

Transforming Pedagogy with AI Tools: Impacts, Challenges, and Opportunities in Higher Education

Chi Ho CHAN ¹, Christy CHAN ², Julia CHEN ³, Pauli LAI ¹

¹ Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China

² Chan Feng Men-ling Chan Shuk-lin Language Centre, City University of Hong Kong, Hong Kong, China

³ Educational Development Centre, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Traditional pedagogy positions teachers as the primary source of knowledge, with educators providing academic guidance on English oral presentations—including both content and paralinguistic features—across university disciplines. However, subject-matter instructors often lack expertise in English language delivery, while large class sizes prevent the provision of timely individual feedback. Advances in technology enable a pedagogical shift to fill this gap. Rather than waiting for teacher feedback, students can leverage AI for immediate learning support. However, AI has limitations regarding nuanced aspects of performance. To complement AI feedback, we implemented **NinjOrAItor**—an AI platform for English oral presentations featuring hybrid intelligence that combines self- and peer-assessment. This hybrid intelligence pedagogy integrates AI capabilities with human insight. The platform addresses key challenges: teacher expertise limitations, scalability issues in large classes, and the need for timely feedback. By integrating AI assessment with structured human intelligence, *NinjOrAItor* demonstrated improved student presentation performance while maintaining pedagogical quality.

Keywords: AI, English oral presentations, hybrid intelligence, pedagogy, self-assessment, peer assessment

1. Introduction

Empirical evidence demonstrates that technological tools consistently enhance student achievement, engagement, and teaching effectiveness across educational levels and disciplines (Hu, 2024). In English oral presentation instruction, where students traditionally rely on delayed teacher feedback, technology offers timely and comprehensive assessment that can accelerate skill development and promote interactive learning experiences.

English serves as the *lingua franca* in tertiary education (Alhasnawi, 2021). Universities now require English oral presentations across disciplines, accompanied by teacher assessment and feedback. However, this traditional approach faces significant learning barriers: subject-matter instructors often lack specialised English language expertise to deliver targeted linguistic feedback, while large classes limit timely individual assessment, which is crucial for language skill development. Although AI platforms can provide faster feedback, current AI systems still struggle to evaluate nuanced presentation qualities, leaving a gap between technological potential and pedagogical needs.

To bridge this gap and optimise learning outcomes, we developed **NinjOrAitor**, an AI platform that employs hybrid intelligence—combining automated feedback with human input and structured learning resources. *NinjOrAitor* is grounded in three pedagogical principles essential for the effective development of English oral presentation skills: active learning, self-regulated learning (SRL), and human-AI hybridity. Details are as follows:

1. **Active Learning:** Emphasises student-centred, iterative practice and engagement over passive lecture-based instruction, enabling deeper skill acquisition through repeated performance, feedback, and refinement (Kozanitis, 2023).
2. **Self-Regulated Learning (SRL):** When properly designed, SRL consistently enhances students' learning outcomes and academic performance, as evidenced by multiple meta-analyses (Panadero, 2017).
3. **Human-AI Hybridity:** Leverages the complementary strengths of AI and human expertise. AI provides scalable, immediate analysis of technical and perceptual elements (e.g., pronunciation, fillers, and delivery), while humans contribute contextual judgment and nuanced decision-making (Holstein et al., 2020).

Together, these principles demand frequent practice cycles supported by timely, multi-layered feedback—conditions that traditional large-classroom instruction struggles to meet. *NinjOrAitor* operationalises them by integrating technology-enhanced active practice, AI-driven immediate feedback with self- and peer-assessment tools, and structured human oversight, creating a scalable environment that accelerates skill development while preserving pedagogical depth.

NinjOrAitor integrates Azure OpenAI GPT-4o and Azure OpenAI GPT-4 to deliver automated rubric-based feedback on eight dimensions: clarity of view, content, eye contact, facial expression, hand gestures, PowerPoint visu-

als, pronunciation, and vocal fillers. The system processes approximately 40 five-minute video submissions per hour to enable rapid feedback cycles while teachers retain final oversight for contextual judgment. Students access the platform through institutional email accounts. They upload presentation videos, receiving both AI and human feedback in the process.

The hybrid feedback loop works as shown in **Figure 1**. Students first receive immediate AI-generated comments. They then complete self-assessment by comparing their own reflections against the AI feedback, followed by peer assessment on linguistic and paralinguistic aspects, as indicated in **Figures 2, 3, and 4**.

