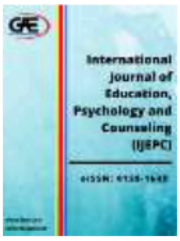
International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
https://gaexcellence.com/ijepc

PRE-SERVICE TEACHERS' EVALUATIVE JUDGMENTS AND THE "HUMAN-IN-THE-LOOP" PARADIGM IN AI-ASSISTED ASSESSMENT DESIGN

Wirawani Kamarulzaman[1*]

[1]Department of Educational Psychology and Counselling
wirawani@iium.edu.my https://orcid.org/0000-0002-5026-8668
*Corresponding Author

Article Info:

Article history:
Received date: 09.02.2026
Revised date: 26.02.2026
Accepted date: 09.03.2026
Published date: 31.03.2026

To cite this document:

Kamarulzaman, W. (2026). Pre-Service Teachers' Evaluative Judgments and the "Human-In-The-Loop" Paradigm in Ai-Assisted Assessment Design. *International Journal of Education, Psychology and Counselling, 11*(62), 1355-1368.

DOI: 10.35631/IJEPC.1162079

Abstract:

The paradigm for education has changed from a traditional teacher-tool approach to a hybrid intelligence framework due to the quick rise of Generative Artificial Intelligence (GenAI). This study examines the assessments of pre-service teachers at a public university in Malaysia as they utilise GenAI to generate subjective assessment items, including essay questions and interpretative exercises. Even though GenAI provides previously unheard-of efficiency for structural scaffolding and brainstorming, research shows that raw AI outputs frequently fall back on Lower-Order Thinking Skills (LOTS) and lack contextual empathy for learning demands. This study examines the ways in which aspiring teachers evaluate AI-generated content through a qualitative reflective thematic analysis of 35 teacher-designed artefacts. The results show that although AI provides a good "starting point" for draughting, human interaction is necessary to guarantee cognitive calibration, cultural sensitivity, and marking reliability. In order to close the gap between generic AI outputs and regional educational realities, pre-service teachers served as "cultural mediators," manually incorporating ideals like *Adl* (Justice) and *Sejahtera* into evaluations. In order to guarantee that automated content creation promotes deep conceptual learning rather than unthinking dependency, this study indicates that the teacher must continue to serve as the "essential architect" and maintain the pedagogical compass.

Keyword:

Classroom Assessment, Generative AI, Human-in-the-loop, Qualitative Method, University Students,

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368

Introduction

The swift progression of Generative Artificial Intelligence (GenAI) has transformed modern educational practices, compelling educators and academics to reevaluate the implementation of assessment, instructional design, and professional judgement within an AI-enhanced context. Unlike earlier digital tools that performed narrow, predefined functions, GenAI systems—powered by large-scale foundational models—can generate novel content, adapting to diverse tasks, and participating in complex pedagogical processes (Bui et al., 2025; Oh, 2025). As these tools become increasingly integrated into everyday teaching workflows, teacher education programs face the pressing challenge of preparing future educators to critically evaluate AI outputs, manage their limitations, and integrate them responsibly into assessment design (Prilop et al., 2025; Kohnke et al., 2025).

For pre-service teachers, this challenge is particularly significant. GenAI tools are now widely used to generate draft questions, interpretive exercises, rubrics, and feedback, yet research consistently shows that AI-generated content often lacks contextual subtlety, cultural grounding, and alignment with higher-order cognitive expectations (Fan et al., 2025; Navío Inglés et al., 2025). While GenAI can accelerate brainstorming and reduce the cognitive load associated with initial drafting, its outputs frequently require substantial human refinement to ensure validity, fairness, and pedagogical coherence (Ishmuradova et al., 2025; Busuttil, 2025). This dynamic underscore the importance of the Human in the Loop (HITL) paradigm, which positions teachers as active evaluators, editors, and cultural mediators who safeguard the integrity of assessment design (Rojas Contreras et al., 2025; Chen & Pi, 2025).

Moreover, GenAI systems are typically trained on large, homogenized datasets that may not reflect the cultural, linguistic, or ethical contexts of diverse learning environments. As a result, pre-service teachers must not only evaluate the cognitive quality of AI outputs but also ensure that assessments remain culturally resonant and ethically aligned with local educational values (Naidu & Sevnarayan, 2025). Understanding how pre-service teachers perform this evaluative work is essential for shaping future teacher education curricula and strengthening AI literacy in ways that support responsible, context-sensitive integration of GenAI into assessment practices.

