Asian Journal of Assessment in Teaching and Learning
*Vol 16, Issue 1, 2026 (68-79) /eISSN **2821-2916***

AI as Rubric Mediator: A Reflective Practice Brief on GPT-Supported Assessment for Learning in Music Education

Sajastanah Imam Koning[1*], Chamil Arkhasa Nikko Mazlan[2]
[1]Faculty of Music and Performing Arts, Sultan Idris Education University,
35900 Tanjong Malim, Perak, Malaysia
[2]Faculty of Music, National Academy of Arts, Culture and Heritage,
464, Jalan Tun Ismail, 50480, Kuala Lumpur, Malaysia

*Corresponding author: chamil@fmsp.upsi.edu.my

Received: 4 June 2026; **Revised**: 19 June 2026
Accepted: 22 June 2026; **Published**: 23 June 2026

To link to this article: https://doi.org/10.37134/ajatel.vol16.1.6.2026

Abstract

Generative artificial intelligence (GenAI) has intensified debates about academic integrity, authorship, assessment validity, and the future of teacher judgement in higher education. Yet the discussion should not be reduced to whether AI should be prohibited or accepted. A more educationally productive question is how AI may be designed and bounded to support assessment for learning, feedback literacy, evaluative judgement, and ethical self-assessment. This brief communication presents a reflective practice account from AMU60104 Assessment in Music Education Research, a master's coursework course involving 18 consenting students, most of whom were practising music teachers in primary, secondary, or private music education settings. A customised GPT, AMU60104 Assessment & Feedback Assistant (M251PM2), was introduced to support students' self-assessment after they had posted weekly reflections in the learning management system. Students were guided to use the tool to understand rubrics, identify strengths and areas for improvement, and strengthen future reflective writing, while being explicitly instructed not to use AI to generate their reflections. Interpreted through feedback literacy, evaluative judgement, relational validity, and situated AI ethics, the practice suggests that AI-supported assessment is educationally defensible only when it preserves student authorship, strengthens human judgement, and deepens assessment understanding. As a reflective practice account, the paper focuses on students' perceptions and experiences of using GPT-supported feedback rather than evaluating the content of individual prompts, AI-generated responses, or the technical accuracy of the feedback provided. The paper argues that the value of AI in assessment lies not in replacing teachers or automating judgement, but in mediating students' engagement with criteria, evidence, feedback, and ethical responsibility.

Keywords: *Generative AI, Assessment for Learning, Feedback Literacy, Evaluative Judgement, Self-Assessment, Music Education, Rubric-Mediated Feedback*

INTRODUCTION

The emergence of generative artificial intelligence (GenAI) has unsettled conventional assumptions about assessment in higher education. Since GenAI systems can produce fluent text, generate explanations, summarise materials, and simulate feedback, educators are increasingly required to

reconsider the relationship between student authorship, academic integrity, assessment validity, and teacher professional judgement. Much of the institutional conversation has focused on detection, prohibition, and the prevention of misuse. Recent assessment scholarship also suggests that the challenge of GenAI should not be reduced to detection alone, because teachers' ability to identify AI-generated work is shaped by institutional, behavioural, contextual, and workload-related factors as much as by the technical sophistication of AI systems (Niloy & Huda, 2025). These concerns are legitimate because assessment loses meaning when submitted work no longer reflects the learner's own understanding, effort, or judgement.

However, a purely defensive response may also narrow the educational imagination. If AI is framed only as a threat, educators may overlook its potential to support assessment for learning, particularly when AI is used not to generate student work, but to help students understand criteria, interpret feedback, and improve future learning. Authentic assessment has therefore become increasingly important because it shifts attention from policing final products toward designing meaningful tasks that require contextual judgement, personal engagement, and evidence of learning (Gedera, 2023). Assessment for learning has long emphasised that assessment should not merely certify achievement but support learning while it is still developing (Black & Wiliam, 1998; Wiliam, 2011). This view is closely related to formative assessment, where feedback helps learners recognise the gap between current and desired performance and take action to improve (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Yet feedback does not automatically lead to learning. Students must understand, evaluate, and use feedback meaningfully. This has led to growing interest in student feedback literacy, defined as the capacities and dispositions needed to make sense of feedback and use it for improvement (Carless & Boud, 2018). In the age of GenAI, feedback literacy must also include the ability to judge AI-generated feedback, because students may increasingly encounter machine-produced comments, suggestions, and evaluations.

The concept of evaluative judgement is equally important. Tai et al. (2018) define evaluative judgement as the capability to make decisions about the quality of one's own work and the work of others. This capability is central to higher education because graduates must eventually make quality judgements beyond the immediate guidance of teachers. Bearman et al. (2024) extend this argument to GenAI, arguing that students now need to develop evaluative judgement not only of human work, but also of AI outputs and AI-supported processes. This is especially relevant in assessment contexts where students may be tempted to treat AI feedback as authoritative. The educational challenge is therefore not simply to allow or ban AI, but to design assessment conditions in which students remain active judges of quality.