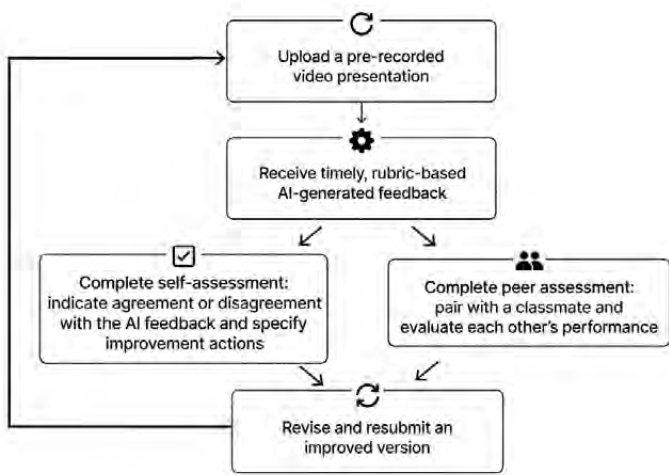


Figure 1: Hybrid feedback loop in *NinjOrAltor*

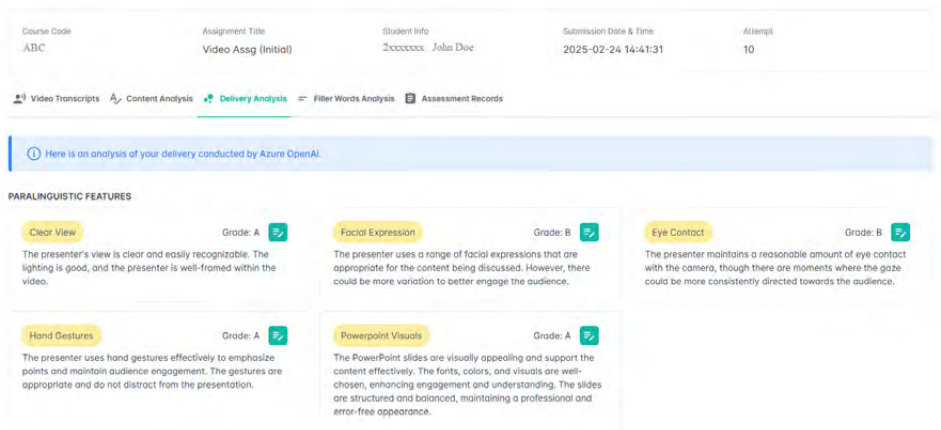


Figure 2: AI feedback interface

Self Assessment

Peer Assessment

Self Assessment Submitted on: 2025-03-04 16:41:33

Q1: Which AI feedback do you AGREE with? Please copy and paste the sentence(s) DIRECTLY.

Additionally, while the presentation covers the advantages of NoSQL databases, it lacks a critical analysis of potential limitations or challenges associated with their use in ride-sharing applications. This would have provided a more balanced view of the topic.

Q2: How would you implement improvements based on the AI feedback that you agree with?

I will add the potential challenges and limitations of NoSQL in application in my presentation and propose corresponding solutions.

Q3: Which specific AI feedback do you DISAGREE with? Please copy and paste them DIRECTLY.

While you have effectively explained the role of NoSQL in Uber's operations, including more detailed examples or case studies could further enrich the presentation. For instance, discussing a specific instance where NoSQL significantly improved Uber's service could make your argument more impactful.

Q4: Why do you disagree with the above-mentioned AI feedback, if any?

I have already provided some examples to illustrate during the explanation process, but due to time constraints, I am unable to add more in this regard.

Q5: What are the feedback that you think AI did not mention?

I don't think AI mentioned whether my presentation and slides are compatible with each other.

Q6: How would you improve the performance?

I will reorganize my content based on the suggestions and feedback provided by AI, add missing parts, and improve my eye contact.

Figure 3: Self-assessment interface

Self Assessment

Peer Assessment

Peer Assessment Submitted by: John Doe (2xxxxxxx d) | Submission on: 2025-03-04 16:31:45

Peer Feedback on Eye Contact

Your eye contact was impressive throughout your presentation. You effectively engaged with the audience by looking at them, which made them feel included and appreciated. Your eye movements were smooth and natural, successfully drawing the audience's attention and fostering a connection with your message.

Peer Feedback on Facial Expression

Your facial expressions were complete and clear, effectively conveying the emotions and themes of your presentation. You displayed a range of expressions that matched your message perfectly, whether it was a warm smile, a look of surprise, or a serious demeanor. This clarity in your facial expressions not only enhanced your communication but also made your presentation more engaging and relatable, allowing the audience to connect with your points on a deeper level.

Peer Feedback on Hand Gesture

Your hand gestures were very effective and added to your communication without being distracting. You used gestures to highlight important points, making your information more engaging and easier to follow. The gestures felt natural and confident, which further enhanced the overall quality of your presentation.

Peer Feedback on Pronunciation

Your pronunciation was clear and accurate, allowing the audience to easily understand every word you spoke. Your effort in pronunciation contributed to a professional delivery, enhancing the effectiveness of your message. Whether using complex terminology or simple sentences, your clarity made the presentation comfortable for the audience to follow.