Problem Statement

Despite the growing availability of GenAI tools, many pre-service teachers remain underprepared to navigate the complexities of AI-assisted assessment design. Studies indicate that future educators often struggle to critically evaluate AI-generated content, balance efficiency with pedagogical rigor, and determine when AI outputs require revision or rejection (Bui et al., 2025; Prilop et al., 2025). GenAI frequently defaults to lower-order thinking

patterns, producing assessment items that emphasize recall rather than analysis or evaluation, which risks undermining deep learning if adopted uncritically (Nguyen et al., 2025; Fan et al., 2025).

Additionally, GenAI systems often lack contextual empathy, failing to incorporate local cultural values, ethical considerations, or learner diversity—elements that are essential for equitable and meaningful assessment (Naidu & Sevnarayan, 2025). Without structured training in AI literacy, prompt engineering, and evaluative judgment, pre-service teachers may produce assessments that are generic, biased, or misaligned with curriculum standards (Busuttil, 2025; Oh, 2025). These gaps highlight the need to investigate how pre-service teachers critique, refine, and contextualize AI-generated assessment items during their formative training.

Research Objective

The objective of this research is to explore and analyse the evaluative judgments of pre-service teachers as they utilise Generative AI for the design and refinement of subjective assessment items. Specifically, the study seeks to identify the themes in their professional reflections regarding the utility, limitations, and irreplaceable role of human expertise in creating valid, fair, and culturally resonant educational measurements.

Literature Review

The rapid integration of Generative Artificial Intelligence (GenAI) into educational environments marks a profound paradigm shift that surpasses the impact of earlier disruptive technologies (Oh, 2025; Bui et al., 2025). Unlike tools such as calculators, which were limited to singular computational functions, GenAI—powered by foundational models like Large Language Models (LLMs)—operates with remarkable flexibility, generality, and creative capacity across diverse educational tasks (Bui et al., 2025). A defining feature of GenAI is its emergent behaviour, where unprogrammed capabilities arise from the scale and complexity of its training data (Bui et al., 2025; O'Dea, 2024). This is complemented by a homogenization effect, enabling a single model to be fine-tuned for a wide spectrum of educational applications, from automated scoring to sophisticated lesson planning (Bui et al., 2025).

Recent work by Lampropoulos and Papadakis (2025) further expands the conversation on AI in education by examining the educational value of intelligent social robots as affective, socially responsive agents capable of supporting personalized learning. Their findings highlight how AI-driven embodied systems can adapt to learners' needs, enhance engagement, and function as human-like tutors across diverse educational contexts. Although their study focuses on social robots rather than text-based generative models, the underlying principles—AI's capacity to personalize learning, respond to learner characteristics, and augment teacher practice—resonate strongly with current discussions on GenAI integration. Their call for clearer standards, ethical guidelines, and more empirical research underscores the broader need for responsible AI adoption in education, reinforcing the importance of teacher judgment and contextual sensitivity in AI-supported assessment design.

Recent research has also begun to examine teachers' perceptions of AI integration in authentic classroom contexts. Uğraş et al. (2024) explored early childhood teachers' experiences using ChatGPT to support STEM education and found that educators viewed GenAI as a valuable tool for providing instant feedback, generating personalized content, and enhancing creativity

and student motivation. Their findings highlight AI's potential to accelerate instructional processes and offer greater flexibility in planning, while also revealing challenges related to technical limitations, student readiness, and the need for stronger pedagogical and security-related support. Although their study focuses on early childhood STEM rather than assessment design, the broader implications resonate with the present study: teachers recognize the benefits of GenAI but emphasize the necessity of professional judgment, pedagogical oversight, and adequate training to ensure meaningful and responsible integration.

Within teacher education, this technological evolution has shifted the educator–tool relationship toward a hybrid intelligence paradigm, where AI functions as a collaborative thinking partner, creative catalyst, and cognitive scaffold (O'Dea, 2024; 2025; Bui et al., 2025). However, the accessibility of tools such as ChatGPT has outpaced formal pedagogical training, resulting in a scenario where students adopt AI tools before educators have established structured frameworks for their use (Bui et al., 2025). This gap underscores the urgent need to strengthen AI literacy among pre-service teachers, conceptualized as a multifaceted set of competencies that enable individuals to critically evaluate, communicate, and collaborate with AI systems (Prilop et al., 2025; MacDowell et al., 2024).

