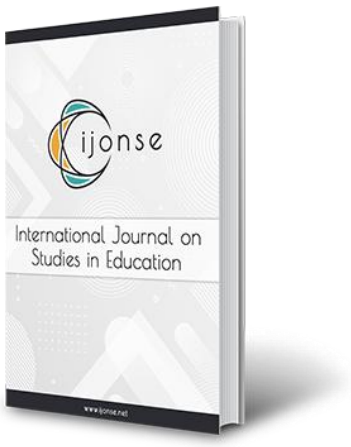

www.ijonse.net

Analysis of AI and Teacher Feedback in Grading Students' Activities

Marjorie Joy Mariano
Rizal Technological University, Philippines

Javilyn Joy Taborete
Rizal Technological University, Philippines

Andrew Gampal
Rizal Technological University, Philippines

Aragorn Dawisan
Rizal Technological University, Philippines

Mary Grace Maraganas
Rizal Technological University, Philippines

Rizaldy Garcia
Rizal Technological University, Philippines

To cite this article:

Mariano, M.J., Taborete, J.J., Gampal, A., Dawisan, A., Maraganas, M.G., & Garcia, R. (2025). Analysis of AI and teacher feedback in grading students' activities. *International Journal on Studies in Education (IJonSE), 7*(4), 1028-1047. https://doi.org/10.46328/ijonse.5148

International Journal on Studies in Education

2025, Vol. 7, No. 4, 1028-1047 https://doi.org/10.46328/ijonse.5148

Analysis of AI and Teacher Feedback in Grading Students' Activities

Marjorie Joy Mariano, Javilyn Joy Taborete, Andrew Gampal, Aragorn Dawisan, Mary Grace Maraganas, Rizaldy Garcia

Article Info

Article History

Received:
3 March 2025
Accepted:
18 July 2025

Keywords

Artificial intelligence
ChatGPT
Student assessment
Teacher's feedback
Technology education

Abstract

Artificial Intelligence has recently emerged as an important topic in the educational landscape and how it affects teachers' professional knowledge. We investigated the alignment of Artificial Intelligence (AI) specifically ChatGPT in providing feedback on student output compared to teachers in the science classroom. Using an explanatory sequential mixed-methods design, we analyzed 47 student submissions assessed by both teachers and ChatGPT, guided by pre-established grading rubrics on two different student outputs of a certain university in Mandaluyong City, Philippines. Using Spearman's Rho and Cohen's Kappa we quantitatively explored the correlations of the evaluations of both AI and Teachers. It revealed weak correlations and low inter-rater agreement, indicating limited alignment between AI-generated and teacher evaluations, especially in subjective and interpretive components. Qualitatively, we interviewed science teachers, the result highlighted that, while AI feedback was efficient, consistent, and structured, it often lacked contextual depth, emotional tone, and pedagogical insight. Teachers valued AI's ability to support routine tasks but emphasized the irreplaceable role of human judgment in assessing higher-order thinking and student-specific needs. We concluded that AI is a valuable supplementary tool rather than a replacement for educators. Our findings contribute to the ongoing discourse on ethical and pedagogically sound integration of AI in classroom assessment practices.

Introduction

The arrival and exponential growth of Artificial Intelligence (AI) are transforming education. ChatGPT of OpenAI is increasingly being applied in schools through automating marking and providing feedback on student work. The application of such systems provides a host of advantages that can make a significant difference to the learning experience of students. Firstly, they assist teachers in saving time, which can be utilized in taking up more significant tasks. Secondly, AI assists in the standardization of grading, removing individual bias from marking work that occurs when a teacher does grades. This minimizes the teacher's workload as they would experience with large class sizes (Delello et al., 2025; Magtalas, 2024). One of the greatest advantages of applying AI in education is that it can apply grading rules uniformly. This is superior to human graders, who might be influenced by fatigue or bias. This aspect is extremely helpful in large classes, where uniform grading can be difficult to attain (Ragolane et al., 2024). Moreover, AI provides feedback, and the students can get instant information about

their performance. This instant feedback can have the power to make the learning process interesting, so that the students can learn and enhance quickly (Seo et al., 2021). Though AI is beneficial in education, there are real problems with the quality and suitability of AI-provided feedback. It is not capable of dealing with subjective feedback that should be thought through more, analyzed more thoroughly, and pictured (Gobrecht et al., 2024; Zhang et al., 2025). For instance, Kemal and Liman-Kaban's (2024) study revealed that AI tends to stick to technical writing norms but differs from human judgment in aspects of comprehension and in-depth assessment.

In science education, students do not merely memorize facts; they must think critically and comprehend concepts thoroughly (Kotsis, 2024). The research "Towards Adaptive Feedback with AI" revealed that although AI provides accurate feedback on science activities, it sometimes lacks the contextual understanding that human feedback can provide (Seßler et al., 2025). This emphasizes what AI cannot do well in contexts where personal insight and profound understanding matter (Vieriu & Petrea, 2025). Another study indicated that students appreciate just how immediate and unbiased AI feedback is, but they do perceive that it is less real or encouraging than feedback from humans (Petersen, 2024).

Educational feedback is not one-dimensional. It is not merely identifying errors; it also encourages students, assists them in persevering, and improves critical thinking (Câmpean et al., 2024; Evans, 2013; Thompson et al., 2023). For teachers, using AI can have both positive and negative sides. The majority of teachers believe it can improve efficiency, but they worry about AI understanding good writing and developing emotional relationships with students (Mogavi et al., 2023; Seo et al., 2021). One study highlighted the importance of correct training and revising AI-curricula to guarantee the effective and responsible use of AI in science education (Chichekian & Benteux, 2022).

Transparency and equity complicate the role of AI in assessments (Memarian & Doleck, 2023). Bulut et al. (2024) identified that AI inherits biases from the training data, which can further perpetuate inequality. Additionally, AI's "black box" creates issues about accountability and trust in AI feedback because the inner mechanisms of such systems are typically unknown to the users. There is extensive research on the use of AI to support education, yet numerous key questions are still unaddressed (Ali et al., 2023; Al-Jahwari & Yousif, 2025; Gayed, 2025). For example, there is minimal evidence comparing teacher feedback to AI feedback, specifically for science projects. More research is needed to establish how effectively AI performs regarding the grading rubrics for objective and subjective exams (Awidi, 2024; Terrazas-Arellanes et al., 2025). In addition, while AI can be effective, there are not enough in-depth studies comparing the time and resources needed for AI feedback and feedback from teachers in real-world classrooms (Memarian & Doleck, 2023).