This brief communication responds to that challenge through a reflective practice account from AMU60104 Assessment in Music Education Research, a master's coursework course. The course focuses on theories, methods, tools, data interpretation, reliability, validity, ethics, and implications of assessment in music learning. This context is significant because music education assessment involves more than technical measurement. It requires situated judgement about musical behaviours, performance, creativity, classroom evidence, assessment fairness, and pedagogical implications. In such contexts, rubrics may support transparency, but they do not remove the need for interpretation. Students must still learn how to connect assessment criteria with authentic teaching examples and professional reasoning.

Within this course, a customised GPT named AMU60104 Assessment & Feedback Assistant (M251PM2) was introduced to support students' self-assessment in Assignment 1, Weekly Reflections. The tool was designed to provide constructive academic feedback based on detailed assignment rubrics. Students used the tool after posting their weekly reflections in the learning management system. Its purpose was to help them understand the rubric, identify strengths and areas for development, and improve subsequent reflective writing. Importantly, students were explicitly instructed not to use GPT to generate their weekly reflections. This ethical boundary is central to the argument of the paper. The GPT was not positioned as a substitute author, autonomous marker, or replacement for lecturer judgement. It was positioned as a rubric-mediated feedback companion.

This paper therefore argues that the educational value of GenAI in assessment lies not in automation, but in mediation. When carefully designed, AI can mediate students' engagement with rubrics, criteria, feedback, and self-assessment. However, such mediation is only defensible if it strengthens rather than weakens human judgement. The paper contributes a critical-reflective account

of how AI-supported assessment may be aligned with assessment for learning, feedback literacy, evaluative judgement, relational validity, and situated AI ethics.

CONCEPTUAL POSITIONING

This brief communication is positioned at the intersection of assessment for learning, feedback literacy, evaluative judgement, relational validity, and situated AI ethics. These perspectives are used not to construct a full theoretical model, but to frame how GPT-supported self-assessment may be understood as a pedagogical and ethical assessment practice. The conceptual positioning is therefore deliberately selective: it focuses on the assessment concepts most relevant to the reflective practice described in this paper, particularly how students interpret criteria, engage with feedback, exercise judgement, and maintain responsibility when using AI-supported feedback.

1. Assessment for Learning and Formative Feedback

Assessment for learning provides the first conceptual anchor for this brief communication. Black and Wiliam's (1998) seminal work established that assessment can improve learning when evidence of student understanding is used to adapt teaching and support learner progress. Later work continued to emphasise that assessment for learning depends on feedback, classroom dialogue, learner involvement, and the use of evidence to inform next steps (Black & Wiliam, 2018; Wiliam, 2011). This is important because AI-supported assessment should not be evaluated only according to speed or efficiency. The more important question is whether AI helps learners understand quality and improve future work.

Feedback research further clarifies this point. Hattie and Timperley (2007) argue that effective feedback addresses where learners are going, how they are progressing, and what they need to do next. Nicol and Macfarlane-Dick (2006) similarly propose that good feedback supports self-regulated learning by clarifying standards, facilitating self-assessment, and encouraging dialogue. This also resonates with broader university assessment scholarship, which emphasises that effective assessment should make standards visible, support learner participation, and create opportunities for students to use feedback for improvement (Carless, 2015). Adarkwah's (2021) narrative review reinforces that feedback is powerful when it is specific, timely, supportive, constructive, and integrated into teaching and learning. Lian Zhao and Ting (2025) also show that supervisory written feedback can function as formative scaffolding, helping students move closer to academic writing conventions across successive drafts.

These perspectives matter because GPT-generated feedback may appear useful due to its immediacy, clarity, and accessibility, but its educational value depends on how students use it. If students receive AI feedback passively, the process may reproduce dependence. If they use AI feedback to compare their work against criteria, identify gaps, and plan improvement, then AI may contribute to assessment for learning.

2. Feedback Literacy and Evaluative Judgement

Feedback literacy offers a second conceptual lens. Carless and Boud (2018) argue that students need to appreciate feedback, make judgements, manage affect, and take action. This reframes feedback as a learner capability rather than a teacher transmission. Recent work also suggests that feedback literacy develops through repeated, purposeful, and coherent practice in which learners seek, generate, process, and act on feedback (Wu et al., 2025). Peer evaluation can also strengthen feedback literacy by requiring students to judge quality and provide constructive comments, not merely receive feedback from authority figures (Song et al., 2025).

