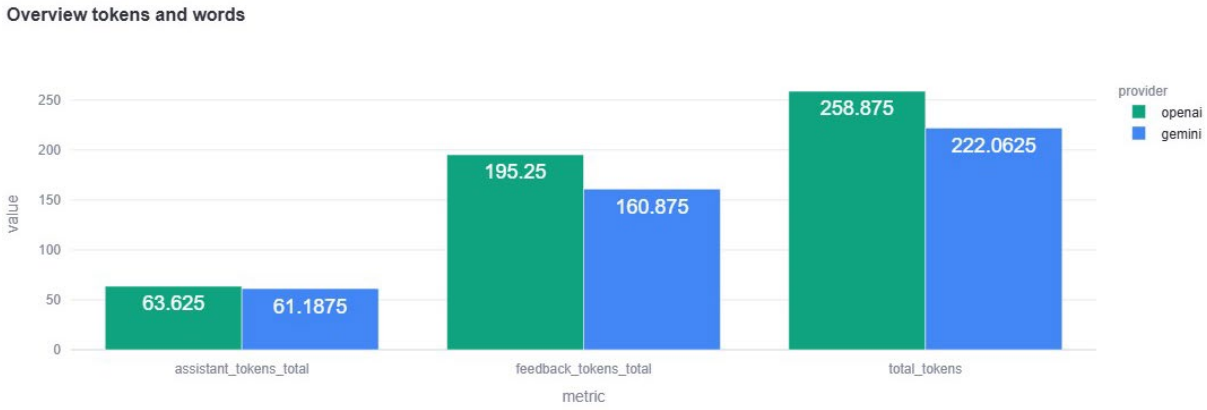


Figure 3: Comparison between OpenAI/ChatGPT and Google/Gemini on average token volumes: assistant tokens, feedback tokens, and total tokens.

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.