This study was designed in response to the growing presence of Artificial Intelligence in education, particularly in the area of assessment. In this study we used ChatGPT to automate grade and provide feedback related to student’s activities. As classrooms become increasingly augmented by intelligent systems, it becomes essential to critically examine how these technologies align with the standards and expectations long held by educators. The study set out to explore the degree to which AI systems adhered to pre-established grading rubrics in comparison to teachers, a question that is central to understanding the validity and fairness of AI-assisted assessment. In doing

so, the research aimed to ground the conversation on AI in a pedagogical framework that respects the complexity of learning, especially in science education, where understanding goes beyond correct answers.

At the same time, the study recognized that assessment is not merely a technical process, but a relational and reflective one. By identifying the potential strengths and limitations of AI feedback, the research aimed to contribute to a more balanced and thoughtful approach to technology integration. Rather than framing AI as a replacement for educators, the study positioned it as a tool whose value must be continually assessed in light of human judgment, pedagogical context, and the deeper purposes of education.

Research Literature

Teacher and AI Alignment with Pre-established Grading Rubrics

AI systems align effectively with pre-established grading rubrics by ensuring the consistent application of criteria without human biases, which is a significant advantage, particularly in large-scale assessments (Gnanaprakasam & Lourdusamy, 2024; Trikoili et al., 2025). Grading consistency is often a challenge in environments with multiple human evaluators, where variability can arise due to unconscious biases or subjective judgment (Malouff & Thorsteinsson, 2016). AI systems can mitigate this issue by adhering strictly to the standardized grading criteria, making AI especially valuable in situations where fairness and uniformity are critical (Gnanaprakasam & Lourdusamy, 2024). This is particularly true when assessing quantifiable metrics, such as grammar, structure, and factual accuracy. AI can apply these criteria with precision, offering a consistent and objective evaluation that aligns with surface-level aspects of the grading rubric (Atasoy & Arani, 2025; Topping et al., 2025).

Studies consistently highlight AI's strength in aligning with pre-established grading rubrics, particularly for objective criteria such as grammar, structure, and factual accuracy (Bouziane & Bouziane, 2024; Myint et al., 2024). Kemal and Liman-Kaban (2024) emphasized that AI systems excel in applying rubric-based standards consistently, reducing human bias and grader variability. This is particularly advantageous in large-scale assessments where maintaining fairness is crucial (Herman & Cook, 2019; Tierney, 2016). A critical component in assessing the validity of feedback systems is their alignment with rubric criteria. While artificial intelligence offers rapid evaluations, recent studies question its precision in nuanced assessments (BediZel, 2023; Swiecki et al., 2022). Literatures also points out AI’s limitations in subjective assessments (Almatrafi et al., 2024). AI systems struggle with evaluating creative outputs, critical thinking, and nuanced argumentation. Human teachers surpass AI in interpreting intent, tone, and contextual relevance, which are often beyond AI's algorithmic scope (Vieriu & Petrea, 2025; Al-Zahrani, 2024). AI systems encounter difficulties when evaluating more subjective aspects of student work, also, tasks that require creativity, critical thinking, or deeper contextual understanding often elude AI’s capabilities (Malik et al., 2023; Lawasi et al., 2024). AI systems can struggle to assess originality, tone, and intent in student submissions, essential components of subjective evaluations (Awidi, 2024; Revell et al., 2024). As a result, while AI is highly effective in grading factual content, it may fall short when required to evaluate higher-order thinking skills, creative expression, or nuanced judgment (Jones-Jang et al., 2025: Malik et al., 2023; Chan & Hu, 2023).

Students' Assessment Feedback of Teachers and AI

AI feedback is widely recognized for its efficiency, clarity, and uniformity (Escalante et al., 2023; Jacobsen & Weber, 2025; Venter et al., 2024). AI-generated responses provide straightforward, linguistically simple feedback that is particularly beneficial for non-native English speakers and large classes (Escalante et al., 2023; Mogavi et al., 2023). The ability to deliver immediate feedback allows students to make timely improvements (Ajogbeje, 2023; O'Neil et al., 2010). Furthermore, AI systems are capable of providing instant feedback, which is especially beneficial in large or online classes where personalized attention from human instructors may be limited (Gligorea et al., 2023; Labadze et al., 2023; Seo et al., 2021; Wang et al., 2024;). In the same manner, AI feedback is often structured, direct, and simple, making it easily understandable for students (Escalante et al., 2023). This clarity is particularly advantageous for students who benefit from clear, concise instructions, such as non-native English speakers (Bolkan, 2015). By its very nature, therefore, AI feedback for student's outputs is uniform, meaning that every student receives feedback based on the same set of predefined rules and criteria, promoting fairness and equity (Singh et al., 2024; Swiecki et al., 2022; Trigo et al., 2024).

In contrast, feedback from teachers is generally richer, more empathetic, and contextually relevant (Adarkwah, 2021; Rowe, 2016; Wang et al., 2021). Teachers can offer personalized insights, drawing on their understanding of each student's learning journey, emotional state, and unique challenges (Frenzel et al., 2021; Gunawardena et al., 2023; Rahman et al., 2024). This feedback is often viewed as more meaningful and engaging because it is delivered with an understanding of the student's circumstances (Adarkwah, 2021; Evans, 2013). Moreover, teacher feedback provides motivation, encouragement, and emotional support, fostering a deeper connection between the student and teacher (Guo et al., 2025; Wang, 2025). Unlike AI, which cannot adapt its tone based on a student's emotional needs, teachers can adjust their feedback to be more supportive or motivational, depending on the situation (Nikitina & Ishchenko, 2024; Ruwe & Mayweg-Paus, 2023). Teacher feedback is also characterized by its personalized, empathetic, and context-aware nature (Aldrup et al., 2022; Rochera et al., 2021). Thus, teacher feedback motivates students, offers deeper insights, and fosters meaningful engagement and teachers consider the learner's journey, emotional state, and socio-cultural background, which enhances the quality of feedback and promotes deeper learning (Morrison & Jacobsen, 2023; Zheng, 2022; Zhang & Hyland, 2021).

Methodology

We aimed to explore how artificial intelligence (AI) compares to human teachers in delivering feedback on student assignments. We focused on two main aspects: the alignment of AI-generated feedback with established grading rubrics and the perceived quality of that feedback from an educational perspective. To achieve this, we employed an explanatory sequential mixed-methods design, beginning with quantitative data collection and analysis, followed by a qualitative phase to provide deeper insight. We worked with two science teachers who provided a set of forty-seven graded laboratory reports at a certain university in Mandaluyong City, Philippines, which we used to compare their assessments with those generated by an AI tool using the same rubric.