AI literacy encompasses cognitive, affective, social, and metacognitive dimensions (Bilbao Eraña & Arroyo Sagasta, 2025). In teacher education, it is triadic in nature, involving AI as a teaching tool, AI as a learning tool, and AI as teaching content (Prilop et al., 2025). This complexity requires pre-service teachers to not only use AI for efficiency but also anticipate student use, evaluate ethical implications, and navigate issues of cultural representation and democratic participation in AI-mediated learning environments.

The integration of GenAI into assessment design is particularly disruptive. While GenAI can provide immediate, real-time feedback that supports learning, its use challenges traditional methods of verifying independent authorship (Nguyen et al., 2025; Navío Inglés et al., 2025). Research indicates that although GenAI excels at generating initial frameworks and brainstorming ideas, its raw outputs often default to Lower Order Thinking Skills (LOTS), such as literal recall and surface comprehension (Nguyen et al., 2025; Oh, 2025). This raises concerns about "metacognitive laziness," where students rely on automated answers rather than engaging in reflective thinking. To counter this, educators must shift toward authentic assessment designs that foreground learning processes and target higher order cognitive gains (Fan et al., 2025; Nguyen et al., 2025).

These challenges intersect directly with the Human in the Loop (HITL) paradigm, which emphasizes the indispensable role of human oversight in AI-assisted assessment design. HITL approaches ensure that human judgment remains central in verifying and validating AI outputs, addressing issues of accuracy, creativity, and ethical integrity (Rojas Contreras et al., 2025; Öz et al., 2022; Tocchetti et al., 2025). Studies show that while AI can generate initial assessment items, substantial human refinement is required to ensure clarity, alignment with learning objectives, and contextual appropriateness (May et al., 2025). For pre-service teachers, this human–AI collaboration enhances evaluative judgment by enabling them to critically interrogate AI-generated content, refine assessment items, and develop a deeper understanding of quality criteria in assessment design (Jho & Lee, 2025; Erdem Coşgun, 2025). At the same time, pre-service teachers express concerns about over-reliance on AI, potential biases, and the need for ethical guidelines, reinforcing the importance of embedding AI literacy and ethical

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368

training within teacher education programs (Kohnke et al., 2025; Oktay & Kilinç, 2025; Dayagbil et al., 2025).

Cultural relevance and contextual sensitivity further complicate the integration of GenAI in assessment. Many GenAI models are trained on homogenized, Western-centric corpora, which risk reproducing cultural biases and marginalizing non-dominant identities (Naidu & Sevnarayan, 2025). This positions teachers as cultural mediators who must audit, localize, and contextualize AI outputs to reflect the diverse lived experiences of their students. For pre-service teachers, this involves a co creation process in which they act as essential architects, ensuring that AI-generated content is not only technically accurate but also ethically grounded and culturally responsive (Oh, 2025; Naidu & Sevnarayan, 2025). Within the HITL framework, this responsibility becomes even more pronounced, as human oversight is the mechanism through which cultural complexity, contextual relevance, and ethical considerations are safeguarded (Chen & Pi, 2025).

Taken together, the convergence of GenAI's transformative capabilities and the HITL paradigm underscores a critical shift in teacher education. Pre-service teachers must learn to navigate AI not as a replacement for human judgment but as a collaborative partner whose outputs require careful interpretation, refinement, and ethical stewardship. This blended approach—anchored in AI literacy, cultural mediation, and human oversight—positions future educators to harness the potential of GenAI while preserving the human values at the heart of teaching and learning.

Methodology

Research Design

This study employed a qualitative reflexive thematic analysis design to investigate how pre-service teachers exercise evaluative judgment when engaging with Generative AI during assessment construction. Conducted as a Scholarship of Teaching and Learning (SoTL) inquiry embedded within a regular course context, the study sought to systematically examine authentic classroom practices while maintaining ethical safeguards appropriate for minimal-risk educational research. Reflexive thematic analysis was selected because it supports an interpretive, iterative, and researcher-involved approach to meaning making, allowing themes to emerge through deep engagement with participants' reflections and revised assessment artifacts (Prilop et al., 2025). The study was conceptually anchored in the Human-in-the-Loop (HITL) paradigm, which positions AI as a co-creative partner while foregrounding the teacher's role in verification, refinement, and contextualization (Bui et al., 2025; Oh, 2025). This framework enabled the analysis to focus on how participants negotiated the boundaries between automated generation and human pedagogical expertise.

