Journal of English Language Teaching and Learning (JELTL)
Vol. 7, No. 1, June 2026, 27-38
E-ISSN: 2723-617X
Available online at: https://publikasi.teknokrat.ac.id/index.php/jeltl

COMPARING AI AND ENGLISH EDUCATION STUDENT PEER FEEDBACK QUALITY IN ASSESSING ENGLISH PRESENTATION SKILLS

Dilah Yuda Rizkia[1], Fela Dafitri[2]
Universitas Singaperbangsa Karawang[1,2]

2210631060073@student.unsika.ac.id[1]
Fela.dafitri@fkip.unsika.ac.id[2]

Received: (31 May 2026) **Accepted**: (8 June 2026) **Published :** (27 June 2026)

Abstract

This study investigated the differences between peer feedback generated by preservice teachers and an artificial intelligence model that simulated the same perspective on assessing student presentation performance in an English education context. Research comparing artificial intelligence and preservice teacher-generated peer feedback in assessing presentation performance rather than writing performance was rare, especially in the Indonesian higher education context. This study aimed to measure and compare the quality, strengths, and limitations of feedback provided by seventh-semester English Education department students versus artificial intelligence-generated feedback. This study used a qualitative descriptive method to analyze 10 peer feedback samples collected from 5 students and 5 artificial intelligence models. A standard presentation review rubric served as a form for both feedback types to review the presentation skills. The researcher used an adapted rubric to compare the two feedback types. The findings showed that the artificial intelligence-generated feedback was more elaborate and detailed but lacked contextual understanding. Human feedback offered contextual understanding and comments shaped by shared classroom experiences but tended to be shorter and surface-level. The study concluded that artificial intelligence-generated feedback was more precise and strong in organizational quality, while human feedback was more authentic and offered interpersonal support. Both feedback types can be used as pedagogical tools. They help students understand how effective feedback should be structured. Teachers can also consider using artificial intelligence as a supplementary tool for grading presentation skills.

Keywords: Artificial intelligence, English education, Feedback quality, Peer feedback, Presentation skills.

To cite this article:
Rizkia, D.Y. & Dafitri, F. (2026). Comparing AI and English education student peer feedback quality in assessing English presentation skills. *Journal of English Language Teaching and Learning*, *7*(1), 27-38.

INTRODUCTION

In English language education, assessing presentation skills is a very important part for future teachers who need to develop both their language skills and teaching ability. Assessing presentation skills helps in developing students' verbal and non-verbal communication abilities. This includes mastery of language competencies and understanding of linguistic registers (Rusu & Bozov, 2024). Feedback is defined as the transmission of evaluative or corrective information about an action or process to the original source. In education, it specifically refers to information provided to students about their performance in relation to learning objectives, aiming to improve student learning and achievement (Al-Ogaili, 2023). Effective feedback is characterized by its concreteness, timeliness, positivity, constructive nature, and dialogicity (Butenko, 2024). As the number of students in English education departments grows, the traditional way of giving detailed and personal feedback for presentations is facing many challenges. These challenges come from teachers having limited time and the need for consistent grading standards.

This problem is very noticeable for seventh-semester English education students. Pre-service teachers face challenges such as classroom management and lesson preparation while simultaneously developing strategies to overcome these issues (Alif et al., 2025). They are a unique group who already have a lot of academic and teaching practice experience. This align with the study of Ficayuma (2022) who identified that a significant portion of English education students scored within the B and C levels on the CEFR scale, suggesting they are competent but not yet at the highest proficiency levels. The teaching practice program for English education students includes a focus on developing pedagogical skills, with students demonstrating good teaching abilities in pre- lesson activities and moderate skills in closing lessons, indicating their growth as educators. (Kasriyati et al., 2022). The practice of peer feedback is generally defined as a process where learners of equal status provide feedback to each other about their performance. In higher education, peer feedback offers several benefits, including promoting deep self-reflection, creating a supportive learning environment, and providing opportunities for peer support through interaction (Lam & Habil, 2020). However, research specific to pre-service teachers' feedback provision skills reveals unique challenges. Rahman et al. (2025) identified that pre-service teachers' feedback skills are influenced by internal factors including pedagogical understanding and teaching experience. Their capacity to provide differentiated and appropriate feedback may still be developing during their final semesters of preparation.

The advancement of technology plays a crucial role in the educational development of these students. These technological developments, especially in artificial intelligence and large language models, have provided new opportunities to create automatic feedback in education. The integration of artificial intelligence into language education has accelerated dramatically. Studies have demonstrated that AI tools can significantly enhance student confidence and engagement in presentations. For instance, AI-assisted tools have been shown to improve presentation quality and boost self-efficacy among students, transforming initial skepticism into acceptance and endorsement of AI technologies (Askari & Rahim, 2024). AI platforms like SmartPresenter have been developed to provide instant feedback on various presentation aspects such as eye contact, facial expressions, and vocal delivery. These platforms are particularly beneficial for students learning English as an additional language, offering tailored feedback that supports self-paced learning and reduces the need for constant teacher intervention (Chen et al., 2022; Lai et al., 2023). Dafitri et al. (2025) highlight this potential by stating that generative artificial intelligence serves as a collaborative medium for developing authentic and adaptive assessments. They argue that when educators combine their pedagogical understanding with specific prompting parameters, they can generate highly personalized materials that meet specific contextual needs. This collaborative approach ensures that the technology supports the human element in educational design. A study involving Turkish high school students found that those who received AI-assisted feedback showed significantly greater improvement in writing performance compared to those who received traditional teacher feedback. The technology operates as an educational companion or personal tutor, delivering tailored and adaptable responses. A study involving Turkish high school students found that those who received artificial intelligence-assisted feedback showed significantly greater improvement in writing performance compared to those who received traditional teacher feedback (Ekizoğlu & Demir, 2025). AI platforms have been developed to provide instant feedback on various presentation aspects such as eye contact, facial expressions, and vocal delivery (Chen et al., 2022).