After the quantitative phase, we conducted interviews with five in-service science teachers who were not involved

in the initial grading. Using a semi-structured interview protocol, we asked them to evaluate and reflect on both teacher and AI feedback. We transcribed and thematically analyzed the interviews to understand participants' perceptions across various dimensions. All tools used in the study such as grading rubrics, the AI prompt, and the interview guide—were reviewed by experts to ensure validity and clarity. Through this design, we sought to build a comprehensive understanding of the potential and limitations of AI in educational feedback, grounded in both numerical alignment and lived user experience.

Participants

We involved two faculty members from the Science Department at Rizal Technological University (RTU). Both are licensed science teachers with a lot of experience teaching hands-on, lab-based courses at the college level. They were intentionally chosen because of their expertise and regular involvement in laboratory instruction, which made them a good fit for what this study set out to explore. One of the professors shared 15 student lab activities, while the other provided 32. These weren't just the student submissions; they came with the full package: the evaluation rubrics, student scores, and the feedback given by the teachers. All of these materials became the foundation for comparing how the teachers and the AI graded and gave feedback. Even though the number of activities varied between the two professors (mostly due to different class sizes and workloads), the process stayed consistent. Each student submission was evaluated using the original rubric from the respective teacher, keeping the assessment fair and aligned across the board.

Research Instrument and Data Gathering

We utilized 2 types of instruments; Instrument 1 is the prompt that was made by the researchers. This prompt is used because it helps with getting the response of AI (ChatGPT) on how it scores the activity of students and provides feedback on their performance in the given lab activity, along with these, the result will help to identify the themes that will be used to compare the work of teachers and the AI. The activities were provided by the 2 faculty and with pre-established rubrics. Instrument 1 or the prompt was validated by three faculty experts. They validated it on the following criteria: clarity, appropriateness of language and sequence and efficiency. The validators provide comments and suggestions which were then adapted by the researchers. The validated prompt was the one input in the AI (ChatGPT) system.

Instrument 2 is an interview guide, specifically a semi structured interview that was used in gathering data for the qualitative part of the study. Specifically, it was focused on the strengths and limitations of the AI system. The same instrument was also validated by the same faculty experts. The validation paper includes the improvement of the questions in terms of clarity, relevance, appropriateness, organization, and the use of language in the set of questions. The validators assessed and provided suggestions.

After all the necessary instruments were validated, we then gathered the data needed in our study. A letter was provided to the College of Education Dean of the Rizal Technological University to request permission for the conduct of interviews and data collection. The letter was then submitted to the school's Data Privacy Office

(DPO). After receiving approval, we proceeded with the two faculty members involved in the study. Letters were given to these faculty members requesting permission to collect their students' lab activities and the rubrics they used as the basis for grading those works. These materials were also intended to be used as input for AI (ChatGPT) assessment. Additionally, individual feedback or comments on each student's work were requested.

The data acquired from the two teachers were as follows: Teacher 1 provided lab activities related to insect displays. A total of 15 activities were collected, along with their scores and the feedback given, including the grading criteria used. These criteria were also used as input when prompting the AI (ChatGPT) system. Teacher 2 provided 32 activities, which involved the creation of lesson plans. These were also submitted along with the corresponding scores, feedback, and a different set of grading criteria used for evaluation.

After collecting all the necessary data from the teachers, we began prompting the activities into the AI using the approved and validated prompt. The student activities provided by the teachers, along with their respective grading criteria, were inputted into the ChatGPT, and the AI was asked to provide feedback on each activity. After processing all 47 activities—previously scored and commented on by the teachers, we successfully obtained AI-generated scores and feedback for all 47-student works using the ChatGPT system. After collecting the data, the researchers proceeded with a semi-structured interview to address the strengths and limitations of ChatGPT. We interviewed the involved science faculty members. The interview consisted of seven questions, all aimed at gaining insights into the use of AI in their field of work, whether it supports their own learning, assists in grading students, and their views on the reliability and overall considerations of using AI in education. With all these components, we were able to obtain a concrete perspective on how they view ChatGPT as an AI for assessing student works. After collecting the responses from the four science faculty members, the researchers transcribed the answers and compiled the data for thematic analysis.

Data Analysis

The quantitative data for the research were subjected to statistical analysis using a freeware statistical software named Jamovi. We used Cohen’s Kappa to get the inter-rater agreement of both the teachers and ChatGPT. To explore more the connection between the assessments of the teachers and ChatGPT, we also computed the Spearman Rho correlation. We used the Spearman Rho correlation as the quantitative data is not normally distributed.

Results and Discussion

The Extent of AI Systems Aligning with Pre-Established Grading Rubrics compared to Teachers.

The level of agreement between Teacher 1 and AI System 1 for activity in insect displays has produced a Spearman's rho measurement of 0.250, this means that there is a weak positive monotonic relationship between the two grading assessments (see Table 1). The corresponding p-value of 0.369 is greater than the conventional alpha that is 0.05, which suggests that the observed relationship between Teacher 1 and AI 1 grading doesn’t present strong evidence of agreement. On the other hand, the Cohen's Kappa measurement that assessed the inter-

rater agreement over chance for categorical data, is 0.23. Based on Landis and Koch's (1977) scale, this is only a "fair" agreement (0.21–0.40). Although not trivial, the agreement indicates a considerable amount of variety on how Teacher and AI give scores, suggesting inconsistency in categorical judgments.

Table 1. Teacher 1 and AI System 1 Cohen's Kappa and Spearman Rho (Insect Displays)

Measure	Value	Df	p-value
Spearman's rho	0.250	13	0.369
Cohen's Kappa	0.23	—	—

The findings gave a clear understanding of the degree to which AI systems align with pre-existing rubrics with those of teachers. The low agreement found between teacher scores and AI scores suggests that AI and teachers were not able to come into agreement in certain categories in giving feedback with students' activities. Although there is some level of agreement, the moderate correlation points out inconsistencies that could be due to different interpretations of the criteria from the rubrics or the AI's inability to investigate contextual and semantic refinements in student activities. Though AI can offer teachers the opportunity to make their work easier, it is not necessarily well-suited to handle the human language and logic that goes into grading more subjective activities (Wang et al., 2024). AI systems might struggle with more structured activities as AI systems cannot fully analyze it than a human teacher that would naturally factor in (Celik et al., 2022; Jie & Kamrozzaman, 2024; Memarian & Doleck, 2024). These results suggest that AI models must be made more context-sensitive through further training or that hybrid evaluation methods must be created that use AI's efficiency with teacher oversight to offer more equitable, accurate grading (Coskun & Alper, 2024; Salih et al., 2024).