Evaluative judgement deepens this argument. Tai et al. (2018) define evaluative judgement as the ability to make informed decisions about quality. In assessment practice, this means students must learn to interpret rubrics, compare examples, recognise evidence, and make reasoned decisions about what counts as good work. Bearman et al. (2024) argue that this capability becomes even more urgent in the age of GenAI because students must evaluate AI outputs and AI-mediated processes. Self-assessment is also relevant here because students need opportunities to monitor the quality of their own

work, interpret standards, and develop confidence in making informed judgements about improvement (Panadero et al., 2016). GenAI does not remove the need for judgement; rather, it increases the need for judgement because students must decide whether AI suggestions are accurate, relevant, ethical, and contextually appropriate. This concern is consistent with Sadler's (2010) argument that integrity in assessment depends not only on grading procedures, but also on the fidelity with which standards, evidence, and quality judgements are interpreted.

This is the central theoretical move in this paper. The AMU60104 Assessment & Feedback Assistant (M251PM2) is not treated as a tool that improves learning by itself. It is interpreted as a potential trigger for feedback literacy and evaluative judgement. Students had to decide whether the GPT feedback was useful, too general, contextually appropriate, aligned with the rubric, or in need of adaptation. In this sense, the pedagogical value was not simply in the feedback produced by AI, but in the judgement students exercised in response to that feedback.

3. Relational Validity and Situated AI Ethics

A third lens is relational validity. Din Eak et al. (2025) propose relational validity as a framework that extends assessment validity beyond technical accuracy toward coherence among evidence, interpretation, dialogue, ethics, fairness, and educational consequence. This is useful for AI-supported assessment because the validity question cannot be reduced to whether GPT comments are grammatically fluent or apparently correct. The more important question is whether the practice strengthens meaningful assessment relationships: between criteria and evidence, between learner and feedback, between teacher judgement and student self-assessment, and between ethical boundaries and educational consequences.

Situated AI ethics also supports this position. Gartner and Krašna (2023) argue that AI ethics in education should address autonomy, privacy, trust, and responsibility. Gaur et al. (2024) similarly emphasise that AI in education creates opportunities for personalised learning and adaptive assessment, but also raises ethical concerns about data privacy, bias, accountability, and inequality. Mazlan et al. (2026a) argue that AI ethics research is shifting from abstract principles toward situated professional judgement, human–AI collaboration, trust, and interpretive practice. This practice-based view is important because educators do not enact AI ethics only through policy statements. They enact it through everyday assessment decisions: what uses are permitted, how students are guided, how authorship is protected, and how human judgement is preserved. AI literacy is therefore not merely a technical skill, but a critical educational capability involving prompt awareness, ethical use, and the ability to evaluate AI-generated information rather than accepting it uncritically (Walter, 2024).

Finally, modern learning literature suggests that AI-supported assessment should be situated within broader transformations in learning design, digital feedback, learner agency, and professional judgement. Bizami et al. (2023) show that heutagogy, peeragogy, and cybergogy can support immersive blended learning when pedagogical principles are aligned with technological affordances. Mohamed Sapawi and Nik Yusoff (2025) further argue that AI is reshaping curriculum development through personalised, efficient, and responsive learning design, while also requiring attention to implementation challenges and strategic direction. Mazlan et al. (2025) similarly show that modern learning research is increasingly shaped by digital technologies, artificial intelligence, adaptive learning systems, competency-based education, digital literacy gaps, and ethical concerns in technology-enhanced learning.

In music education, Mazlan et al. (2024) demonstrate how bite-sized learning and TikTok can support accessible, engaging, and technology-mediated music learning within the broader orientation of Education 5.0. Within the specific context of assessment and feedback, Ngu et al. (2025) show that students' engagement with ChatGPT-generated feedback is shaped by motivation, personal learning goals, institutional policies, and assessment practices. Sankaran and Low (2025) similarly emphasise that feedback in continuous assessment should support student understanding, future improvement, and engagement, while remaining clear, constructive, and useful. Together, these studies support the proposition that AI-supported assessment is not an isolated technical issue. It is part of a wider shift in curriculum, pedagogy, digital feedback, learner agency, assessment design, and professional judgement.

4. Context of Practice

This reflective practice took place in AMU60104 Assessment in Music Education Research, a four-credit master's coursework course. The course covers assessment planning, implementation, evaluation, data interpretation, standardised music tests, reliability, validity, ethics, musical behaviours, aptitude, attitude, achievement, behavioural objectives, item construction, and implications for music learning. The students were 18 consenting master's coursework students, most of whom were practising music teachers in primary, secondary, or private music education settings. Their professional backgrounds made the reflections practice-rich because students were not only learning assessment concepts but also relating those concepts to their own teaching contexts.

Assignment 1, Weekly Reflections, carried 15% of the course assessment. Students were required to post weekly reflections in the learning management system from Week 1 to Week 7. Each reflection could be submitted as a written post of approximately 300 words or as a video reflection of three to four minutes. Each reflection needed to include a key concept from the weekly topic, a classroom example from the student's teaching, and a reflection on challenges or implications. Students were also required to reply to at least one peer each week using the Acknowledge–Extend–Question model.