Despite these promising capabilities, research has identified critical limitations in artificial intelligence feedback. Mostafapour et al. (2024) found that while artificial intelligence demonstrated impressive breadth of knowledge, it exhibited limitations in depth of contextual understanding. Furthermore, Chen et al. (2025) revealed a complex relationship between artificial intelligence feedback and learning experiences, finding that technical capabilities alone do not guarantee effective learning outcomes because affective dimensions require careful consideration. The effectiveness of the technology also depends critically on prompt engineering. Qian (2025) identified that technique-based strategies targeting specific learning goals are key prompting frameworks for generating quality responses.

These studies have clearly shown the abilities and limitations of artificial intelligence feedback. However, several important gaps still exist in the current literature. The first gap is the lack of comparisons between artificial intelligence and human feedback for presentation assessment in higher education, specifically for future teachers. The second gap is that most feedback research has focused on written assessment with little investigation into oral presentation skills. The third gap is the limited research conducted in Indonesian English education department settings. Therefore, this study aims to fill these research gaps through a systematic analysis of artificial intelligence feedback versus human peer feedback in presentation skills assessment. The main objective is to measure and compare the quality of feedback provided by seventh-semester English Education department students versus artificial intelligence feedback specifically prompted to simulate the perspective of an English education student.

Based on the problems, this research raised two questions to address. The first question explores the differences in the quality of feedback provided by seventh-semester English Education students compared to artificial intelligence feedback designed to simulate the same perspective. The second question explores the strengths and limitations of each peer feedback type. The findings from this research could fill theoretical gaps by examining the intersection of technology capabilities and human teaching expertise. Practically, the findings could inform curriculum development and offer evidence-based recommendations for integrating feedback systems into language education programs.

RESEARCH METHOD

This research used a qualitative descriptive design to explore and compare the characteristics of feedback produced by artificial intelligence and human peers in the context of presentation assessment. The qualitative descriptive approach was selected because the study aimed to provide a detailed and factual description of how each feedback type demonstrates quality rather than to test hypotheses or measure numerical differences. This makes it ideal for capturing the reality of a situation from the participants' perspectives (Tamayo et al., 2020). This study focused on presenting an in-depth account of patterns, tendencies, and contrasts between the two feedback sources.

This research was conducted at Singaperbangsa State University, specifically within the English Education Department. The study focused on seventh-semester students who were undertaking a Research Proposal Seminar course requiring the oral presentation of research proposals. The participants were selected based on specific criteria. They had to be currently enrolled as seventh-semester students in the program. They also had to have completed approximately 75 percent of their degree requirements, including teaching practicum courses. A total of five students participated as peer reviewers. The research proposal presentation videos were obtained from five different students in the same semester whom the peer reviewers evaluated. Each participant provided informed consent for their presentation videos to be used in this research.

The artificial intelligence feedback system used the Gemini 2.5 Pro model with a carefully engineered prompt. This prompt was designed to simulate the perspective, knowledge base, and feedback provision capabilities of a seventh-semester English Education department student. The prompt instructed the model to act as a peer feedback review specialist and embody the persona of a student with prior teaching experience. The model was instructed to analyse the presentation video details and deliver high-quality, constructive, and actionable feedback based on established pedagogical practices. The prompt required the model to structure its feedback clearly, maintain a professional and supportive tone, and pair every critique with a suggestion for improvement.

This research used two main instruments for data collection and analysis. The first instrument was the standardized presentation review rubric developed by Paltridge & Starfield (2020). Both the human feedback and the artificial intelligence model used this rubric to score the presenters based on clarity of presentation, introduction quality, research questions, literature review, theoretical framework, and research methodology. The second instrument was a Qualitative Descriptive Rubric for Feedback Quality adapted from Hansen (2025). This adapted rubric focused on seven main categories, including content relevance, use of explanations or examples, specificity, questioning, advice or suggestions, style, and formulation or encouragement.

The data collection was carried out by compiling written feedback samples from both human and artificial intelligence sources. Human feedback was obtained from students' authentic peer assessment forms. The artificial intelligence feedback was generated using identical presentation details and assessment criteria. Each piece of feedback was anonymized and labelled according to its source to maintain systematic organization. The collected feedback was analysed using qualitative content analysis guided by the adapted Feedback Quality Rubric. Each feedback comment was read carefully and coded under one or more relevant categories. Patterns were then identified within and across the two feedback sources to reveal similarities, contrasts, and distinctive tendencies.

FINDINGS

This study analyzed ten written feedback samples, five from an artificial intelligence system and five from human peers. Using the qualitative descriptive framework adapted from Hansen (2025), the findings reveal consistent contrasts between the two sources of feedback in terms of depth, tone, and structure. Artificial intelligence feedback demonstrated higher levels of elaboration, specificity, and constructiveness. The findings are categorized into two groups according to the research questions that guide this research. Human feedback tended to focus on surface-level observations and brief evaluative comments.