Teacher 2 and AI System 2 for student activity in lesson planning present a virtually non-existent monotonic relationship, as reflected in a Spearman's rho of – 0.017, showing an essentially non-existent correlation between rankings of student scores made by the teacher and the AI system (see Table 2). Additionally, the p-value of 0.926 is significantly distant from statistical significance, further suggesting that the result obtained is likely the result of random variation and does not represent a meaningful relationship.

Table 2. Teacher 2 and AI System 2 Cohen's Kappa and Spearman Rho (Lesson Plans)

Measure	Value	Df	p-value
Spearman's rho	-0.017	30	0.926
Cohen's Kappa	0.111	—	—

The Cohen's Kappa of 0.111 indicates poor agreement between the two raters. On the Landis and Koch (1977) scale, this is in the range indicating extremely limited consistency in categorical judgments. In effect, the teacher

and the AI system disagreed rather strongly about how they were judging and classifying student work.

The findings highlight significant concern regarding the validity of the artificial intelligence system to imitate human processes of evaluation, especially in settings that require subjective judgment. The observed inconsistency may result from inherent variations in the interpretations and application of grading rubrics by the teacher and the AI (Fuller & Bixby, 2024; Taylor et al., 2024; White & Klette, 2023). The teacher, for example grading lesson plans, may consider factors like creativity, effort, or relevance to the situation—qualities the AI may not have been trained to identify or assess adequately (Ayanwale et al., 2022; Gnanaprakasam & Lourdusamy, 2024). The training data and assessment frameworks employed by the AI system may not be also aligned with the context-dependent and subjective standards typically employed by teachers. This points to the risks of implementing such systems in sensitive educational settings without human intervention. The limited correlation and agreement emphasize the requirement for widespread calibration of artificial intelligence software (Díaz-Rodríguez et al., 2023; Hradecky et al., 2022; Perifanis & Kitsios, 2023). Until these machines are properly calibrated to reflect human assessment methods, particularly in subjective areas, their deployment should be confined to assisting roles under human oversight, not as independent grading mechanisms (Ho-Dac & Martinez, 2025).

AI Systems Feedback compared to Feedback from Teachers

The teacher and AI-assisted feedback for insect collection activity focused on similar aspects, like presentation quality, specimen variety and pinning technique (see Table 3). However, each feedback has different parts of student learning. The key difference lies in how each feedback method works and the benefits they provide for assessment and student growth. Human feedback is highly effective because it notices contexts and uses sensory judgment. It can pinpoint specific problems, like missing materials (such as mothballs or scratch tape), improper spread wings, or specimens that are still moving. Manual feedback also accounts for the behaviors, such as whether submissions are on time. It is something AI currently cannot measure. Furthermore, human evaluators often use a direct and sometimes blunt approach, which may seem harsh, but gives clear and straightforward guidance. This feedback helps students understand exactly what is wrong and what they need to improve.

Table 3. Comparison of Teacher 1 and AI 1 in Insect Collection: Thematic Analysis

Themes Present in Both	Unique to Teacher 1	Unique to AI (ChatGPT)
-Presentation and Organization (well-organized display, visually appealing arrangement) **-Specimen Diversity (**presence or lack of variety) **-Pinning Technique** (accuracy and correctness of specimen mounting)	**-Missing Materials** (e.g., no mothballs, scratch tape) **-Still Moving Specimens** **-Outspread Wings** **-Late Submission** **-Blunt Tone** **-Direct Correction**	**-Scientific Naming and Data Detail** (encouragement to include full scientific names, collection dates, locations) **-Encouraging Tone and Praise** (“Great job,” “Outstanding work,” etc.) **-Focus on Rare/Uncommon Species** (Suggestion to collect rarer species)

AI-assisted feedback is strong in consistency, detail, and motivation. It helps students by encouraging them to include scientific names, accurate data, and even rare species. One of AI's unique strengths is its positive way of framing suggestions. This positive tone helps reduce anxiety and boosts student confidence. The strengths of both teacher feedback and AI suggest they work well together (Escalante et al., 2023; Malik et al., 2023). Teacher feedback is essential for handling real-world issues, observing complex behaviors, and providing guidance while AI is great at analyzing data, giving positive reinforcement, and ensuring (Lin & Chen, 2024; Lin et al., 2023. Using them together can create a feedback system where machines improve teaching and human insights enhance algorithmic assessments (Memarian & Doleck, 2023; Shum et al., 2023). Educators and institutions should think about combining these methods to offer feedback (Jin et al., 2025).

Comparison of the analysis of the lesson plan feedback of teacher 2 and AI shows similarities and differences in terms of depth of understanding (see Table 3). They recognized the main teaching strengths as well-organized lessons, appropriate scientific content with emphasis on the students. For instance, they valued the systematic order of lesson elements from introduction to conclusion and proper scientific use of motivational techniques like puzzles, games, and real-life examples. Weaknesses noted by them were transitions from one subject to another being smoother and the use of more varied assessment instruments for improved understanding of student learning. These similarities show an agreement of opinion on significant practices of teaching essential for effective learning.

Table 4. Comparison of Teacher 2 and AI 2 Feedback in Lesson Plan: Thematic Analysis

Shared Themes	**Unique to Teacher 2**	**Unique to AI**
-Formative Assessment Tools: Present but needing enhancement **-Summative Assessment Structures**: Require better structure and alignment **-Creative Motivation Activities**: Games, puzzles, real-life tasks (e.g., “2 Truths and a Lie”) - **Student-Centered Learning**: Emphasized by both **-Content-Student Connection**: Both mentions improving relevance	-Suggestions for rubrics, checklists, and oral reflections -Inclusion of metacognitive tools like journals and reflective questions -Activities deeply tied to content (e.g., dart throwing to explain principles) -Assessment of motivation’s relevance to learning outcomes -Feedback links pre-activity (e.g., PowerPoint) to student backgrounds -Emphasizes learner autonomy through reflection and exploration	-Identifies over-reliance on simple recall (e.g., T/F questions) -Recommends performance tasks but in broad, non-specific terms -Recognition of humor and cultural references (e.g., vinegar, memes) -Encouragement for developmentally appropriate tasks -Praises relatable items like scavenger hunts or introductory demos -Focuses on accessibility and developmental fit, less on deep alignment to learner needs

AI feedback tends to recognize correctness of content but provides generic feedback on how to improve without

pointing out particular gaps in concepts or student confusion. It reflects a weakness in AI to deal with complex subject matter. While AI feedback identifies the use of SMART goals and interactive features, it also tends to comment on vague verbs like "explain" without recommending functional substitutes. Teacher feedback, on the other hand, judges the alignment of these goals with Bloom's Taxonomy and the way support develops at different levels of understanding. Teacher feedback provided explicit solutions in the form of rubrics, checklists, short quizzes, and reflective methods like journals and oral reflection. It highlighted ensuring alignment of formative and summative assessment to learning objectives and instructional strategies.