The rubric assessed four components: theory–practice connection, classroom application, critical reflection, and peer replies using the AEQ model. These criteria require students to move beyond summary. A strong reflection needed to demonstrate conceptual understanding, authentic classroom application, critical evaluation of practice, and dialogic peer engagement. This made the assignment suitable for GPT-supported self-assessment because students needed to understand not only what to write, but how quality was being judged.

The AMU60104 Assessment & Feedback Assistant (M251PM2) was introduced as a customised GPT developed by the course lecturer. It was designed to evaluate coursework submissions using assignment-specific rubrics and provide clear, constructive academic feedback. For Assignment 1, students were encouraged to use the tool after posting their weekly reflections. The purpose was to help them review the quality of their reflection, understand the rubric, identify strengths and areas for development, and improve future reflections. The ethical instruction was explicit: students could use GPT to understand the rubric and improve future reflections, but they could not use GPT to generate the weekly reflections on their behalf. Their reflections had to come from their own teaching experience and learning.

5. Reflective Practice Approach

This paper adopts a reflective practice approach rather than a formal experimental or intervention design. The evidence was drawn from anonymised student reflections and voluntary feedback on the use of the AMU60104 Assessment & Feedback Assistant (M251PM2). It is important to clarify the boundary of the evidence. The study did not collect students' exact prompts, GPT response logs, or full human–AI interaction histories. Therefore, the analysis does not evaluate the technical accuracy of GPT-generated feedback or compare AI feedback against lecturer feedback. Instead, the analysis focuses on how students reflected on the usefulness, limitations, and ethical implications of using the GPT assistant as a self-assessment support. This boundary is consistent with the paper's positioning as a reflective practice brief rather than a formal validation study of an AI assessment system.

Within this boundary, the intended use of the GPT assistant was framed through a human-in-the-loop feedback protocol. Students were guided to complete and post their original weekly reflection in the learning management system before using the GPT assistant. They could then use the assistant to obtain rubric-linked feedback on the posted reflection and compare the feedback with the four assessment criteria: theory–practice connection, classroom application, critical reflection, and AEQ peer engagement. Students were encouraged to treat the AI feedback as provisional rather than authoritative by deciding whether suggestions should be accepted, adapted, questioned, or rejected. The feedback was directed toward improving subsequent reflections rather than rewriting earlier posts for grade advantage. In this way, GPT was positioned as one feedback source within a broader assessment ecology involving self-assessment, peer interaction, lecturer judgement, and students' own teaching contexts. The human-in-the-loop process is summarised in Figure 1, which illustrates how the GPT assistant was positioned as a rubric-mediated feedback support rather than as an authoring tool,

autonomous marker, or replacement for lecturer judgement.

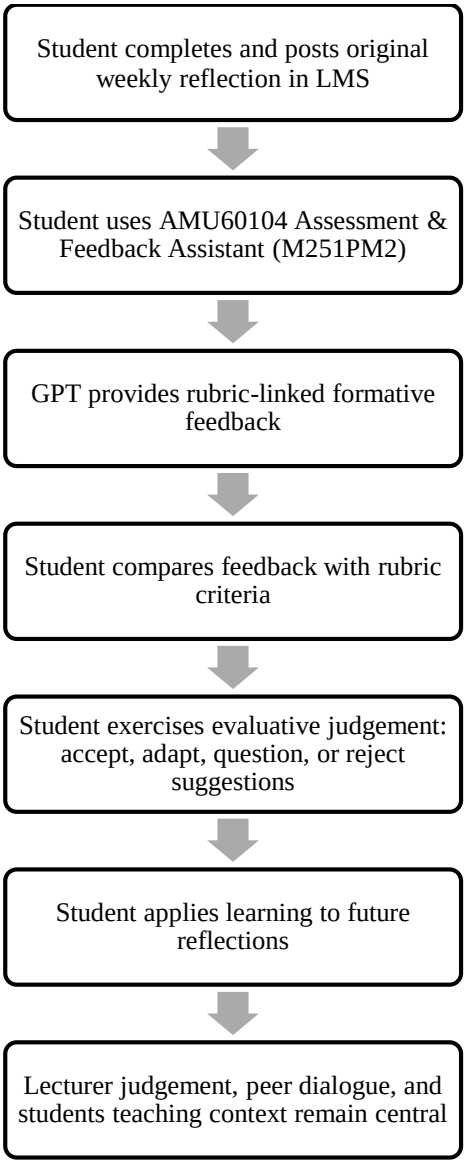

Figure 1 Rubric-mediated human-in-the-loop feedback cycle in GPT-supported self-assessment

Note The cycle illustrates the intended formative and ethical sequence of GPT use. Students first completed and posted their own reflections before using the GPT assistant for rubric-linked feedback. GPT feedback was treated as provisional and interpretive, while lecturer judgement, peer dialogue, and students' teaching contexts remained central.