Participants

Participants were students of bachelor's in education program from a public university in Selangor. They were enrolled in the program of Teaching of English as a Second Language, Teaching of Arabic as a Second Language, Guidance and Counselling and Islamic Education. These individuals were in the formative stages of their professional development and had varying levels of familiarity with AI tools. Their engagement with assessment design formed part of a broader course requirement focused on testing and evaluation. This diversity provided

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368

a rich dataset for examining how emerging educators interact with GenAI and apply evaluative judgment across disciplinary contexts.

Data Collection Procedure

One of the core assignments within the class assessments required participants to design interpretive exercises and essay questions—this specific assignment formed the basis of the study. Data were collected through a structured assessment task in which participants designed subjective assessment items, specifically interpretive exercises and essay questions, using GenAI as an initial drafting tool. The task unfolded in three stages:

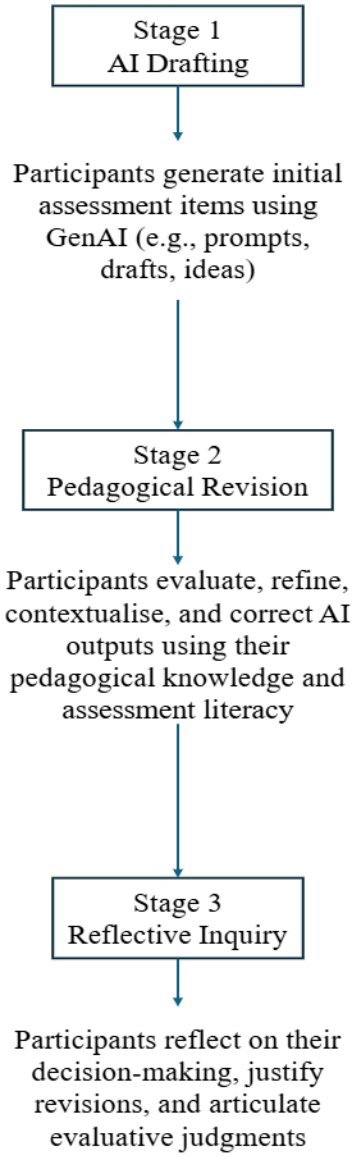


Figure 1: Data Collection Procedure

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368

1. **AI-Assisted Drafting:** Participants used GenAI platforms (e.g., ChatGPT, Gemini, Copilot) to generate preliminary versions of interpretive exercises and essay prompts. These drafts included items targeting inference, assumptions, relevance, principles, and pictorial interpretation.
2. **Pedagogical Revision:** Participants critically evaluated the AI-generated drafts against learning outcomes, Bloom's Taxonomy, and principles of fairness and clarity. They manually revised items to improve cognitive demand, linguistic precision, and alignment with curriculum expectations.
3. **Reflective Inquiry:** Participants produced written reflections documenting their critique of AI outputs, the rationale behind their revisions, and their perceptions of AI's strengths and limitations in assessment design.

These artifacts—AI drafts, revised items, and reflective commentaries—constituted the primary dataset.

Data Analysis Procedure

Data were analysed using inductive and deductive coding strategies consistent with reflexive thematic analysis (Bui et al., 2025; Prilop et al., 2025). The analysis proceeded through the following stages:

1. **Familiarization:** All reflections and revised items were read repeatedly to gain a holistic understanding of participants' experiences and evaluative processes.
2. **Initial Coding:** Data segments were coded to capture recurring ideas such as "AI as scaffolding," "LOTS vs. HOTS," "technical precision," and "cultural blind spots."
3. **Theme Development:** Codes were clustered into broader themes that reflected participants' evolving AI literacy and their transition from passive consumers to active co-designers of assessment.
4. **Triangulation:** AI-generated drafts were compared with participants' revised versions to validate claims about AI limitations and the necessity of human intervention.