Differences in Feedback Quality

The findings of this study reveal clear and distinct differences between the artificial intelligence outputs and the human peer feedback in terms of explanatory depth, structural organization, and pedagogical function. The artificial intelligence demonstrated a higher capacity for elaboration. It systematically linked observed presentation behaviors to broader pedagogical implications. The system justified its evaluations by explaining the complex reasoning behind its critiques. For example, the system provided comments such as "reading from notes reduced eye contact," which clearly establishes a cause-and-effect relationship for the presenter. This finding suggests that large language models possess an inherent ability to model structured and reflective feedback discourse. In contrast, the human peer feedback reflected a highly observational approach with minimal elaboration. The seventh-semester students tended to identify noticeable surface issues without connecting them to deeper communication impacts. Comments such as "mispronounced words" or "relied on reading" correctly identify areas of concern but do not explain why these behaviors reduced presentation quality. This pattern showed that the artificial intelligence feedback was more explanatory, with the depth of its reasoning reflecting a high degree of simulated feedback literacy.

Another significant difference is found in the specificity and constructiveness of the evaluations. The artificial intelligence system provided specific references to presentation slides and offered clear, actionable suggestions for improvement to function as a highly formative assessment tool. The system included exact slide numbers, precise descriptors such as "dense paragraphs," and direct references such as "Slide 6 clearly defines...". This precision made the comments immediately actionable for the presenter. The artificial intelligence also advised the presenter to "Use bullet points instead of full sentences on slides". Human feedback, on the other hand, was more limited to observable surface details such as voice volume, general note reliance, or basic pronunciation. While these comments remained relevant to overall performance, they did not include follow-up reasoning. The artificial intelligence also provided practical suggestions, such as advising the presenter to "use bullet points instead of full sentences on slides," while human comments typically concluded after identifying an issue without proposing a solution.

In terms of delivery style, the artificial intelligence consistently used a direct and professional tone. The system addressed the presenters in the second person using phrases like "You might..." or "I suggest..." to maintain a professional but supportive demeanor. It also included reflective questions such as "How will you assess performance qualitatively?" to encourage the presenter to engage in deeper critical thinking about their own research proposal. Human feedback rarely utilized such reflective questions. The pre-service teachers tended to employ distant third-person phrasing, such as "She was good" or "He relied on notes". This phrasing created a more distant formal relationship between the reviewer and the recipient on paper. However, the human peers often conveyed a strong sense of sincerity and authenticity in their praise, particularly at the very beginning of their evaluation forms. This indicates that while artificial intelligence excelled in maintaining a balanced and pedagogically appropriate academic tone, the human peers contributed a highly personal and context-sensitive voice that technology cannot easily replicate.

This difference means that the artificial intelligence feedback was highly explanatory. The depth of the machine's explanations reflects a high degree of simulated feedback literacy, which is a skill that the student reviewers are clearly still developing. The feedback generated by the pre-service teachers exhibited differences rooted in their shared classroom context. While their feedback lacked analytical precision, it heavily emphasized immediate and observable performance aspects such as pronunciation or vocal confidence. The human reviewers usually stopped after identifying an issue and rarely offered concrete follow-up steps. To summarize the findings, a table will be presented below:

Table 1. Comparison of AI-Generated and Human-Generated Peer Feedback Characteristics

Feedback Quality Category	AI-Generated Feedback Characteristics	Human-Generated Feedback Characteristics
Content and Explanatory Depth	High rubric alignment. Justifies evaluations pedagogically. Offers deep reasoning.	Identifies issues without elaboration. Provides purely observational judgments.

Specificity and Constructiveness	References specific slides and details. Offers practical suggestions. Functions formatively.	Focuses on observable surface details. Lacks actionable follow-up steps.
Style, Questions, and Encouragement	Uses direct second-person tone. Includes reflective questions.	Uses distant third-person tone. Conveys authentic personal context.

Strengths and Limitations of Each Feedback Type

To address the second research question, it is important to combine the positive and negative attributes of both evaluation methods. Table 2 provides a summary of the distinct strengths and limitations of each source of peer feedback, alongside their broader pedagogical implications for language education.

Table 2. Strengths and Limitations of AI and Human Peer Feedback

Feedback Source	Primary Strengths	Notable Limitations	Pedagogical Implications
Artificial Intelligence	High structural consistency. Deep analytical precision. Actionable suggestions. Formative modeling.	Lacks situational awareness. Cannot perceive emotional nuance. Highly dependent on user prompt quality.	Serves as an excellent scaffolding tool to model proper academic critique and structure for developing teachers.
Human Peers	High authenticity. Empathetic engagement. Shared contextual awareness. Genuine interpersonal support.	Brief and surface-level analysis. Descriptive rather than analytical. Limited by developing pedagogical expertise.	Fosters a supportive classroom environment and helps students recognize immediate, observable delivery traits.

The main strength of artificial intelligence feedback is its structural consistency and analytical precision. The model provides highly readable and detailed evaluations that align closely with standard academic rubrics. However, the artificial intelligence presents notable limitations regarding contextual and affective understanding. The system cannot perceive the classroom atmosphere or the emotional state of the presenter. Furthermore, the effectiveness of the system depends entirely on the quality of the prompt it receives.