Peer Feedback on Vocal Filler

You managed the use of vocal fillers exceptionally well, with minimal distractions throughout your speech. Your delivery was smooth, and you almost entirely avoided using fillers like "um" or "uh," which made your presentation more coherent and professional. This performance not only showcased your familiarity with the content but also helped the audience focus on your key points.

Figure 4: Peer-assessment interface

By prompting students to critique AI-generated feedback through structured self- and peer-assessment, *NinjOrAI*tor cultivates *evaluative judgement*—the capacity to make informed decisions about the quality of one’s own and others’ work (Tai et al., 2018). This active process aligns with constructivist principles in language pedagogy, as students reconcile automated outputs with their own reflections and peer perspectives instead of receiving feedback passively. Such engagement fosters deeper metacognition and autonomy compared with the passive consumption of AI feedback.

2. Methodology

The study adopted a mixed-methods design (Creswell & Plano Clark, 2018) to evaluate *NinjOrAI*tor’s impact on pedagogical practice and student learning in English oral presentations. Our dataset came from a database subject involving 171 students; over 98% of whom use English as an Additional Language (EAL). They were required to submit presentation videos through *NinjOrAI*tor.

Quantitatively, pre- and post-surveys were administered to measure students' understanding of rubric criteria, confidence, and the use of rubrics for self-learning. Responses were coded on a 5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). These were analysed using weighted means ($\Sigma [\text{response value} \times \text{frequency}] / \text{total responses}$) to capture both the direction and magnitude of change. The survey instrument employed the same eight rubric components as those used in *NinjOrAItor*, with no additional differential weighting applied beyond the 5-point Likert scale itself. A total of 104 students completed the pre-survey, while 59 filled in the post-survey. Using this approach, the study observed improvements in students' understanding of rubric criteria and their use of rubrics for self-learning after interacting with *NinjOrAItor*.

Qualitatively, semi-structured interviews with students examined user experience, perceived specificity and actionability of AI feedback, the value of learning units, and the complementarity of self/peer assessment. Survey, performance, and interview data were collected after students received immediate, rubric-aligned AI feedback and completed self/peer assessments embedded in course workflows.

3. Results and Discussion

The video marks were calculated based on eight rubric criteria with the following weightings: Content (35%), Clarity of View, Eye Contact, Facial Expression, Pronunciation, Vocal Fillers, and PowerPoint Visuals (10% each), and Hand Gestures (5%).

Students showed improved performance in English oral presentations after using *NinjOrAItor*. Among the 161 students who submitted both video assignments, average scores rose by 4.80% from $M = 81.07$ ($SD = 4.41$) on the first assignment to $M = 84.96$ ($SD = 3.74$) on the second. A paired-samples t-test confirmed the improvement was statistically significant, $t(160) = -13.01$, $p < 0.001$ (two-tailed).

Students received hybrid feedback from *NinjOrAItor* on their first submission and then revised and resubmitted an improved version as the second assignment. At the end of the semester, post-survey results (compared with pre-survey baseline) showed improved attitudes across most assessment criteria. The largest gains occurred in students' perceptions of vocal fillers (weighted mean increased by 16.67% from a pre-survey baseline of 3.33 to a post-survey value of 3.88), followed by paralinguistic techniques (weighted mean increased by 7.17% from a pre-survey baseline of 3.61 to a post-survey value of 3.86).

However, attitudes toward pronunciation assessment showed a modest decline (with the weighted mean decreased by 0.61% from a pre-survey baseline of 3.62 to a post-survey value of 3.59).

Students also reported improved understanding of rubric criteria and an enhanced ability to apply rubrics for self-directed learning. Perceptions of *NinjOrAitor* were mixed: while appreciation for the platform’s feedback features increased (weighted mean up by 15.38%), confidence in its assessment accuracy decreased (weighted mean down by 8.60%), primarily attributed to technical limitations in hand gesture detection. Meanwhile, in the post-survey, 71.19% of students indicated that the platform was helpful for self-improvement.

Interviews with five consenting students revealed generally positive perceptions of *NinjOrAitor*, with participants describing it as intuitive and user-friendly. Students reported effective use of rubrics in preparing their presentations and consistent improvements in vocal filler reduction, pronunciation, and speech pacing after receiving AI feedback. The hybrid intelligence approach provided comprehensive guidance that helped students identify specific areas for improvement and enhance confidence in presentation preparation and delivery. One student described a complementary peer-AI loop: peers use AI’s rubric-based overview (e.g., on eye contact) as a foundation but added contextual depth, such as content simplification for non-experts, creating a collaborative revision process.

