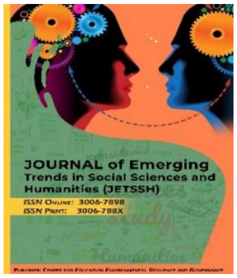

Journal of Emerging Trends in Social Sciences and Humanities

A Double-Blind, Peer-Reviewed, HEC recognized Y-category Research Journal

E-ISSN: 3006-7898 **P-ISSN:** 3006-788X

Volume 3, Issue 2, 2026, 113-121

Retrieval-Augmented Generative AI for Academic Viva Preparation: Effects on Anxiety, Confidence, and Oral Performance

Shahab Qaisar[1] Airah Mehmood[2] Arslan Haider[3]

[1] MS Scholar, Muhammad Nawaz Shareef University of Agriculture, Multan, Punjab, Pakistan.
[2] MPhil Scholar, University of Education, Lahore, Multan Campus, Punjab, Pakistan.
[3] MPhil Education, University of Education Lahore, Multan Campus, Punjab, Pakistan.

Corresponding Author: Airah Mehmood, **Email:** airahmehmood786@gmail.com

Article Information

Received:

2026-04-13

Revised:

2026-05-11

Accepted:

2026-06-16

Keywords

Generative Artificial Intelligence; Retrieval-Augmented Generation; Academic Viva; Oral Assessment; Academic Anxiety; Student Confidence; Formative Feedback; Higher Education; Thesis Defense.

ABSTRACT

University students need to face high-stakes oral exams - such as academic viva examinations, thesis defenses, proposal presentations and others - that can cause a significant amount of anxiety. This methodological training manuscript illustrates the evaluation procedure of a retrieval-augmented generative artificial intelligence system to prepare a personalized academic viva. The system accesses uploaded academic documents and automatically creates questions that could be asked by an examiner, grades the responses, gives formative feedback to the student, and produces a personal performance summary. A synthetic pre-post dataset was created to demonstrate the proposed analysis for 30 students. In the results, mean academic anxiety decreased from 6.18 ($SD = 0.86$) to 3.44 ($SD = 0.92$), $t(29) = -23.48$, $p < .001$, Cohen's $d_z = -4.29$. Mean academic confidence increased from 4.17 ($SD = 0.69$) to 6.65 ($SD = 0.86$), $t(29) = 19.87$, $p < .001$, $d_z = 3.63$, while rubric-based oral performance increased from 2.95 ($SD = 0.46$) to 4.07 ($SD = 0.47$), $t(29) = 19.17$, $p < .001$, $d_z = 3.50$. The possibility of 4 themes emerged from the synthetic reflections: increased preparedness, decreased fear of questioning, improved response structure, and ongoing need for human oversight. The findings provided here are not empirical and are meant to be used for data collection, statistical analysis and reporting in APA style.

© 2026 by the authors. Licensee Qaisar, Mehmood & Haider. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license https://creativecommons.org/licenses/by/4.0/

Publisher: Institute for Educating Environmental Resilience and Governance

1. Introduction

In higher education, the oral examination is a common tool for evaluating students' conceptual knowledge, research skills, critical thinking, oral communication, and decision-making on academic issues. The skills students develop in thesis viva, research proposal defence, synopsis presentation, comprehensive examination and seminar presentation go beyond disciplinary knowledge; these skills involve the ability to communicate their knowledge in an evaluative context (Grubert et al., 2025).

Many students experience anxiety when facing supervisors, external examiners, evaluation committees or audiences, even though they are well schooled academically. This fear can come from fears of questions that will be raised, worries about being judged negatively, lack of practice defending arguments, or lack of opportunity to practice in realistic situations. High levels of anxiety can hinder recall, explanation, organisation, voice control, confidence and critical response (Ansari et al., 2025).

Typical viva preparation typically includes repeatedly reading the thesis or proposal, anticipating questions, practicing with fellow students and holding mock vivas organized by supervisors. These are useful techniques but students may not be able to have supervision and adequate peers all the time. Some feedback may also be vague, generic or weak. Students may therefore go into high stakes oral assessments with inadequate exposure to diverse and document-specific questioning (Huang, 2026).