Permission to use assignment-related reflections for research and publication purposes was requested separately from normal coursework participation. Students were informed that participation was voluntary and would not affect their grades. Only consenting students' reflections were considered in the reflective analysis. The purpose was not statistical generalisation. Instead, the aim was to examine the pedagogical meanings emerging from students' experiences of GPT-supported self-assessment within an authentic postgraduate music education assessment context.

The reflections were read interpretively to identify recurring patterns related to rubric understanding, feedback use, reflective writing, professional judgement, and ethical awareness. Four reflective themes were developed: GPT as rubric translator, GPT as formative feedback companion, GPT as evaluative judgement trigger, and GPT as ethical tension. To make the interpretive logic of the reflective analysis more transparent, Table 1 summarises how each theme relates to the assessment issue

it foregrounds, the possible risk or tension identified, and the design response used to preserve the formative and ethical purpose of the GPT-supported practice.

Table 1 Reflective themes, assessment meanings, and design responses in GPT-supported self-assessment

Reflective theme	Meaning in the practice	Assessment concept supported	Risk or tension	Design response
GPT as rubric translator	GPT helped students interpret rubric language such as theory–practice connection, classroom application, critical reflection, and AEQ peer engagement.	Feedback literacy; assessment for learning	Students may treat AI explanation as authoritative.	Students were guided to compare GPT feedback with the official rubric and lecturer expectations.
GPT as formative feedback companion	GPT provided immediate, constructive feedback on strengths and areas for improvement.	Formative feedback; self-regulated learning	Feedback may be generic or insufficiently music-specific.	GPT feedback was used for future improvement, not as final grading or replacement for lecturer judgement.
GPT as evaluative judgement trigger	Students had to decide whether AI suggestions were relevant, accurate, ethical, and contextually appropriate.	Evaluative judgement; AI literacy	Students may over-rely on AI or accept suggestions uncritically.	Students were encouraged to accept, adapt, question, or reject GPT feedback.
GPT as ethical tension	Students recognised benefits and risks, including authorship, dependency, accuracy, and loss of personal voice.	Situated AI ethics; relational validity	AI use may weaken authorship and assessment integrity.	The ethical boundary was explicit: GPT could support feedback understanding but not generate reflections.

Table 1 is not intended to present quantified findings or claim causal impact. Rather, it provides an organising map for the reflective discussion that follows. This approach is appropriate for a brief communication because the intention is not to overclaim causal improvement in achievement scores. The paper argues more cautiously that the practice created a pedagogical space for students to engage with assessment criteria, feedback, judgement, and ethical self-regulation. This distinction matters because AI-supported assessment should be studied critically rather than celebrated uncritically.

REFLECTIVE THEMES AND DISCUSSION

1. GPT as Rubric Translator

The first theme was GPT as rubric translator. Students reported that the tool helped them understand rubric criteria more clearly, particularly in relation to what counted as a strong reflection, how to connect theory with practice, and how to identify areas for improvement. This theme is significant because rubrics are often assumed to be transparent, yet students may still struggle to interpret what criteria mean in practice. Terms such as “critical reflection,” “theory–practice connection,” and “classroom application” require interpretation, especially when students are moving between

professional experience and academic writing.

The GPT appeared to mediate between the formal rubric and the student's understanding of quality. Rather than functioning merely as a scoring device, the rubric became more dialogic through AI-supported explanation. This aligns with feedback literacy scholarship, which argues that students must learn to interpret and use feedback rather than simply receive it (Carless & Boud, 2018). It also aligns with assessment for learning because the tool was used after submission to the forum and before future reflections, allowing feedback to inform subsequent learning rather than only justify a grade (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006).

However, the idea of GPT as rubric translator must be treated critically. Translation is not the same as judgement. AI may explain criteria in accessible language, but students still need to decide whether the explanation fits the assignment, their teaching example, and the lecturer's expectations. The educational value therefore lies in the interaction among rubric, student, AI feedback, and human judgement.

2. GPT as Formative Feedback Companion

The second theme was GPT as formative feedback companion. Students described the feedback as useful for identifying strengths, weaknesses, and possible improvements. Some valued its clarity and immediacy, while others noted that it helped them organise ideas, improve academic expression, and think more critically about their teaching practice. This reflects wider feedback research showing that useful feedback should be timely, specific, constructive, and oriented toward improvement (Adarkwah, 2021; Hattie & Timperley, 2007).