Regarding relevance and student-centeredness, AI feedback focused on accessibility, while teacher feedback emphasized student autonomy, reflection, and incorporating students' past experiences. It suggested connections between pre-activities and students' backgrounds and included reflective moments that encourage ownership of learning. It highlights the teaching insight that human evaluators offer in assessing how lessons support diverse learning needs. In conclusion, while AI feedback provides a structured and efficient evaluation, teacher feedback offers a more thoughtful and pedagogical sound analysis (Escalante et al., 2023; Memarian & Doleck, 2023; Pang et al., 2024).

AI as Used in Grading Students' Activities as Perceived by Teachers

Table 5. AI Systems in Providing Feedback and Grades as Perceived by Teachers

Theme	Code
Usefulness and Efficiency	AI facilitates faster task completion and easier access to information or evaluation. Use of AI to design materials, activities, or assess outputs.
Limitations and Misalignment	AI feedback may produce irrelevant, incorrect, or misaligned suggestions. AI mistakenly labels original work as AI-generated or plagiarized.
Human vs. AI Feedback	Teacher feedback includes compassion, context, and consideration of student conditions. Preference for using both AI and teacher input for better outcomes,
Improvement and Suggestions	The importance of crafting accurate prompts to get effective AI feedback. Need for AI to align feedback with real-world rubrics or learning standards.

Teachers expressed a strong appreciation for the usefulness and efficiency of AI in supporting their professional responsibilities specifically grading student outputs. Many respondents emphasized how AI tools enhanced their productivity by assisting in the generation of instructional materials and feedback. A strong appreciation for the usefulness and efficiency of AI was seen in the answers of the teachers, which resonates about AI supporting professional responsibilities. The respondents echo how AI tools enhanced their productivity. One respondent

noted,

"it helped me improve my work efficiently" (Respondent 1).

While another stated,

"it makes me feel convenient in my part" (Respondent 2).

These insights are in line with findings from Zawacki-Richter et al. (2019) who observed that AI systems can reduce the time teachers spend on routine tasks, allowing more focus on student-centered learning. AI was also praised for improving time management and lesson planning, as one respondent shared,

"Allows us to use our time efficiently" (Respondent 3).

Beyond efficiency, teachers described AI as a tool that supports creativity and instructional design. Respondent 1 mentioned that AI helps

"craft questions, activities for my students,"

while Respondent 5 remarked that it

"helped me construct comprehensive feedback."

Another respondent added the following highlighting its versatility,

"use AI to guide presentation" (Respondent 2)

These perspectives align with Memarian and Doleck (2023), who noted that AI facilitates content generation and can scaffold pedagogical tasks. However, researchers such as Luckin et al. (2016) caution that while AI enhances convenience, overreliance may limit teachers' professional growth and pedagogical decision-making. Despite the perceived advantages, teachers also identified several limitations and misalignments in AI-generated feedback. Key concerns involved the accuracy, contextual fit, and potential for erroneous assessments. As one teacher explained,

"output differs from expectations" (Respondent 1),

while another observed,

"AI did not give accurate information" (Respondent 2).

A third respondent reported,

"emphasized aspects not aligned with rubric" (Respondent 4),

indicating that AI feedback may fail to reflect instructional goals or assessment standards. These findings reinforce critiques by Ng et al. (2023), who argues that AI lacks the situational awareness necessary to interpret learning outcomes meaningfully. Furthermore, some participants reported negative experiences with plagiarism detection systems, which incorrectly flagged original work. For example, Respondent 3 stated,

"AI may mistakenly identify your work as AI-generated,"

and added,

"detected paraphrased words as plagiarized."

This aligns with concerns raised by Zhai et al. (2024) and Wang and Tahir (2020), who warned that algorithmic approaches can misclassify content, creating distrust and anxiety among students and teachers. Such incidents highlight the need for human oversight and validation in AI-assisted evaluation processes. In discussing the differences between human and AI feedback, respondents consistently favored the human approach for its contextual and emotional richness. Teachers valued feedback that is empathetic, student-sensitive, and aligned

with professional philosophy. For example, Respondent 1 expressed that

"teacher feedback is coupled with emotions"

and

"considers student sensitivity, struggles, and needs."

Likewise, Respondent 4 noted that

"teacher feedback includes teaching philosophy,"

suggesting that educators view their feedback as deeply connected to their values and the socio-emotional context of learners. These findings echo Henderson et al.'s (2019) assertions that feedback is most effective when it considers both cognitive and affective dimensions. While some respondents acknowledged AI's efficiency, they advocated for a blended model. Respondent 5 recommended the

"co-intelligence of teachers and AI,"

while Respondent 3 stated,

"get feedback from both,"

and Respondent 2 acknowledged,

"both… but there are limitations in using technology."

This co-intelligence model is supported by Luckin et al. (2016) who argue that AI should augment human insight rather than substitute it. Teachers see value in combining the scalability and speed of AI with the human capacity for empathy, contextualization, and moral judgment.

Teachers offered recommendations for improving AI systems, primarily focusing on prompt engineering and rubric alignment. Several noted that the effectiveness of AI feedback depends on how well users phrase their inputs. As Respondent 1 advised,

"we should learn to give the correct prompt,"

while Respondent 5 emphasized,

"be more specific in prompting."

These views align with Zhai et al. (2024), who highlights prompt literacy as an emerging skill in AI-integrated classrooms. Furthermore, respondents suggested enhancements in AI customization to align with educational standards. For instance, Respondent 4 proposed that

"customizable rubric input would be helpful,"

and Respondent 5 recommended that AI should

"consider patterns of student performance."

These insights reflect a broader need for context-aware and flexible AI systems capable of adapting to authentic classroom settings and performance data. Overall, these suggestions indicate that while teachers welcome AI, they advocate for systems that are more responsive, transparent, and aligned with real-world educational frameworks.

Conclusion

We conclude that a weak AI alignment was seen in Spearman Rho measurement with basic rubric components

and also strained with complex interpretive elements. Cohen's Kappa results indicated low agreement between AI and teacher grading, highlighting discrepancies in nuanced assessment. In the end, teacher judgment remains essential for evaluating higher-order thinking and contextual depth in student work. Furthermore, we explored that while both teacher and AI feedback referred to underlying instructional elements, teacher feedback posed richer, more personalized insights rooted in instructional proficiency. In contrast, AI responses remained universal and lacked depth, context sensitivity, and educational nuance essential for meaningful improvement.