Generative AI opens the door to new opportunities in personalized academic preparation. Conversational question, explanation, summary and feedback generation are possible with LLM's. When used without being based on the student's actual academic document, however, general-purpose models can produce inaccurate, overly broad or contextually irrelevant content. To overcome this limitation, retrieval-augmented generation (RAG) is a method that takes relevant text from a provided knowledge source and uses it to aid the model's response-generation (Yusuf et al., 2025).

The system examined in this study was designed to retrieve content from students' uploaded theses, proposals, synopses, and presentations. It then generated examiner-style questions, assessed written responses, provided immediate formative feedback, and produced a performance summary. The intervention therefore combined document-grounded questioning, repeated rehearsal, individualized feedback, and a low-pressure practice environment (Lee & Eronen, 2025).

University students frequently experience anxiety and low confidence during academic oral examinations. Conventional preparation methods may provide insufficient exposure to unpredictable examiner-style questions and may not offer immediate, personalized feedback. Although generative AI tools can support practice, their responses may be disconnected from the student's actual research document. There is therefore a need to examine whether retrieval-augmented generative AI can provide relevant, document-grounded viva preparation and whether such preparation is associated with improvements in anxiety, confidence, and oral performance. The purpose of this study was to evaluate the use of a retrieval-augmented generative AI system as a preparatory intervention for academic vivas and related oral assessments. Specifically, the study examined changes in students' presentation-related anxiety, academic confidence, preparedness, and rubric-based oral performance after structured interaction with the system.

Research Questions

1. How does retrieval-augmented AI-supported viva practice affect students' academic oral-assessment anxiety?
2. How does the intervention affect students' confidence in responding to examiner questions?
3. What changes occur in students' rubric-based oral performance after using the system?
4. How do students perceive the relevance and usefulness of AI-generated questions and formative feedback?

2. Literature Review

The rapid growth of generative AI has reached its boom phase in higher education. AI-based systems provide support when brainstorming, explaining content, improving language, preparing for assessments, giving feedback, tutoring, and providing personalized learning.

Such applications can help in providing immediate support and enhance access to learning support outside the classroom time (Beale, 2025).

The educational benefits of generative AI rely on its outputs being of good quality, relevant, and correct. General purpose systems can give plausible but incorrect answers, especially when questions are of a kind that involves information in a specific thesis, course document, institutional policy or research proposal. This is a significant point of limitation in academic vivas as questions should be related to the research problem, methodology, theory, evidence, and conclusions of the student's research (Smith, 2025).

Retrieval augmented generation is the integration of retrieval with generation. A document is broken up into smaller parts, or chunks. The chunks are then converted to numerical representations and stored in a vector database. The system looks for the most relevant pieces when a user queries a question or initiates a task and provides these pieces as context to the language model (Blessing, 2026).

In the case of document-grounded questioning, there's a place for retrieval augmentation in viva preparation. It could pose more specific questions about the student's research design, sampling procedure, variables, results, or theory rather than asking a generic question like ‘What is your methodology. This can enhance the authenticity and value of the rehearsal experience (Puleio et al., 2024).

Academic anxiety is defined as a combination of emotional stress, apprehension, worry and physical discomfort in evaluative situations. Oral examinations may generate greater anxiety than some written assessments because students must respond immediately, communicate in front of evaluators, and manage follow-up questions (Perez Verástegui et al., n.d.).

While moderate anxiety can spur preparation, high anxiety can negatively affect working memory, concentration, verbal fluency, confidence and organization. Some students may have the correct answer but be unable to articulate their response. Practice in a psychologically safe setting could decrease uncertainty and familiarity with questioning structure (Németh et al., 2025).

Self-efficacy is related to academic confidence, which is an individual's belief in his or her ability to perform a certain task. Self-efficacy is associated with higher willingness to try difficult questions, manage anxiety, get going from a mistake and persevere when faced with a challenge (Zapata, 2025).