The word "companion" is important. The GPT was not positioned as the final evaluator. It provided formative support within a broader assessment ecology that included weekly prompts, lecturer-designed rubrics, peer interaction, and ethical instructions. This distinction prevents the practice from becoming a form of automated judgement. In a music education context, assessment requires attention to musical behaviour, pedagogical context, creativity, performance evidence, and classroom realities. These dimensions cannot be fully captured by an AI system. Parkes and Burrack (2020) emphasise that music classroom assessment should connect assessment processes with learning outcomes and teaching improvement. Therefore, GPT feedback is useful only when it remains connected to real teaching evidence and human interpretation.

This theme also resonates with modern learning design. Bizami et al. (2023) argue that immersive blended learning requires alignment between technological affordances and pedagogical principles such as self-determined, peer-supported, and digitally mediated learning. In the present practice, GPT did not stand alone as technology. It was embedded within a reflective assessment structure, peer dialogue, and self-evaluation process. This is what made it educationally meaningful.

3. GPT as Evaluative Judgement Trigger

The third and most important theme was GPT as evaluative judgement trigger. Students did not simply accept the GPT feedback. Some recognised that the feedback could be too general, too long, or not fully aligned with their teaching context. Others noted that they needed to adapt the suggestions, preserve their own academic voice, and check whether the feedback matched the assignment requirements. This pattern is pedagogically valuable because it shows students practising judgement.

Evaluative judgement is central to this brief communication. Tai et al. (2018) argue that students need to learn how to make decisions about quality. Bearman et al. (2024) further argue that GenAI makes this capability more urgent because students must learn to judge AI outputs and AI-supported processes. In the present practice, GPT feedback became a stimulus for judgement. Students had to ask whether the feedback was relevant, whether it reflected their actual teaching experience, whether it aligned with the rubric, and whether it should be accepted, adapted, or rejected.

This finding challenges simplistic assumptions about AI in assessment. The problem is not only that AI may do too much for students. The deeper risk is that students may stop judging. Therefore, the best AI-supported assessment designs may be those that require students to critique AI feedback rather than merely consume it. This turns AI use into a form of assessment literacy development.

4. GPT as Ethical Tension

The fourth theme was ethical tension. Students recognised that GPT was useful, but they also identified risks such as overdependence, generic feedback, accuracy concerns, loss of personal voice, and the need to relate feedback to real classroom experience. This tension is not a weakness of the practice. It is part of the learning. Ethical AI use should not require pretending that AI is risk-free. Instead, students and educators should learn how to identify risks and negotiate responsible boundaries.

The ethical rule in this practice was clear: students could use GPT to understand the rubric and improve future reflections, but they could not use it to generate the weekly reflections. This distinction protected student authorship while allowing AI-supported feedback. It also reflected a situated approach to AI ethics. Gartner and Krašna (2023) emphasise autonomy, privacy, trust, and responsibility in AI education ethics, while Gaur et al. (2024) highlight privacy, bias, accountability, and inequality as key concerns. Mazlan et al. (2026a) extend this discussion by arguing that AI ethics must be understood through professional judgement and practice-based reasoning, not only through abstract policy principles.

Relational validity provides an additional critical lens. Din Eak et al. (2025) argue that assessment validity should include coherence among evidence, interpretation, dialogue, ethics, fairness, and educational consequence. Applying this to GPT-supported self-assessment, the key question is not simply whether the GPT feedback was "accurate." The deeper question is whether the practice strengthened students' relationship with assessment criteria, enhanced their capacity to interpret feedback, preserved ethical authorship, and supported meaningful learning consequences. From this perspective, AI-supported assessment should be judged by its educational relationships, not only by its technical outputs.

5. Implications for Assessment Practice

This reflective practice brief suggests several implications. First, AI-supported assessment should begin with assessment design, not with the tool. The AMU60104 Assessment & Feedback Assistant (M251PM2) was meaningful because it was linked to a specific assignment, rubric, course learning context, and ethical instruction. Without these anchors, AI feedback may become generic or misaligned. Second, AI feedback should remain formative and provisional. Students should be encouraged to treat AI feedback as something to interpret rather than something to obey. This supports assessment for learning because it encourages students to use feedback for future improvement while preserving responsibility for judgement (Nicol & Macfarlane-Dick, 2006; Wiliam, 2011).

Third, AI-supported assessment should deliberately develop feedback literacy and evaluative judgement. Students can be asked to compare AI feedback with rubric criteria, identify which suggestions are useful, reject suggestions that are inappropriate, and explain how they will apply feedback in future work. Such activities help students become active participants in feedback processes rather than passive recipients (Carless & Boud, 2018; Song et al., 2025; Tai et al., 2018). Fourth, AI integration should include ethical boundaries that are clear, teachable, and assessment specific. General statements such as "use AI responsibly" may be insufficient. Students need to know what forms of AI use are permitted, what forms are not permitted, and why these distinctions matter for authorship, learning, and fairness. This caution is particularly important in music education, where recent scholarship shows that AI applications may support teaching, learning, and assessment, but also raise pedagogical challenges related to musical interpretation, teacher expertise, contextual relevance, and ethical implementation (Mazlan et al., 2026b).