We also conclude that teachers recognized the value of AI in enhancing efficiency, instructional support, and feedback generation, yet remained cautious of its limitations in accuracy, emotional depth, and contextual alignment. While the teachers appreciated AI's role in routine tasks, they emphasized the irreplaceable value of human insight, empathy, and professional judgment in educational feedback. At the end, the teachers advocated for a blended, co-intelligent model where AI supports, but does not replace teacher expertise, guided by better prompt design and customizable systems aligned with classroom truths.

Recommendations

Based on our findings, we recommend a thoughtful integration of AI through a co-intelligent feedback model where AI supports but does not replace teacher judgment. Teachers should remain the primary assessors of tasks requiring higher-order thinking, interpretation, and contextual insight, as AI showed weak alignment with complex rubric components. To improve AI feedback relevance, we suggest developing context-aware prompting frameworks that allow teachers to customize prompts and align AI outputs with classroom realities. AI systems must be designed to adapt to subject-specific and instructional goals rather than offering generic responses.

We also recommend sustained professional development to enhance teachers' AI literacy, enabling them to critically interpret and ethically use AI-generated feedback. Training should include understanding metrics like Spearman Rho and Cohen's Kappa to assess alignment between AI and human grading. Ultimately, we see AI as a valuable tool for improving efficiency, but the depth, empathy, and insight of teachers remain irreplaceable in meaningful student assessment.

References

Adarkwah, M. A. (2021). The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature. *SN Social Sciences, 1*(3). https://doi.org/10.1007/s43545-021-00086-w

Ajogbeje, O. J. (2023). Enhancing Classroom Learning Outcomes: The Power of Immediate Feedback Strategy. *International Journal of Disabilities Sports & Health Sciences,* 6(3), 453–465. https://doi.org/10.33438/ijdshs.1323080

Aldrup, K., Carstensen, B., & Klusmann, U. (2022). Is Empathy the Key to Effective Teaching? A Systematic Review of Its Association with Teacher-Student Interactions and Student Outcomes. *Educational Psychology Review, 34*(3), 1177–1216. https://doi.org/10.1007/s10648-021-09649-y

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2023). The effects of artificial intelligence

applications in educational settings: Challenges and strategies. *Technological Forecasting and Social Change, 199*, 123076. https://doi.org/10.1016/j.techfore.2023.123076

Al-Jahwari, M., & Yousif, M. J. (2025). The impact of AI tools on Education: ChatGPT in focus. *Artificial Intelligence & Robotics Development Journal, 4*(4), 314–336. https://doi.org/10.52098/airdj.20244430

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open, 6,* 100173. https://doi.org/10.1016/j.caeo.2024.100173

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon, 10*(9), e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Atasoy, A., & Arani, S. M. N. (2025). ChatGPT: A reliable assistant for the evaluation of students' written texts? *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13553-1

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education Artificial Intelligence, 6,* 100226. https://doi.org/10.1016/j.caeai.2024.100226

Ayanwale, M. A., Sanusi, I. T., Adelana, O. P., Aruleba, K. D., & Oyelere, S. S. (2022). Teachers' readiness and intention to teach artificial intelligence in schools. *Computers and Education Artificial Intelligence, 3*, 100099. https://doi.org/10.1016/j.caeai.2022.100099

Bedizel, N. R. T. (2023). Evolving landscape of artificial intelligence (AI) and assessment in education: A bibliometric analysis. *International Journal of Assessment Tools in Education, 10*(Special Issue), 208–223. https://doi.org/10.21449/ijate.1369290

Bolkan, S. (2015). The importance of instructor clarity and its effect on student learning: Facilitating elaboration by reducing cognitive load. *Communication Reports, 29*(3), 152–162. https://doi.org/10.1080/08934215.2015.1067708

Bouziane, K., & Bouziane, A. (2024). AI versus human effectiveness in essay evaluation. *Discover Education, 3*(1). https://doi.org/10.1007/s44217-024-00320-6

Bulut, O., & Beiting-Parrish, M. (2024). The rise of artificial intelligence in educational measurement: opportunities and ethical challenges. *Chinese/English Journal of Educational Measurement and Evaluation, 5*(3). https://doi.org/10.59863/miql7785

Câmpean, A., Bocoș, M., Roman, A., Rad, D., Crișan, C., Maier, M., Tăușan-Crișan, L., Triff, Z., Triff, D., Mara, D., Mara, E., Răduț-Taciu, R., Todor, I., Baciu, C., Neacșu, M., Dumitru, I., Colareza, C. C., & Roman, C. E. (2024). Examining teachers' perception on the impact of positive feedback on school students. *Education Sciences, 14*(3), 257. https://doi.org/10.3390/educsci14030257

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends, 66*(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00411-8

Chichekian, T., & Benteux, B. (2022). The potential of learning with (and not from) artificial intelligence in education. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.903051

Coskun, T. K., & Alper, A. (2024). Evaluating the evaluators: a comparative study of AI and Teacher Assessments in Higher Education. *Digital Education Review, 45,* 124–140. https://doi.org/10.1344/der.2024.45.124-140

Delello, J. A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., & Giuseppe, T. D. (2025). AI in the Classroom: Insights from Educators on Usage, Challenges, and Mental Health. *Education Sciences, 15*(2), 113–113. https://doi.org/10.3390/educsci15020113

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., De Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion, 99,* 101896. https://doi.org/10.1016/j.inffus.2023.101896

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00425-2

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120. https://doi.org/10.3102/0034654312474350

Frenzel, A. C., Daniels, L., & Burić, I. (2021). Teacher emotions in the classroom and their implications for students. *Educational Psychologist, 56*(4), 250–264. https://doi.org/10.1080/00461520.2021.1985501

Fuller, L. P., & Bixby, C. (2024). The theoretical and practical implications of OpenAI System Rubric assessment and feedback on higher Education written assignments. *American Journal of Educational Research, 12*(4), 147–158. https://doi.org/10.12691/education-12-4-4

Gayed, J. M. (2025). Educators' perspective on artificial intelligence: equity, preparedness, and development. *Cogent Education, 12*(1). https://doi.org/10.1080/2331186x.2024.2447169

Gligorea, I., Cioca, M., Oancea, R., Gorski, A., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature review. *Education Sciences, 13*(12), 1216. https://doi.org/10.3390/educsci13121216

Gnanaprakasam, J., & Lourdusamy, R. (2024). The role of AI in Automating grading: Enhancing feedback and efficiency. In *Artificial intelligence*. https://doi.org/10.5772/intechopen.1005025