On the other hand, human-generated feedback offers distinct strengths, lying in authenticity and shared situational awareness. The seventh-semester pre-service teachers had a deep understanding of the specific classroom context and the interpersonal relationships among their peers. This shared experience allows them to provide empathetic and highly contextualized encouragement. They can highlight nuanced behavioral aspects that digital models entirely overlook. Despite these interpersonal strengths, human peer feedback is limited by the reviewers' developing pedagogical expertise. The students often focus solely on easily observable traits such as pronunciation or vocal volume rather than providing analytical depth. Consequently, their capacity to provide highly constructive and differentiated feedback is still developing during their final semesters of preparation.

DISCUSSION

The primary objective of this research is to measure and compare the quality of peer feedback provided by seventh-semester English Education students against feedback generated by an artificial intelligence model. To provide a comprehensive interpretation of the findings, this discussion section is structured to systematically address the two central research questions guiding this investigation. Drawing on established theoretical frameworks and empirical evidence from comparative feedback research, each subsection examines the reasons for the observed patterns and relates them within a broader scholarly research on feedback quality, assessment literacy, and technology integration in language education.

Differences in Feedback Quality

The findings reveal a consistent pattern of qualitative divergence between artificial intelligence and human peer feedback across all three dimensions of analysis: explanatory depth, specificity, and dialogic engagement. These differences are interpretable through multiple theoretical lenses and carry significant implications for how feedback provision is understood as a developmental competency in pre-service teacher education.

Explanatory Depth and Feedback Literacy

The most fundamental distinction between the two feedback sources lies in the degree to which evaluative comments were supported by reasoning and pedagogical justification. Effective feedback, as defined by Butenko (2024), is characterized by its concreteness, constructiveness, and dialogic nature. The artificial intelligence feedback showed these characteristics through its consistent practice of linking observed behaviors to their communicative or pedagogical effects, therefore producing comments that functioned as explanatory texts rather than evaluative decisions. This explanatory tendency relates to what Sunarto et al. (2024) described as feedback literacy, which included the ability to interpret, engage with, and act upon evaluative information for sustained improvement. The artificial intelligence model exhibited a high degree of simulated feedback literacy by constructing evaluations that made explicit the reasoning chain from observation to implication, a capacity that the student reviewers demonstrably had not yet fully developed.

The human feedback demonstrated the descriptive tendencies that Hansen (2025) identified as characteristic of underdeveloped feedback literacy among language learners. Comments such as "mispronounced words" or "relied on reading" correctly identify observable issues but do not explain how these behaviors affect the quality of the overall presentation. This pattern suggests that while the seventh-semester students had sufficient observational competencies to identify surface-level performance issues, they had not yet developed the evaluative reasoning capacity needed to articulate the broader communicative impact of those issues. Driscoll (2025) established that feedback performs its highest educational function when it operates as a dialogic and interactive process, promoting individual learning through explicit communication between evaluator and the receiver. The short observation of the human feedback limits this dialogic potential and, as a result, limits the transformative value of the peer review process for the presenters involved.

This developmental gap in feedback literacy is positioned within a recognized path of professional preparation. Although seventh-semester students have accumulated experience through teaching practicums and theoretical coursework, a feedback-related course is a discipline-specific competency that requires explicit instruction and deliberate practice to develop. Rahman et al. (2025) identified that pre-service teachers' feedback skills are shaped by both internal factors, including pedagogical understanding, and external factors such as institutional support and structured preparation for assessment tasks. The feedback profiles in this study suggest that explicit instruction in analytical feedback construction has not been fully integrated into the participants' professional preparation, despite its importance for their future professional roles as educators who will regularly assess and respond to student performance.

Specificity and the Formative Assessment Function

The differences in specificity observed between artificial intelligence and human peer feedback reflect a fundamental distinction between summative and formative assessment tendencies. Hattie et al. (2021) demonstrated that feedback oriented toward "where to next," combining specific diagnosis with concrete guidance for improvement, produces the most significant learning gains in comparison to feedback that remains at the level of general evaluation. The artificial intelligence feedback consistently embodied this formative orientation by pairing each identified issue with a specific and actionable recommendation. Comments referencing particular slide numbers, precise presentation details, and concrete corrective strategies provided the presenter with an immediately actionable roadmap for improvement. The human feedback, which typically concluded after identifying an issue, fulfilled an evaluative function but did not fully use the formative potential that the peer review process is theoretically designed to deliver.

The multifaceted nature of presentation performance further showed this specificity gap. Setiawan (2021) proposed a comprehensive assessment framework for student presentations including language proficiency, delivery method, and content mastery, illustrating the analytical complexity that rigorous oral performance

evaluation demands. The artificial intelligence model engaged with this complexity by addressing multiple aspects of the presentation simultaneously, producing evaluations that reflected an integrative understanding of performance quality. The human reviewers mostly focused on a narrower range of observable performance features, particularly those related to vocal delivery and note reliance. This selective focus aligns with Khorwal et al.'s (2024) observation that peer assessors of student presentations tend to prioritize the most observable and recognizable performance features rather than engaging systematically with the full range of assessment criteria, defaulting to a perceptual rather than analytical evaluative stance.

The structured rubric used in this study, adapted from Paltridge and Starfield (2020), provided both human and artificial intelligence reviewers with an identical framework for evaluation. The differential outcomes suggest that access to a structured rubric alone is not enough to produce analytical peer feedback among pre-service teachers. Gao et al. (2023) found that structured rubric training sessions significantly improve the quality of peer feedback in higher education contexts, reinforcing the position that rubric-guided training, rather than rubric access alone, is the mechanism through which evaluative precision develops. The artificial intelligence model demonstrated this precision naturally through its training on large-scale instructional data, whereas the human reviewers required a form of structured preparation for the task that was apparently not provided within the course design.