Finally, music education assessment should resist full automation because music learning involves contextual, expressive, technical, cultural, and pedagogical judgement. AI can support clarity, structure, and reflection, but it cannot replace the teacher's interpretive understanding of musical learning. The future of assessment in music education should therefore be human-centred and AI-aware rather than AI-driven.

6. Limitations

This brief communication has limitations. It is based on one master's coursework course and a small group of 18 consenting students. The reflections provide practice-based insight but do not support statistical generalisation. The paper also does not include pre- and post-measures of reflective writing quality, so it cannot claim measurable improvement in achievement. In addition, the study did not collect students' exact prompts, GPT response logs, or full interaction histories. As a result, the paper cannot determine whether the GPT assistant provided consistently accurate, rubric-aligned, or music-specific feedback across all student uses. The analysis is therefore limited to students' reflective accounts of their experience with GPT-supported self-assessment.

Another limitation concerns positionality. The course lecturer designed the GPT assistant and taught the course, which creates possible power-relation and social desirability concerns despite voluntary consent. Future research could compare lecturer feedback, peer feedback, AI feedback, and student uptake to examine alignment, accuracy, specificity, and learning consequences more systematically. Future studies could also collect prompt-output samples, conduct expert validation of GPT feedback, and investigate whether GPT-supported feedback literacy develops across multiple assignments.

CONCLUSION

This brief communication has argued that a customised GPT can function as a rubric mediator when embedded within clear assessment design, ethical instruction, and reflective practice. In the AMU60104 context, the AMU60104 Assessment & Feedback Assistant (M251PM2) was not used to replace the lecturer, automate grading, or generate student work. Instead, it supported students in interpreting rubrics, identifying areas for improvement, and reflecting on the quality of their own writing and assessment understanding.

The central argument is that AI in assessment should not be evaluated only through efficiency, automation, or technological novelty. Its deeper educational value lies in whether it strengthens human judgement. When students use AI feedback to understand criteria, question suggestions, preserve their own voice, and improve future work, AI may support assessment for learning. When students accept AI uncritically or use it to bypass thinking, it weakens assessment.

This is the time for education to rethink, redefine, and refine its stand on assessment in the age of AI. Educators and students are being asked not only to learn new tools, but also to relearn the purposes of feedback and unlearn narrow assumptions that assessment is merely grading, detection, or compliance. The emerging AI-mediated learning condition may be tentatively described as tetagogical: a convergence of teacher-guided pedagogy, adult learning, self-determined learning, peer-supported learning, cyber-mediated learning, and AI-supported feedback. The term is used here not as a settled theory, but as a provocation for further scholarly discussion. If pedagogy asks how teachers guide learning, and heutagogy asks how learners determine learning, tetagogy may ask how human and non-human agents now participate in feedback, judgement, and learning design.

The future of assessment should therefore not be framed as a choice between human judgement and artificial intelligence. It should be framed as a principled redesign of how both can coexist without weakening the ethical, reflective, and human purposes of education. In music education, where assessment is deeply tied to performance, interpretation, creativity, and professional judgement, this principle is especially important. AI may help students see the rubric more clearly, but educators must ensure that students still learn to hear, judge, question, and decide for themselves.

ACKNOWLEDGEMENT

The authors would like to thank the AMU60104 Assessment in Music Education Research students (M251PM2) for their reflective engagement and permission to use their assignment-related reflections

for scholarly purposes.

FUNDING

This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The data used in this reflective practice brief consist of anonymised student reflections and feedback from a coursework learning management system. Due to ethical and privacy considerations involving student assignment data, the full dataset is not publicly available. Anonymised excerpts may be made available from the author upon reasonable request, subject to institutional and participant confidentiality requirements.

CONFLICT OF INTEREST

The authors declare no financial conflicts of interest. One of the authors designed the AMU60104 Assessment & Feedback Assistant (M251PM2) and taught the course in which the reflective practice was implemented. This positionality is acknowledged as part of the reflective practice context.

DECLARATION OF GENERATIVE AI

During the preparation of this manuscript, ChatGPT was used to support drafting, language refinement, organisation of ideas, and checking of academic tone. The authors reviewed, edited, verified, and approved the final content and take full responsibility for the integrity, accuracy, and originality of the manuscript.