Gobrecht, A., Tuma, F., Möller, M., Zöller, T., Zakhvatkin, M., Wuttig, A., Sommerfeldt, H., & Schütt, S. (2024). *Beyond human subjectivity and error: a novel AI grading system.* ArXiv.org. https://arxiv.org/abs/2405.04323

Gunawardena, M., Bishop, P., & Aviruppola, K. (2023). Personalized learning: The simple, the complicated, the complex and the chaotic. *Teaching and Teacher Education, 139,* 104429. https://doi.org/10.1016/j.tate.2023.104429

Guo, W., Wang, J., Li, N., & Wang, L. (2025). The impact of teacher emotional support on learning engagement among college students mediated by academic self-efficacy and academic resilience. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-88187-x

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research & Development, 38*(7), 1401–1416. https://doi.org/10.1080/07294360.2019.1657807

Herman, J., & Cook, L. (2019). Fairness in classroom assessment. In *Routledge eBooks* (pp. 243–264).

https://doi.org/10.4324/9780429507533-14

Ho-Dac, M., & Martinez, B. (2025). *Human Oversight of Artificial Intelligence and Technical Standardisation*. https://doi.org/10.2139/ssrn.5228774

Hradecky, D., Kennell, J., Cai, W., & Davidson, R. (2022). Organizational readiness to adopt artificial intelligence in the exhibition sector in Western Europe. *International Journal of Information Management, 65*, 102497. https://doi.org/10.1016/j.ijinfomgt.2022.102497

Jacobsen, L. J., & Weber, K. E. (2025). The Promises and Pitfalls of large language models as Feedback Providers: A study of prompt engineering and the Quality of AI-Driven Feedback. *AI, 6*(2), 35. https://doi.org/10.3390/ai6020035

Jie, A. L. X., & Kamrozzaman, N. A. (2024). The Challenges of Higher Education Students Face in Using Artificial Intelligence (AI) against Their Learning Experiences. *Open Journal of Social Sciences, 12*(10), 362–387. https://doi.org/10.4236/jss.2024.1210025

Jin, F. J., Dai, W., Maheshi, B., Martinez-Maldonado, R., Gašević, D., & Tsai, Y. (2025). Feedback in K-12 and higher education: Educators' perspectives. *Teaching and Teacher Education, 156*, 104933. https://doi.org/10.1016/j.tate.2025.104933

Jones-Jang, S. M., Chung, M., Choi, J., Kim, N., & Lee, S. (2025). Fairness perceptions of AI in grading systems: Examining how discontent with the status quo and outcome favorability reduce AI reluctance. *Computers and Education Artificial Intelligence*, 100419. https://doi.org/10.1016/j.caeai.2025.100419

Kemal, S., & Liman-Kaban, A. (2024, November 18). *Comparative analysis of human graders and AI in Assessing Secondary School EFL Journal writing*. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/796?utm_source=chatgpt.com

Kotsis, K. T. (2024). Significance of experiments in inquiry-based science teaching. *European Journal of Education and Pedagogy, 5*(2), 86–92. https://doi.org/10.24018/ejedu.2024.5.2.815

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00426-1

Landis J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lawasi, M. C., Rohman, V. A., & Shoreamanis, M. (2024). The use of AI in improving student's critical thinking skills. *Proceedings Series on Social Sciences & Humanities, 18,* 366–370. https://doi.org/10.30595/pssh.v18i.1279

Lin, C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments, 10*(1). https://doi.org/10.1186/s40561-023-00260-y

Lin, H., & Chen, Q. (2024). Artificial intelligence (AI) -integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. *BMC Psychology, 12*(1). https://doi.org/10.1186/s40359-024-01979-0

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education.* Pearson. https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf

Magtalas, S. A. (2024). Teacher's Workload in Relation to Burnout and Work Performance. *International Journal of Multidisciplinary Applied Business and Education Research, 5*(10), 4111–4123. https://doi.org/10.11594/ijmaber.05.10.24

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki, N. (2023). Exploring Artificial Intelligence in Academic Essay: Higher Education Student's perspective. *International Journal of Educational Research Open, 5,* 100296. https://doi.org/10.1016/j.ijedro.2023.100296

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*(3), 245–256. https://doi.org/10.1177/0004944116664618

Memarian, B., & Doleck, T. (2023). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior Artificial Humans, 2*(1), 100040. https://doi.org/10.1016/j.chbah.2023.100040

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., Nacke, L. E., & Hui, P. (2023). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior Artificial Humans, 2*(1), 100027. https://doi.org/10.1016/j.chbah.2023.100027

Morrison, L., & Jacobsen, M. (2023). The role of feedback in building teaching presence and student self-regulation in online learning. *Social Sciences & Humanities Open, 7*(1), 100503. https://doi.org/10.1016/j.ssaho.2023.100503

Myint, P. Y. W., Lo, S. L., & Zhang, Y. (2024). Harnessing the power of AI-Instructor Collaborative Grading Approach: Topic-Based effective grading for Semi Open-Ended Multipart questions. *Computers and Education Artificial Intelligence*, 100339. https://doi.org/10.1016/j.caeai.2024.100339

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development, 71*(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6

Nikitina, I., & Ishchenko, T. (2024). The impact of AI on teachers: support or replacement? *Academia Polonica., 65*(4), 93–99. https://doi.org/10.23856/6511

O'Neil, H., Chuang, S., & Huang, T. (2010). Instructional system provided feedback. In *Elsevier eBooks* (pp. 226–230). https://doi.org/10.1016/b978-0-08-044894-7.00312-2

Pang, T. Y., Kootsookos, A., & Cheng, C. (2024). Artificial Intelligence use in Feedback: A Qualitative analysis. *Journal of University Teaching and Learning Practice, 21*(06). https://doi.org/10.53761/40wmcj98

Perifanis, N., & Kitsios, F. (2023). Investigating the Influence of Artificial intelligence on business Value in the digital Era of Strategy: a literature review. *Information, 14*(2), 85. https://doi.org/10.3390/info14020085

Petersen, T. (2024, September 17). *Students prefer teacher feedback over AI feedback, research finds.* Phys.org. https://phys.org/news/2024-09-students-teacher-feedback-ai.html

Ragolane, M., Patel, S., & Salikram, P. (2024). AI versus Human Graders: Assessing the role of large language models in Higher education. *Asian Journal of Education and Social Studies, 50*(10), 244–263. https://doi.org/10.9734/ajess/2024/v50i101616

Rahman, M. H., Amin, M. B., Yusof, M. F., Islam, M. A., & Afrin, S. (2024). Influence of teachers' emotional intelligence on students' motivation for academic learning: an empirical study on university students of Bangladesh. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186x.2024.2327752