Structured practice can bolster confidence through mastery experiences. As students provide responses to progressively more challenging questions, they are given corrective feedback and are able to see their own improvement, they may become more confident in their performance when it comes to the viva. Feedback, however, needs to be constructive, understandable, specific and linked to clear performance criteria (Pramanik & Husin, 2025).

Formative feedback helps learners by revealing the difference between what they do and what they should do. The following constructive feedback is helpful for a viva preparation task: factual accuracy, conceptual depth, coherence, academic language, argumentation, relevance, completeness (Philippakos et al., 2026).

Students can get instant feedback to make corrections as they are still thinking about the question and reasons. A conversational AI system can give repeated practice without embarrassment or fear of judgement. However, automated feedback should not be seen as definitive feedback due to the potential of the AI system to miss disciplinary subtleties, to have incorrect interpretations of correct answers, and to employ varying standards (Voultsiou & Moussiades, 2026).

Research Gap

There have been some previous studies on AI tutoring, automated question generation, personalized learning, speaking assessment, and presentation feedback. There has been less research that has examined systems that incorporate retrieval from a student's academic document, the generation of examiner-style questions, the evaluation of the responses, the provision of immediate formative feedback, anxiety reducing rehearsal and the reporting of the performance. The present study addressed this gap by evaluating a retrieval-augmented generative AI system as a structured intervention for academic viva preparation.

3. Theoretical Framework

Self-efficacy theory and principles of formative assessment were used to inform the study.

Self-efficacy theory suggests that the self-efficacy of learners affect their efforts, persistence, emotional responses and performance on the task. The intervention was predicted to boost SE by providing practice, manageable challenges, feedback and opportunities to adjust responses (Pramanik & Husin, 2025).

Formative assessment focuses on the use of evidence of learners' performance to inform improvement prior to summative assessment. In the current system, students' understanding was identified by answering questions generated by AI, and their revisions were identified by feedback generated by AI (Philippakos et al., 2026).

The hypothesis of the relationship between the study variables is as follows: retrieval-augmented viva practice to document-grounded rehearsal and formative feedback to prepare and build up individuals' confidence, reduce anxiety and enhance oral performance. This model does not require that reducing anxiety will lead to an improvement in performance. Instead, it suggests that frequent rehearsal, greater familiarity with what questions are likely to come up, better ways of organizing the content, and revision that is supported by feedback all help to improve performance (Voultsiou & Moussiades, 2026).

4. Research Methodology

This study used a pre–post design with a single group with qualitative student reflections. It was quantitatively analyzed the change of anxiety, confidence and oral performance, and qualitatively collected the student's comments on the use of the preparation system supported by AI. The study was a pilot intervention, and thus was not designed to establish definitive causal effects, but to gain a preliminary evaluation of feasibility and educational usefulness.

The subjects included 30 students from a public sector university in Pakistan. A total of 18 pre-service and 12 postgraduate students from education and STEM programmes were selected for the sample. Participants were purposively selected because they were preparing for academic presentations, proposal defenses, synopsis defenses or thesis vivas. There was no compulsion for participants and informed consent was obtained prior to data collection.

Description of the Intervention

The intervention employed an AI-based retrieval-augmented generative system created for academic viva preparation. Students submitted an academic document (e.g., thesis, proposal, synopsis or presentation). The system extracted specific parts of the document and created questions tailored to the examiner's style, based on the uploaded text.

Usually there were five questions that were given in a practice sequence, but this could be varied. Questions included research problem, literature review, theoretical framework, methodology, findings, originality, limitations and implications. This system has been put together in Python and the front end has been created using Gradio. The text and structured information in uploaded documents were extracted using PDF-processing libraries. A large language model created questions and feedback, and a vector database enabled semantic retrieval.

The technological workflow consisted of four main layers: a document-processing layer, in which documents were processed for extraction, cleaning, segmentation, and preparation; a retrieval layer, in which document-grounded questions were created, followed-up, and fed back to the document, and relevant chunks were retrieved and selected; a generation layer, in which document-grounded questions were produced, followed-up, given feedback, and suggestions for improving them; and an interaction layer, where the documents were uploaded, the response was entered, questions were navigated, the session was managed, and the performance report was exported.