Despite these benefits, AI limitations in visual analysis (hand gestures and slide content) and occasional content feedback inaccuracies were noted as areas needing refinement. Overall, students appreciated the objective baseline and structured support, particularly for inexperienced presenters, with rubrics and Learning Units facilitating targeted skill development. Instructors valued *NinjOrAitor*’s vocal filler detection for identifying expressions like “so,” “like,” and “uh.” However, occasional misclassifications (e.g., flagging “but” as a filler) require human oversight. The platform enhanced grading efficiency while maintaining assessment quality by providing AI evaluations as an objective starting point. Educators retained full authority to adjust scores and feedback for specific pedagogical contexts, creating an optimal human-AI partnership that leverages automation while preserving personalised instruction.

This study has several limitations. Although 171 students were enrolled, only 104 respondents (60.82%) completed the pre-survey and 59 respondents (34.50%) completed the post-survey, resulting in a 43.27% dropout rate. This substantial dropout may introduce survivorship bias, as more motivated or en-

gaged students were more likely to persist with both surveys and video assignments. Consequently, the observed improvements in presentation scores and perceptions may be positively skewed.

4. Implications

The hybrid model implemented in *NinjOrAItor* resolves the core weaknesses identified in the literature and observed in our data (see Table 1). Students found the platform effective for independent presentation skill development, yet the slight decline in confidence in AI evaluation highlights the continued importance of human judgment. Through integrated self- and peer-assessment features, students learn to critically evaluate AI feedback while learning collaboratively from peers. These findings align with the theoretical advantages of hybrid intelligence outlined earlier.

Table 1 compares the four main feedback models currently used in English oral presentation instruction. It summarises the performance outcomes of four primary feedback models: instructor-only, automated (AI-only), peer/self, and the hybrid model. Since performance gains for the AI-only and teacher-only models could not be quantified in this study, they are described qualitatively as ‘moderate’ based on prior literature and observed classroom practices (Kaliisa et al., 2025).

Table 1: Comparative Analysis of Feedback Models in English Oral Presentation Instruction

Feedback Model	Timeliness of feedback	Scalability	Nuance in assessment	Performance gains observed	Data Source
Instructor-only	No	No	Yes	Moderate	Kaliisa et al. (2025); classroom observations
AI-only	Yes	Yes	No	Moderate	Kaliisa et al. (2025)
Peer/Self only	Partially	Partially	Partially	Moderate	Kaliisa et al. (2025)
Hybrid (NinjOrAItor)	Yes (AI + selective human)	Yes (AI + selective human)	Yes (AI + human contextual judgment)	High (+4.80% overall score, p <0.001)	Present study (empirical, paired t-test)

Note: Non-NinjOrAItor rows draw on Kaliisa et al. (2025); NinjOrAItor results are from the present study.

The hybrid model implemented in *NinjOrAItor* (rightmost column) resolves the core weaknesses identified in the literature and observed in our data. While AI-only feedback delivers the timeliness and scalability demanded by large EAL classes (171 students, >98% EAL), it cannot provide the nuanced judgment on content relevance, cultural pragmatics, or motivational support

that human instructors supply. By embedding immediate AI analysis within a human-feedback loop, students not only improved presentation scores by 4.80% ($t(160) = -13.01, p < 0.001$) but also developed metacognitive skills through actively evaluating AI output.

Acknowledgments

This research was supported by the TDG Inter-institutional Collaborative Activities (IICA) 2022–25 grant through a partnership between The Hong Kong Polytechnic University (PolyU) and City University of Hong Kong (CityUHK). The authors extend their gratitude to the English instructors at CityUHK and the technical development team at PolyU, whose collaborative efforts have been instrumental in advancing this study on hybrid intelligence in English language education.

Ethical Declaration Statement

The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) (or its Delegate) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20231206007).

Declaration of Use of AI in the Writing Process

The authors used Grammarly during the preparation of this work to check grammar and the use of English. After utilising the tool/service, the author(s) thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication's content.

References

- Alhasnawi, S. (2021). English as an academic lingua franca: Discourse hybridity and meaning multiplicity in an international Anglophone HE institution. *Journal of English as a Lingua Franca*, 10(1), 31–58. <https://doi.org/10.1515/jelf-2021-2054>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education (AIED 2020)* (Lecture Notes in Computer Science, Vol. 12163, pp. 240–254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20
- Hu, S. (2024). The effect of artificial intelligence-assisted personalised learning on student learning outcomes: A meta-analysis based on 31 empirical research papers. *Science Insights Education Frontiers*, 24(1), 3873–3894.

- Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2025). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, 46(1), 80–111. <https://doi.org/10.1080/01443410.2025.2553639>
- Kozanitis, A., & Nenciovici, L. (2023). Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: A meta-analysis. *Higher Education*, 86(6), 1377–1394. <https://doi.org/10.1007/s10734-022-00977-8>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>