Revell, T., Yeadon, W., Cahilly-Bretzin, G., Clarke, I., Manning, G., Jones, J., Mulley, C., Pascual, R. J., Bradley, N., Thomas, D., & Leneghan, F. (2024). ChatGPT versus human essayists: an exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities. *International Journal for Educational Integrity, 20*(1). https://doi.org/10.1007/s40979-024-00161-8

Rochera, M. J., Engel, A., & Coll, C. (2021). The effects of teacher' feedback. *Revista De Educación a Distancia (RED), 21*(67). https://doi.org/10.6018/red.476901

Rowe, A. D. (2016). Feelings about feedback: The role of emotions in Assessment for learning. In *The enabling power of assessment* (pp. 159–172). https://doi.org/10.1007/978-981-10-3045-1_11

Ruwe, T., & Mayweg-Paus, E. (2023). "Your argumentation is good", says the AI vs humans – The role of feedback providers and personalised language for feedback effectiveness. *Computers and Education Artificial Intelligence, 5*, 100189. https://doi.org/10.1016/j.caeai.2023.100189

Salih, S., Husain, O., Hamdan, M., Abdelsalam, S., Elshafie, H., & Motwakel, A. (2024). Transforming Education with AI: A Systematic Review of ChatGPT's Role in Learning, Academic Practices, and Institutional Adoption. *Results in Engineering*, 103837. https://doi.org/10.1016/j.rineng.2024.103837

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1). https://doi.org/10.1186/s41239-021-00292-9

Sebler, K., Bewersdorff, A., Nerdel, C., & Kasneci, E. (2025, February 18). *Towards Adaptive Feedback with AI: Comparing the Feedback Quality of LLMs and Teachers on Experimentation Protocols*. arXiv.org. https://arxiv.org/abs/2502.12842

Shum, S. B., Lim, L., Boud, D., Bearman, M., & Dawson, P. (2023). A comparative analysis of the skilled use of automated feedback tools through the lens of teacher feedback literacy. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00410-9

Singh, T. M., Reddy, C. K. K., Murthy, B. V. R., Nag, A., & Doss, S. (2024). AI and Education. In *Advances in educational technologies and instructional design book series* (pp. 131–160). https://doi.org/10.4018/979-8-3693-8151-9.ch005

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education Artificial Intelligence, 3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Taylor, B., Kisby, F., & Reedy, A. (2024). Rubrics in higher education: an exploration of undergraduate students' understanding and perspectives. *Assessment & Evaluation in Higher Education, 49*(6), 799–809. https://doi.org/10.1080/02602938.2023.2299330

Terrazas-Arellanes, F. E., Strycker, L., Alvez, G. G., Miller, B., & Vargas, K. (2025). Promoting Agency Among Upper Elementary School Teachers and Students with an Artificial Intelligence Machine Learning System to Score Performance-Based Science Assessments. *Education Sciences, 15*(1), 54. https://doi.org/10.3390/educsci15010054

Thompson, Z., Yoon, H., & Booth, P. (2023). Dispersed assessment: A novel approach to enhancing student engagement during and beyond Covid-19. *The International Journal of Management Education, 21*(2), 100811. https://doi.org/10.1016/j.ijme.2023.100811

Tierney, R. D. (2016). Fairness in educational assessment. In *Springer eBooks* (pp. 1–6).

https://doi.org/10.1007/978-981-287-532-7_400-1

Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-024-00501-1

Trigo, A., Stein, N., & Belfo, F. P. (2024). Strategies to improve fairness in artificial intelligence: A systematic literature review. *Education for Information, 40*(3), 323–346. https://doi.org/10.3233/efi-240045

Trikoili, A., Georgiou, D., Pappa, C. I., & Pittich, D. (2025). *Critical Thinking Assessment in Higher Education: A Mixed-Methods Comparative Analysis of AI and human Evaluator.* Taylor & Francis. https://www.tandfonline.com/doi/full/10.1080/10447318.2025.2499164

Venter, J., Coetzee, S. A., & Schmulian, A. (2024). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. *Assessment & Evaluation in Higher Education,* 1–21. https://doi.org/10.1080/02602938.2024.2415649

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences, 15*(3), 343. https://doi.org/10.3390/educsci15030343

Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education, 149,* 103818. https://doi.org/10.1016/j.compedu.2020.103818

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems With Applications, 252,* 124167. https://doi.org/10.1016/j.eswa.2024.124167

Wang, X., Zhang, L., Peng, Y., Lu, J., Huang, Y., & Chen, W. (2021). Development and validation of the empathy scale for teachers (EST). *Studies in Educational Evaluation, 72,* 101112. https://doi.org/10.1016/j.stueduc.2021.101112

Wang, Z. (2025). The Impact of Teacher Feedback on Student Motivation in Online Learning Environments: A Study Based on Self-Determination Theory. *Journal of Education, Humanities, and Social Research*, *2*(2), 13-27.

White, M., & Klette, K. (2023). What's in a score? Problematizing interpretations of observation scores. *Studies in Educational Evaluation, 77*, 101238. https://doi.org/10.1016/j.stueduc.2023.101238

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1). https://doi.org/10.1186/s41239-019-0171-0

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments, 11*(1). https://doi.org/10.1186/s40561-024-00316-7

Zhang, A., Gao, Y., Suraworachet, W., Nazaretsky, T., & Cukurova, M. (2025). *Evaluating Trust in AI, Human, and Co-produced Feedback Among Undergraduate Students.* ArXiv.org. https://arxiv.org/abs/2504.10961

Zhang, Z., & Hyland, K. (2021). Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51,* 100586. https://doi.org/10.1016/j.asw.2021.100586

Zheng, F. (2022). Fostering Students' Well-Being: The mediating role of teacher interpersonal behavior and Student-Teacher Relationships. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.796728

Author Information

Marjorie Joy Mariano
https://orcid.org/0009-0009-4588-152X
Rizal Technological University
Philippines

Javilyn Joy Taborete
https://orcid.org/0009-0005-9068-1499
Rizal Technological University
Philippines

Andrew Gampal
https://orcid.org/0009-0001-4193-6791
Rizal Technological University
Philippines

Aragorn Dawisan
https://orcid.org/0009-0003-9032-2088
Rizal Technological University
Philippines

Mary Grace Maraganas
https://orcid.org/0009-0009-7213-516X
Rizal Technological University
Philippines

Rizaldy Garcia
https://orcid.org/0000-0001-9208-8975
Rizal Technological University
Philippines
Contact e-mail: *regarcia@rtu.edu.ph*