Instruments

A self-reported measure was used to assess presentation-related anxiety. The items addressed emotional tension, apprehension about examiner questioning, fear of making mistakes, concern about negative evaluation, and discomfort during oral academic assessment. Participants completed the measure before and after the intervention, and higher scores represented greater anxiety.

Academic confidence was assessed through items related to readiness, perceived understanding, ability to explain research decisions, confidence in answering follow-up questions, and belief in the ability to perform successfully during an oral assessment. Higher scores represented greater confidence.

Participants' mock oral responses were evaluated through a rubric covering conceptual understanding, accuracy and relevance, depth of explanation, organization and coherence, academic language, justification of research decisions, and confidence of response. The same rubric dimensions were applied during the pre-intervention and post-intervention assessments.

Reflective Responses

Participants provided brief reflections on the relevance of the generated questions, usefulness of the feedback, confidence during practice, perceived anxiety, and overall preparedness.

Data-Collection Procedure

Before using the system, participants completed the anxiety and confidence measures and participated in a mock oral-assessment activity. Their responses were evaluated through the performance rubric.

Participants were then oriented to the system and completed structured practice sessions using their academic documents. During the practice, they answered generated questions, reviewed feedback, and revised selected responses.

After the intervention, participants completed the same anxiety and confidence measures and participated in a second mock assessment. They also submitted reflective comments regarding their experience.

Synthetic reflective responses were analysed through a worked example of thematic analysis. The process involved repeated reading, initial coding, comparison of related codes, development and review of broader themes, and selection of representative excerpts. The synthetic dataset was screened for missing values, impossible values, and extreme scores. Means and standard deviations were calculated for academic anxiety, academic confidence, and rubric-based oral performance at the pre-intervention and post-intervention stages. Internal consistency was examined using Cronbach's alpha. Paired-samples t tests were conducted because the same participants were measured twice. Mean paired differences were accompanied by 95% confidence intervals, and Cohen's dz was calculated by dividing the mean difference by the standard deviation of the paired differences. Statistical significance was evaluated at alpha = .05. For the actual study, the assumptions of the paired-samples t test should be assessed using the distribution, histogram, Q-Q plot, and outliers of each change score. A Wilcoxon signed-rank sensitivity analysis may be reported when serious violations occur.

Reflective responses were examined through thematic analysis. Initial codes were developed from repeated ideas in the responses, after which related codes were grouped into broader themes. The final report should include exact means, standard deviations, test statistics, degrees of freedom, probability values, confidence intervals, reliability coefficients and effect sizes from the original data.

Ethical Considerations

Students were told about the purpose of the study and participation was voluntary. They were told that the feedback they received from the AI was not to be used as an actual assessment of their academic work. PI was not needed for system interaction. Students were also encouraged to delete any sensitive, confidential or unpublished information from documents as appropriate. Formal academic guidance and evaluation were kept with human supervision.

Results

Table 1: *Pre- and Post-Intervention Results*

Outcome	Pre M (SD)	Post M (SD)	Mean change [95% CI]	t(29)	p	Cohen's dz
Academic anxiety	6.18 (0.86)	3.44 (0.92)	-2.74 [-2.98, -2.50]	-23.48	< .001	-4.29
Academic confidence	4.17 (0.69)	6.65 (0.86)	2.47 [2.22, 2.73]	19.87	< .001	3.63

Oral performance	2.95 (0.46)	4.07 (0.47)	1.12 [1.00, 1.24]	19.17	< .001	3.50

Note. All values are synthetic and are included solely to demonstrate the proposed analysis and APA-style reporting. Cohen's dz is the paired-samples standardized mean difference.

Preliminary Data Screening and Reliability

There were 30 complete cases and no out-of-range responses in the synthetic data set. The Cronbach's alpha of the pre- and post-test anxiety and confidence scores were above .70, which is considered acceptable for internal consistency. Cronbach's alpha level for pre- and post-test anxiety and confidence scores were above .70, which is considered acceptable for internal consistency. These coefficients are illustrative only and will be different for each set of data.