Tone, Style, and Dialogic Engagement

The tone and style differences between the two feedback sources reflect a significant way in which the feedback relationship was constructed and how the presenter was positioned within the evaluative exchange. The artificial intelligence model's consistent use of second-person address and its inclusion of reflective questions, such as "How will you assess performance qualitatively?", position the presenter as an active participant in the evaluation process. This orientation aligns with the dialogic feedback principles articulated by Lam and Habil (2020), who identified the co-construction of understanding through feedback interaction as central to the effectiveness of peer review in higher education. Feedback that invites reflection rather than delivering judgments and positions the receiver as a person capable of engaging critically with their own performance therefore deepens the learning value of the evaluative encounter.

The theoretical foundation for this dialogic tendency comes from the social constructivist principles established by Vygotsky (1962), who demonstrated that knowledge construction occurs most effectively through purposeful social interaction. Peer feedback, as a structured form of social interaction centered on performance evaluation, embodies this constructivist principle by creating opportunities for reflective dialogue between learner-evaluators. The artificial intelligence model approximated this dialogic function through its conversational tone and reflective questioning, despite the fundamentally non-social nature of machine-generated output. The human reviewers, by employing distant third-person phrasing in their evaluations, constructed a formal evaluative distance that worked against the dialogic and socially constructive principles that peer feedback is theoretically designed to enact. This finding is particularly significant because it suggests that the pre-service teachers, despite their theoretical familiarity with constructivist pedagogy through coursework, did not transfer those principles into their evaluative practice.

Ladyshewsky (2012) established that the defining value of peer feedback lies in the co-learner relationship between reviewer and recipient, which generates an evaluative engagement qualitatively different from instructor or automated assessment. The human reviewers offered genuine co-learner recognition in their opening remarks, demonstrating an authentic understanding of the interpersonal context that a machine cannot replicate. This interpersonal authenticity constitutes a distinct and irreplaceable quality of human peer feedback, even where the analytical depth falls short of pedagogical standards. The central challenge for English education programs is to leverage this authentic relational quality while simultaneously developing the analytical and explanatory competencies that would allow peer feedback to reach its full formative potential.

Strengths and Limitations of Each Feedback Type

The findings revealed the strengths and limitations of artificial intelligence feedback and human feedback provided by the pre-service teachers. The implications and interpretations are categorized into two sections that are structured to show each feedback type's strengths and limitations.

Strengths and Limitations of Artificial Intelligence Feedback

The artificial intelligence feedback's principal strength lies in its structural consistency and analytical depth across all five samples in the dataset. Dai et al. (2024) observed that automated feedback systems demonstrate

strong utility in educational contexts because of their systematic assessment criteria and their alignment with rubric-based evaluation standards. This consistency is particularly valuable in large contexts, where the artificial intelligence feedback addresses the practical constraints that limit individualized teacher feedback delivery. Chen et al. (2022) and Lai et al. (2023) similarly documented that AI-assisted feedback platforms offer self-paced and tailored learning support that reduces dependence on constant teacher intervention, positioning artificial intelligence as an efficient way for maintaining consistent feedback quality across large groups of learners.

The capacity of artificial intelligence to model analytically elaborated feedback further positions it as a pedagogical resource for developing feedback literacy among pre-service teachers. Klimova et al. (2024) documented that large language models function effectively as educational companions and personal tutors, delivering accessible, adaptable, and pedagogically complex responses that simulate expert guidance. The artificial intelligence feedback in this study showed this tutoring function by producing comments that made visible the reasoning patterns, structural conventions, and formative tendencies of high-quality academic critique. Sustained exposure to this model-like feedback during pre-service preparation may accelerate the development of evaluative reasoning skills by providing a concrete exemplar of the analytical standards toward which peer feedback provision should develop.

The limitations of artificial intelligence feedback are equally significant for a comprehensive interpretation of the findings. Mostafapour et al. (2024) found that while large language models demonstrate an impressive range of knowledge and rapid response capacity, they showed critical weakness in contextual depth and occasionally produce assessments that lack situational accuracy and nuance. In this study, the artificial intelligence model evaluated each presentation as just a data input, addressing the content and delivery of the performance in isolation from the specific institutional, interpersonal, and cultural context in which it occurred. This detached approach produces technically precise feedback that may nonetheless fail to resonate with presenters who experienced the performance as a socially embedded classroom event, therefore limiting its motivational and interpersonal impact regardless of its analytical quality.

The affective limitations of artificial intelligence feedback represent a related but distinct concern. Chen et al. (2025) revealed that technically proficient artificial intelligence feedback does not automatically generate positive affective responses in learners, as motivational and emotional dimensions require deliberate consideration in instructional design. Feedback that addresses the intellectual content of a performance without acknowledging its interpersonal dimensions may be perceived as impersonal, potentially reducing its motivational value. This limitation reinforces the pedagogical position that artificial intelligence feedback is most appropriately used as a supplementary analytical resource rather than as a comprehensive replacement for the relational dimensions of human peer assessment.