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*, Article 75. https://doi.org/10.1007/s43545-021-00086-w

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Bizami, N. A., Tasir, Z., & Kew, S. N. (2023). Innovative pedagogical principles and technological tools capabilities for immersive blended learning: A systematic literature review. *Education and Information Technologies, 28*, 1373–1425. https://doi.org/10.1007/s10639-022-11243-w

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. Routledge. https://doi.org/10.4324/9781315740621

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Din Eak, A., Ooi, L. H., & Tan, S. F. (2025). Relational validity in practice: A decade of research on alternative assessment in English language teaching higher education (2014–2024). *Asian Journal of Assessment in Teaching and Learning, 15*(2), 96–110. https://doi.org/10.37134/ajatel.vol15.2.7.2025

Gartner, S., & Krašna, M. (2023). Ethics of artificial intelligence in education. *Journal of Elementary Education,*

16(2), 221–237. https://doi.org/10.18690/rei.16.2.2846

Gaur, A. S., Sharan, H. O., & Kumar, R. (2024). AI in education: Ethical challenges and opportunities. In *The ethical frontier of AI and data analysis* (pp. 39–54). IGI Global. https://doi.org/10.4018/979-8-3693-2964-1.ch003

Gedera, D. (2023). A holistic approach to authentic assessment. *Asian Journal of Assessment in Teaching and Learning, 13*(2), 23–34. https://doi.org/10.37134/ajatel.vol13.2.3.2023

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Lian Zhao, & Ting, S.-H. (2025). Formative assessment through supervisory feedback on undergraduate thesis writing. *Asian Journal of Assessment in Teaching and Learning, 15*(1), 126–141. https://doi.org/10.37134/ajatel.vol15.1.9.2025

Mazlan, C. A. N., Abdullah, M. H., Othman, M. A., Md Noor, A. R., & Jamnongsarn, S. (2026a). From abstract ethics to situated practice: A bibliometric analysis of AI ethics and professional judgement. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-026-10630-1

Mazlan, C. A. N., Abdullah, M. H., Sulong, M. A., Abas, A., Uyub, A. I., Pisali, A., Hidayatullah, R., & Daud, I. S. (2024). Revolutionizing music education: Bite-sized learning and TikTok in the context of Education 5.0. *The International Journal of Arts, Culture & Heritage, 12*(4), 79–122. https://doi.org/10.62312/asw.ijach.12.4.2024

Mazlan, C. A. N., Hanafi, H. F., Sarifin, M. R., Md Noor, A. R., Sadykova, S. A., Hidayatullah, R., & Jamnongsarn, S. (2026b). Artificial intelligence applications and pedagogical challenges in music education. *Discover Education, 5*, Article 140. https://doi.org/10.1007/s44217-026-01127-3

Mazlan, C. A. N., Mohd Yusoff, M. Y., Safian, A. R., Koning, S. I., Saearani, M. F. T. B., Md Noor, A. R., Sadykova, S. A., & Sinaga, F. S. S. (2025). The evolution of modern learning: A dual systematic review of emerging trends and challenges. *Asian Journal of University Education, 21*(3), 793–813. https://doi.org/10.24191/ajue.v21i3.54

Mohamed Sapawi, M. S., & Nik Yusoff, N. M. R. (2025). Artificial intelligence in curriculum development: A global systematic review of trends, challenges, and strategic directions. *Journal of Curriculum Studies Research, 7*(2), 466–497. https://doi.org/10.46303/jcsr.2025.30

Ngu, I.Y., Lau, S.Y., & Schellini, M. (2025). Student's attitudes and motivations toward using ChatGPT feedback for personalised learning. *Asian Journal of Assessment in Teaching and Learning, 15*(2), 43–57. https://doi.org/10.37134/ajatel.vol15.2.3.2025

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Niloy, A. C., & Huda, T. (2025). Is AI only to blame? Assessing teachers' perceived challenges in AI detectability. *Asian Journal of Assessment in Teaching and Learning, 15*(2), 21–42. https://doi.org/10.37134/ajatel.vol15.2.2.2025

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803–830. https://doi.org/10.1007/s10648-015-9350-2

Parkes, K. A., & Burrack, F. (Eds.). (2020). *Developing and applying assessments in the music classroom*. Routledge. https://doi.org/10.4324/9780429202308

Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education, 35*(6), 727–743. https://doi.org/10.1080/02602930902977756

Sankaran, S., & Low, S. F. (2025). Effectiveness of feedback on continuous assessment: Students' views. *Asian Journal of Assessment in Teaching and Learning, 15*(1), 49–62. https://doi.org/10.37134/ajatel.vol15.1.4.2025

Song, M. H., Park, J., Lim, J., & Park, J. (2025). Improving feedback literacy through peer evaluation: A quantitative and qualitative analysis. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2025.2593371

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*, Article 15. https://doi.org/10.1186/s41239-024-00448-3

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Wu, S. P., Choi, Y. F., & Patel, L. (2025). Transforming feedback into learning throughout the trajectory of competency based medical education. *Indian Pediatrics, 62*(3), 221–227.

https://doi.org/10.1007/s13312-025-00012-w