Academic Anxiety

For the synthetic data, academic anxiety decreased after the intervention. The pre-intervention mean was 6.18 (SD = 0.86), whereas the post-intervention mean was 3.44 (SD = 0.92). Overall, the mean change for the post-minus-pre was -2.74 points, or -44.3% of the pre mean.

A paired-samples to test indicated that the reduction was statistically significant, t (29) = -23.48, p < .001, 95% CI [-2.98, -2.50], Cohen's dz = -4.29. The negative sign means that the anxiety after the intervention was less than before the intervention. This very large effect was created with the purpose of being included in the teaching, and cannot be found in the real study.

APA-style example: Academic anxiety decreased significantly from pre-intervention (M = 6.18, SD = 0.86) to post-intervention (M = 3.44, SD = 0.92), t (29) = -23.48, p < .001, 95% CI [-2.98, -2.50], dz = -4.29.

Academic Confidence

Academic confidence increased in the synthetic dataset from a pre-intervention mean of 4.17 (SD = 0.69) to a post-intervention mean of 6.65 (SD = 0.86). The average increase was 2.47 points.

The paired-samples analysis showed a statistically significant increase, t (29) = 19.87, p < .001, 95% CI [2.22, 2.73], Cohen's dz = 3.63. This finding can be interpreted as an increase in confidence, but this will need to be substantiated with actual scores and accounts of the participants in the final study.

APA-style example: Academic confidence increased significantly from pre-intervention (M = 4.17, SD = 0.69) to post-intervention (M = 6.65, SD = 0.86), t (29) = 19.87, p < .001, 95% CI [2.22, 2.73], dz = 3.63.

Oral Performance

Oral performance also showed an improvement in rubric based oral performance in the data set. The mean performance score rose from 2.95 (SD = 0.46) prior to the intervention to 4.07 (SD = 0.47) after the intervention, with an average gain of +1.12 points on the five-point rubric.

The paired-samples to test indicated a statistically significant improvement, t (29) = 19.17, p < .001, 95% CI [1.00, 1.24], Cohen's dz = 3.50. Criterion-level patterns may also be explored descriptively for conceptual understanding, accuracy, explanation depth, organization, academic language, research justification, and response confidence; however, the prespecified total or mean rubric score should remain the primary performance outcome.

APA-style example: Oral performance increased significantly from pre-intervention (M = 2.95, SD = 0.46) to post-intervention (M = 4.07, SD = 0.47), t (29) = 19.17, p < .001, 95% CI [1.00, 1.24], dz = 3.50.

Perceived Relevance of Generated Questions

In the qualitative component, several reflections described the questions as closely connected to research objectives, methodological choices, findings, theoretical assumptions, and limitations. These examples illustrate how perceived relevance can be analysed qualitatively. In the real study, researchers should also consider adding a short relevance scale and reporting the number and percentage of students who rate the questions as relevant or highly relevant.

Qualitative Themes

Greater Preparedness

The synthetic reflections suggested greater preparedness. Example statements described exposure to previously unanticipated questions and increased attention to the justification of research choices. This theme may be coded from comments concerning readiness, identification of weak sections, or broader anticipation of examiner questions.

Reduced Fear of Questioning

A second theme was reduced fear of questioning. Example reflections described the low-pressure environment as helpful for becoming familiar with difficult questions and for reducing fear of an imperfect first answer.

Improved Response Structure

A third theme was improved response structure. Example comments indicated that students learned to start with a direct answer, provide concise justification, connect the response to their own study, and avoid unnecessary detail. A fourth theme, human oversight and limitations, reflected the need to verify important feedback with supervisors and the limitations of text-based rather than voice-based practice.

Discussion

The analyses illustrate the pattern that would support the educational usefulness of retrieval-augmented generative AI: lower anxiety, higher confidence, and stronger rubric-based oral performance after structured document-grounded rehearsal. Because the data are synthetic, the numerical results cannot establish effectiveness; rather, they demonstrate how actual findings could be analysed and reported (Grubert et al., 2025).