The prompt dependency of the artificial intelligence system introduces an additional qualification to its perceived advantages. Qian (2025) identified that structured, goal-oriented prompting frameworks significantly enhance the quality of artificial intelligence outputs, while insufficiently specified prompts yield correspondingly less effective responses. Dafitri et al. (2025) extended this argument by establishing that prompt engineering constitutes an essential pedagogical competency for educators seeking to leverage generative artificial intelligence as an assessment and instructional tool. Their research demonstrates that when educators apply clear, context-specific parameters to configure artificial intelligence systems, the technology can function as a collaborative partner in the design of meaningful assessments. The carefully engineered prompt employed in this study, which instructed the model to simulate the perspective and evaluative behavior of a seventh-semester English Education student, was instrumental in producing the pedagogically aligned feedback observed in the dataset. This dependency on prompt quality means that the educational value of artificial intelligence feedback is not an inherent property of the technology itself, but a function of the user's capacity to configure it effectively, introducing an additional competency requirement for both educators and students who seek to utilize artificial intelligence as a feedback tool.

Strengths and Limitations of Human Peer Feedback

The authentic and contextually grounded quality of human peer feedback represents a pedagogical resource that artificial intelligence systems are structurally unable to replicate. The seventh-semester students who participated as reviewers in this study shared the institutional context, academic pressures, coursework history, and interpersonal relationships of the presenters they evaluated. This shared experiences enabled them to provide recognition and encouragement that carried genuine social meaning within the specific learning community. Lam and Habil (2020) identified the creation of a supportive learning environment as a primary benefit of peer feedback in higher education, and Ladyshewsky (2012) established that the co-learner relationship between reviewer and

recipient generates a form of evaluative engagement qualitatively different from instructor or automated feedback. The authentic interpersonal support demonstrated in the opening remarks of the human reviewers showeed this relational quality, offering presenters recognition from individuals who genuinely understood the demands of the shared academic context.

The human reviewers also demonstrated perceptual attunement to the lived dimensions of the presentation that extends beyond the scope of standardized rubric criteria. Suharni et al. (2022) documented that peer feedback on oral presentations frequently captures nuanced behavioral and relational aspects of classroom performance that structured instruments may not fully address. This perceptual sensitivity reflects the reviewer's direct experience as a co-present observer within the shared learning environment, allowing them to attend to contextually meaningful performance aspects that go beyond the analytical framework of a rubric. This capacity for contextually sensitive observation merits explicit recognition as a dimension of peer feedback quality, even where the overall analytical depth of the feedback remains limited.

The limitations of the human peer feedback in this study are best understood within the developmental framework appropriate to the participants' stage of professional preparation. Rahman et al. (2025) established that pre-service teachers' feedback provision skills continue to develop in response to evolving pedagogical knowledge and accumulated practical experience, with skills maturing throughout the final semesters of degree completion. The surface-level analytical quality observed in the human feedback suggests that the explicit application of pedagogical principles to evaluative practice is an area requiring further development, even among students who have completed teaching practicum requirements. This aligns with Hansen's (2025) observation that feedback literacy remains underdeveloped among language learners at this stage, indicating that the transition from observer to analytical evaluator requires instructional support beyond what standard coursework currently provides.

Wanchid and Charoensuk (2024) further demonstrated that student-generated feedback quality varies significantly in relation to reviewer achievement levels, introducing a reliability concern for peer assessment systems that assign evaluative authority to student reviewers without structured preparation. Gao et al. (2023) identified that rubric-guided training substantially improves peer feedback quality in higher education, reinforcing the position that the surface-level evaluations produced by the student reviewers in this study are partly attributable to the absence of structured preparation for the feedback task. This finding suggests that the developmental limitations observed in the human feedback are at least partially addressable through targeted feedback literacy instruction, rather than being an invariable feature of this stage of professional development.

Pedagogical Implications

The complementary profile of strengths and limitations identified across both feedback types indicates that an integrated pedagogical approach, incorporating both artificial intelligence and human peer feedback in a deliberate instructional context, offers the most comprehensive framework for developing presentation skills and feedback literacy in English education programs. Artificial intelligence feedback can serve as a modeling and scaffolding resource, providing pre-service teachers with consistent exposure to analytically elaborated and structurally coherent examples of high-quality evaluative commentary. Klimova et al. (2024) established that large language models are effective in this tutoring function, and sustained exposure to model-like feedback may develop students' understanding of what analytical standards in peer evaluation require, therefore informing and progressively improving their own feedback provision practice.

Human peer feedback, in its authentic and interpersonally grounded form, provides a relational and contextually sensitive dimension of evaluation that artificial intelligence cannot replicate. Lam and Habil (2020) identified this relational quality as central to the supportive learning environment that effective peer assessment is designed to create. Positioning artificial intelligence and human peer feedback as complementary ways rather than competing alternatives allows educators to design assessment systems that address both the analytical and relational dimensions of the feedback process. A structured integration model, in which students first engage with artificial intelligence feedback to develop an understanding of analytical standards and then participate in human peer review to cultivate relational and contextually responsive evaluative skills, could produce a more balanced and developmentally appropriate feedback ecology for pre-service teacher education.

Furthermore, such an integration model requires explicit feedback literacy instruction as a curriculum component. Sunarto et al. (2024) established that feedback literacy is foundational to long-term academic and professional competence, and Gao et al. (2023) demonstrated that structured rubric training substantially improves peer feedback quality. Incorporating formal feedback literacy instruction into the English Education program curriculum, alongside explicit training in prompt engineering for artificial intelligence utilization, would address the identified skill gaps and better prepare pre-service teachers for the evaluative responsibilities of their future professional roles.