One factor that could contribute to this decrease in anxiety is the decrease in uncertainty. Pupils often worry about questions they haven't thought of or gaps in their knowledge being identified. Predictability was not eliminated in the system but rather made students familiar with several question types relating to their own documents. This could have enabled an unknown assessive environment to become more familiar and manageable (Ansari et al., 2025).

This increase in confidence is in line with the self-efficacy theory. Mastery experiences may have occurred during multiple opportunities to answer questions, give feedback and refine responses. Students had the opportunity to monitor their own work without having to face a formal examination panel (Huang, 2026).

The results also point to the importance of formative feedback. The system did not just check to see if an answer was right or wrong, it also gave feedback on clarity, depth, organization and relevance. This can prompt students to consider how well they express what they know, and what they know (Yusuf et al., 2025)

One aspect of the intervention that is relevant is the employment of retrieval augmentation. General generative AI systems may generate general and formulaic viva questions. The system was able to extract data from uploaded documents to create questions that were more related to specific research projects. This relevance probably heightened students' engagement and perceived authenticity (Lee & Eronen, 2025).

If similar patterns are noted in the real data set, results should be viewed with caution. Repeated exposure to the research document, more study time, becoming familiar with the rubric, regression to the mean, or practice effects can all lead to improvement. Without any comparison group, the changes cannot be solely attributed to the AI system (Smith, 2025).

Like automated evaluation, automated testing has its limitations. The feedback from a language model can be convincing but wrong, or it may not identify legitimate disciplinary alternatives, or it may be more inclined toward academic forms of expression. The use of AI-generated model answers can result in students relying more on rote-learning than building an understanding (Puleio et al., 2024).

Therefore, the use of AI for rehearsals should be used in conjunction with rather than instead of human supervision. These are tasks that automated systems can't reliably do, and supervisors can do: Disciplinary judgment, challenge of assumption, originality assessment, ethical concerns and nonverbal communication evaluation.

Limitations

There were a number of limitations to the study. The study sample was small and limited to one public sector university. The results may therefore not reflect students in other disciplines, at other schools, in other languages or at other educational systems. There was no control or comparison group for the one group pre–post design. Changes could be due to multiple assessments, preparation by others, or familiarity with the assessment procedures. Self-reported outcomes were used for some outcomes. The positive responses may have been due to the fact that the intervention was novel or that a positive response was anticipated.

Mostly text-based responses were assessed. It did not provide a thorough analysis of pronunciation, voice quality, pacing, eye contact, posture, gestures with a focus on confidence, or spontaneous communication with a human panel. Automated feedback may be valid and reliable only for certain types of documents, retrieval, prompt construction, model updates and disciplinary complexity. Lastly, all the values have to be substituted with real data statistics. The final article should include sample sizes, missing data, means, standard deviations, reliability coefficients, test statistics, degrees of freedom, exact p values (if applicable), 95% confidence intervals, effect sizes, and transparent qualitative procedures. Empirical articles simply cannot be based on percentage improvements.

Future Research

Future studies should use more and varied samples and controls/comparisons. Randomized or quasi-experimental study design might include comparing retrieval-augmented AI practice to traditional preparation, peer rehearsal, supervisor-led mock vivas and general use generative AI.

Further research is needed on long-term confidence and performance gains, discipline differences, inter-subject differences, effectiveness for undergraduates and post-graduates, multilingual viva preparation, automated scoring accuracy and fairness, supervisors' perception of AI generated questions, student reliance on AI-generated model answers, local deployment of AI for privacy reasons, verbal and nonverbal performance indicators, and speech-based questioning and response analysis.

Future versions will cover score explanations, discipline specific rubrics, voice interaction, follow up questions, adjustable examiner difficulty, and support for multiple languages.

Conclusion

This study tested the use of retrieval-augmented generative AI as a preparatory tool for academic vivas and related oral examinations. The system retrieved the information from students' uploaded academic documents, created examiner style questions, evaluated the responses and gave students immediate formative feedback.

The synthetic findings demonstrated that, after structured practice, the students' anxiety about presentation, their academic confidence, and their oral competence measured by the rubric could be explored. The accompanying examples of qualitative information showed how they can be combined with quantitative results in the context of document relevance, preparedness, response organization, and human oversight. These are methodological examples and should be substantiated with empirical data of participants.

Retrieval-augmented generative AI thus holds promise for accessible, personalized, and pressure-free practice for high-stakes, academic oral evaluations. It is the repeated questioning, reflection, and formative feedback that is educational, and it is the link between document-specific questioning and repeated practice, reflection, and formative feedback that has educational significance.

However, an AI-powered preparation should not be the only component of a larger supervisory and institutional process. The contributions of human academic judgment, authentic interpersonal practice, ethical oversight and disciplinary expertise are still vital. A good use of such systems is as an additional formative device which complements and supports the student-supervisor relationship and formal assessment.

Conflict of Interest

The authors showed no conflict of interest.

Funding

The authors did not mention any funding for this research.

References

Ansari, G. J., Wasti, S. H., & Sarwar, M. N. (2025). MumtahinGPT: A preparatory tool for reducing anxiety and enhancing students' academic performance with retrieval-augmented generation and AI feedback. *Social Sciences Spectrum*, *4*(2), 906–918.

Beale, R. (2025). The revolution has arrived: What the current state of large language models in education implies for the future. *ArXiv Preprint ArXiv:2507.02180*.

Blessing, E. (2026). Artificial intelligence in medical education: A systematic review of teaching, learning, and academic integrity issues. *International Journal of Applied Resilience and Sustainability*, *2*(3), 16–48.

Grubert, J., Sedelmaier, Y., & Landes, D. (2025). Towards embodied conversational agents for reducing oral exam anxiety in extended reality. *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 561–564.

Huang, J. W.-T. (2026). *AI as a debate coach: A mixed-methods analysis of student self-efficacy and perceptions in an AI-assisted debate*.

Lee, S., & Eronen, J. (2025). Transforming English education in Japan by utilizing large language models for tailored learning and diverse skill development. *Discover Education*, *4*(1), 403.

Németh, R., Tátrai, A., Szabó, M., Zaletnyik, P. T., & Tamási, Á. (2025). Exploring the use of retrieval-augmented generation models in higher education: A pilot study on artificial intelligence-based tutoring. *Social Sciences & Humanities Open*, *12*, 101751.

Perez Verástegui, J. F., Ortega Rojas, Y. K., Casazola Cruz, O., Villar, L. P., Castro Chiroque, R., & Morote, S. (n.d.). *Guide to the use of generative artificial intelligence in education and research*.

Philippakos, Z., Katsarou, E., & Sipitanos, K. (2026). *Artificial Intelligence in Literacy Education: Foundations, Practices and Innovations*. Taylor & Francis.

Pramanik, S., & Husin, M. H. Bin. (2025). The Retrieval-Augmented Pedagogical Assistant (RAPA): A Methodology for Enhancing Critical Thinking and Equity in AI-Augmented Education. *International Journal of Advanced Computer Science and Applications*, *16*(12), 274–284.

Puleio, F., Lo Giudice, G., Bellocchio, A. M., Boschetti, C. E., & Lo Giudice, R. (2024). Clinical, research, and educational applications of ChatGPT in dentistry: a narrative review. *Applied Sciences*, *14*(23), 10802.

Smith, S. (2025). Enhancing English language education with retrieval-augmented generation (RAG): A new paradigm for AI-powered teaching and learning. *International Journal of Education Humanities and Social Science*, *8*(4), 235–252.

Voultsiou, E., & Moussiades, L. (2026). A Systematic Review of Large Language Models in Mental Health: Opportunities, Challenges, and Future Directions. *Electronics*, *15*(3), 524.

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Towards reducing teacher burden in Performance-Based assessments using aivaluate: an emotionally intelligent LLM-Augmented pedagogical AI conversational agent. *Education and Information Technologies*, *30*(17), 24649–24693.

Zapata, G. C. (2025). *Generative AI technologies, multiliteracies, and language education*. Taylor & Francis.