CONCLUSION

This study examined the difference in quality between human-generated peer feedback and artificial intelligence peer feedback when reviewing student presentation videos. The research used a qualitative descriptive design to analyze multiple aspects of quality, including clarity, specificity, constructiveness, tone, and pedagogical value. The results showed that artificial intelligence feedback exhibited a greater depth and consistency that explained the reasoning behind its evaluations. It also included proper suggestions for improvement for the presenter. The human-generated feedback showed a contextual understanding and sincerity. This implies that artificial intelligence excels at providing structured and analytical precision, while human feedback shows strength in authenticity and interpersonal support.

These findings indicate that artificial intelligence can serve as a structured analytical commentary model for students. Human feedback can serve as commentary from an actual person who shares the same learning perspective. This means that artificial intelligence can function as a complementary tool to human feedback to further a student's understanding of their performance. For teachers, artificial intelligence can serve as a supplementary tool to help students learn what effective feedback looks like. However, artificial intelligence cannot solely serve as a standalone commentary and cannot replace human feedback. Future researchers could explore how a mixed feedback approach influences student presentation skills over time by expanding the samples across different academic backgrounds.

ACKNOWLEDGEMENTS

Thank you to all the participants who have taken part in this research. This acknowledgement also applies to the lecturer who have supported and have helped in this research. Lastly, thank you to the state University of Karawang that has allowed this research to be conducted.

REFERENCES

Aliakbari, M., Barzan, P., & Sayyadi, M. (2021). Automated Feedback vs. Human Feedback: A Study on AI-Driven Language Assessment. AI And Tech in Behavioral and Social Sciences, 3(2). https://doi.org/10.61838/kman.aitech.3.2.9

Alif, A. L. S., Hamid, A. I. R., & Khatimah, K. (2025). Pre-service English Teachers' Perception on The Challenges and Strategies During Their Teaching Practice Program. English Language Teaching Methodology, 5(2), 124–138. https://doi.org/10.56983/eltm.v5i2.1867

Al-Ogaili, N. (2023). Feedback in Teaching English Electronically to TESOL. Social Science Research Network. https://doi.org/10.2139/ssrn.4531858

Butenko, L. (2024). The Phenomenon of Feedback in the Educational Process of Higher Education Institution: Explicit and Implicit Contexts. Osvìta Ta Pedagogìčnì Nauki, 2(186), 24–37. https://doi.org/10.12958/2227-2747-2024-2(186)-24-37

Chen, J., Lai, P., Chan, A., Man, V., & Chan, C.-H. (2022). AI-Assisted Enhancement of Student Presentation Skills: Challenges and Opportunities. Sustainability, 15(1), 196. https://doi.org/10.3390/su15010196

Chen, Z., Wei, W., & Zou, D. (2025). Generative AI technology and language learning: global language learners' responses to ChatGPT videos in social media. Interactive Learning Environments, 1–14. https://doi.org/10.1080/10494820.2025.2511248

Dafitri, F., Hakim, P. K., Wigati, F. A., Pujiawati, N., & Rahmawati, M. (2025). Penulisan Butir Soal Mata Pelajaran Bahasa Inggris dengan Memanfaatkan Generative Artificial Intelligence (GenAI): Pelatihan Peningkatan Kompetensi Guru pada MGMP Bahasa Inggris Kecamatan Cipayung Jakarta Timur. Ihsan: Jurnal Pengabdian Masyarakat, 7(1), 24-32. https://doi.org/10.30596/ihsan.v7i1.24073

Dai, W., Tsai, Y., Lin, J., Aldino, A. A., Jin, F., Li, T., Gašević, D., & Chen, G. (2024). Assessing the Proficiency of Large Language Models in Automatic Feedback Generation: An Evaluation Study. https://doi.org/10.35542/osf.io/s7dvy

Driscoll, A. (2025). Feedback for students and teachers (pp. 187–208). https://doi.org/10.4324/9781003445234-8

Ekizoğlu, M., & Demir, A. N. (2025). AI-Assisted Writing Feedback for Enhancing Secondary Students' Writing Skills: An Experimental Study. https://doi.org/10.21203/rs.3.rs6430737/v1

Ficayuma, L. A. (2022). Identifying language proficiency and sla of english education department students for designing an intercultural communication e-book. International Research-Based Education Journal, 4(2), 83. https://doi.org/10.17977/um043v4i2p83-99

Gao, X., Noroozi, O., Gulikers, J., Biemans, H. J. A., & Banihashem, S. K. (2023). A systematic review of the key

components of online peer feedback practices in higher education. *Educational Research Reviews*. https://doi.org/10.1016/j.edurev.2023.100588

Hansen, R., Prilop, C. N., Nielsen, T. A., Møller, K. L., Hougaard, R. F., & Lindberg, A. B. (2025). The effects of an AI feedback coach on students' peer feedback quality, composition, and feedback experience. *Tidsskriftet Læring Og Medier (LOM)*, *17*(31). https://doi.org/10.7146/lom.v17i31.148831

Hattie, J., Crivelli, J., Van Gompel, K., West-Smith, P., & Wike, K. (2021). Feedback that leads to improvement in student essays: Testing the hypothesis that "where to next" feedback is most powerful. *Frontiers in Education*, *6*. https://doi.org/10.3389/FEDUC.2021.645758

Kasriyati, D., Herdi, H., & Abbas, M. F. F. (2022). Identifikasi Kemampuan Mahasiswa dalam Praktek Mengajar pada Mata Kuliah Teaching English as a Foreign Language (TEFL). Lectura, 13(1), 99–107. https://doi.org/10.31849/lectura.v13i1.7494

Khorwal, G., Meshram, R., Vaibhav, V., Sharma, R., Singh, B., Chandran, S., & Sharma, K. (2024). Peer assessment of student presentations: Key takeaways and lessons learned. *Cureus*. https://doi.org/10.7759/cureus.59809

Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology*, *15*, 1269319. https://doi.org/10.3389/fpsyg.2024.1269319

Ladyshewsky, R. K. (2012). The role of peers in feedback processes (pp. 174–189). Routledge. https://doi.org/10.4324/9780203074336-16

Lai, P. P. Y., Chen, J., Man, V., & Chan, C.-H. (2023). A New Frontier in AI-Assisted English Oral Presentation Assessment. 1–8. https://doi.org/10.1109/tale56641.2023.10398320

Lam, C. N. C., & Habil, H. (2020). Peer Feedback in Technology-Supported Learning Environment: A Comprehensive Review. The International Journal of Academic Research in Business and Social Sciences, 10(9). https://doi.org/10.6007/IJARBSS/V10-I9/7866

Mostafapour, M., Fortier, J. H., Pacheco, K., Murray, H., & Garber, G. (2024). ChatGPT vs. Scholars: A comparative examination of literature reviews conducting by humans and AI (Preprint). JMIR AI, 3, e56537. https://doi.org/10.2196/56537

Paltridge, B., & Starfield, S. (2020). *Thesis and dissertation writing in a second language a handbook for students and their supervisors*. https://doi.org/10.4324/9781315170022

Qian, Y. (2025). Prompt Engineering in Education: A Systematic Review of Approaches and Educational applications. Journal of Educational Computing Research. https://doi.org/10.1177/07356331251365189

Rahman, H., Irfan, M., Yusuf, F., Ali, A. M., & Abadi, A. U. (2025). Analysis of pre-service teachers' skills in providing feedback to students during field experience practice in school. IJORER International Journal of Recent Educational Research, 6(2), 544–564. https://doi.org/10.46245/ijorer.v6i2.799

Rusu, O. C., & Bozov, P. (2024). Assessment of Oral Presentation Skills in Teaching English for Specific Purposes. Dialogos. https://doi.org/10.24818/dlg/2023/40/09

Setiawan, D. (2021). Improvement of presentation performance through feedback. *Journal of Applied Studies in Language*, *5*(1), 156–164. http://ojs.pnb.ac.id/index.php/JASL

Suharni, S., Imelwaty, S., Sesmiyanti, S., & Perpisa, L. (2022). The impact of peer feedback on higher education students' oral presentation: A case study. *Tell-Us Journal: Teaching-English-Linguistics-Literature-Usage*, *8*(3), 141–151. https://doi.org/10.22202/tus.2022.v8i3.6181

Sunarto, M. J. D., Kisoworo, A. Y., & Lemantara, J. (2024). The road to better assessment: Unleashing the potential of feedback literacy through bibliometric analysis in education. *IJORER*, *5*(5), 1065–1083. https://doi.org/10.46245/ijorer.v5i5.649

Tamayo, N., Lane, A., & Dewart, G. (2020). Qualitative Description Research: An Examination of a Method for Novice Nursing Researchers. International Journal of Nursing, 7. https://journalhosting.ucalgary.ca/index.php/ijnss/article/view/71786

Vygotsky, L. (1962). *Thought and language.* (E. Hanfmann & G. Vakar, Eds.). Boston Review. https://doi.org/10.1037/11193-000

Wanchid, R., & Charoensuk, V. (2024). Effect of feedback types and English achievement on EFL oral presentation performance of Thai undergraduates. *PASAA*, *69*, 153–194. https://doi.org/10.58837/chula.pasaa.69.6

APPENDICES

Appendix 1. Paltridge & Starfield (2020) presentation review rubric.

Category	Indicators	Score Range
Clarity of Presentation	Clear structure, smooth flow, understandable delivery	1-5
Introduction	Relevant background, clear problem statement, justified research gap	1-5
Research Questions/Objectives	Well-formulated, researchable, aligned with problem and method	1-5
Literature Review	Relevant, current sources, critical synthesis, shows gap in knowledge	1-5
Framework (Theoretical/Conceptual)	Clearly explained and aligned with study purpose	1-5
Research Methodology	Appropriate design, justified method, logical data collection plan	1-5

Score Key: 1 = Poor; 2 = Fair; 3 = Good; 4 = Very Good; 5 = Excellent
In addition to numerical ratings, reviewers were required to provide qualitative feedback in three structured sections:

- **A. Strengths I Observed**: Identification of effective presentation elements
- **B. Suggestions for Improvement**: Constructive recommendations for enhancement
- **C. One Question I Have About Your Research**: Engagement through critical inquiry
- **D. Clarity of Presentation**: Holistic rating (Poor to Excellent)
  E. Feasibility/Readiness for Data Collection: Assessment of research preparedness (Poor to Excellent)

Appendix 2. Qualitative Descriptive Rubric for Feedback Quality (Adapted from Hansen, 2025)

Category	Indicators	Score Range
Clarity of Presentation	Clear structure, smooth flow, understandable delivery	1-5
Introduction	Relevantbackground, clear problem statement, justified research gap	1-5
Research Questions/Objectives	Well-formulated, researchable, aligned with problem and method	1-5
Literature Review	Relevant, current sources, critical synthesis, shows gap in knowledge	1-5
Framework (Theoretical/Conceptual)	Clearly explained and aligned with study purpose	1-5
Research Methodology	Appropriate design, justified method, logical data collection plan	1-5