This analytic process enabled a complex understanding of how pre-service teachers enacted evaluative judgment within HITL workflows.

Ethical Considerations

The study was carried out as a Scholarship of Teaching and Learning (SoTL) project within the author's regular teaching context. In line with institutional guidelines for minimal-risk classroom research, SoTL inquiries of this nature do not require formal ethics committee approval; however, the institution was fully aware of the study and its procedures. All participants were informed about the purpose of the research and the voluntary nature of their involvement, and written informed consent was obtained prior to data collection. To ensure confidentiality, all data were anonymised, and no identifying information was collected. Participation had no impact on students' grades, assessment outcomes, or course standing. These measures ensured that the study adhered to ethical standards for research involving

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 62 (March 2026) PP. 1355-1368

human participants while respecting the pedagogical environment in which the inquiry took place.

Findings and Discussion

Following the reflexive thematic analysis of the assessment artifacts and reflective commentaries, several salient themes emerged that illuminate how pre-service teachers exercised evaluative judgment within Human-in-the-Loop (HITL) workflows. These themes reflect the subtle ways in which participants interacted with Generative AI—both as a productive scaffolding tool and as a source of pedagogical, cognitive, and cultural limitations requiring human intervention. The findings are presented below, each accompanied by a discussion that situates the theme within existing scholarship.

Efficiency and AI as a Scaffolding Partner

Pre-service teachers consistently found that Generative AI serves as an "incredibly powerful tool for initial drafting" and acts as a "high-speed assistant" that helps overcome the "blank page hurdle" [P29, P21, P8]. Participants noted that the technology is "exceptionally effective" in producing structured item frameworks, such as True/False/No Inference formats, in "a matter of seconds" [P1, P25]. By providing a "clear item frame and starting point," AI significantly reduces the "preparatory time" and "brainstorming phase" required to develop complex interpretive exercises [P4, P11, P28]. One student highlighted that it "excelled at the preliminary stage" by suggesting a variety of relevant themes and wording options that might not surface as quickly through manual effort [P11, P16, P22].

This thematic finding is strongly aligned with the research of Bui et al. (2025), which established that "ideation and content creation" are the most common GenAI-supported tasks among future educators, who value the tool for its "creative support" despite accuracy concerns. This synergy is further reflected in the work of Oh (2025), who posits that automated problem generation allows teachers to move away from "time-consuming problem creation" and toward more meaningful "instructional facilitation." By leveraging AI for the "heavy lifting" of formatting, teachers can explore a broader range of instructional strategies (Oh, 2025). This reflects a transition toward the "hybrid intelligence paradigm" where AI acts as a collaborative thinking partner and intellectual scaffolding for human thought (O'Dea, 2024, 2025).

Defaulting to Lower-Order Thinking Skills (LOTS)

A significant number of students observed that AI "consistently defaulted to lower order thinking skills" such as literal recall and surface comprehension [P21, P25, P12]. Reflections indicated that AI "tends to give obvious false or correct items," which "significantly reduces the challenge for students" [P3, P36]. Furthermore, participants noted that interpretive questions generated by AI often "copy exact phrases from the stimulus," creating items that can be answered via "eye scanning" rather than genuine reasoning [P25, P18, P31]. AI was found to struggle in creating "pedagogical tension" or complex dilemmas where multiple answers might seem plausible, often failing to challenge students with "higher-order cognitive processes" [P35, P21, P14].

These observations confirm the warnings of Fan et al. (2025) regarding "metacognitive laziness," wherein an uncritical reliance on AI content can diminish reflective thinking and deep cognitive engagement. This highlights a critical need, as expressed by Nguyen et al. (2025), to "rethinks conventional assessment goals" to ensure they capture learning gains that AI cannot easily automate. Because traditional assessments are increasingly "susceptible to cheating" via AI, policy guidelines from the OECD (2023) and UNESCO (Miao & Holmes, 2023) advocate for a shift toward evaluating higher-order thinking and analysis over rote memorization. The pre-service teachers' findings reinforce that while AI can simulate a test structure, it often lacks the pedagogical depth required to foster "deep conceptual learning" (Oh, 2025).

The Need for Technical Precision and Reliability

Participants identified that AI-generated essay prompts were frequently "too broad" and lacked "specific and explicit criteria, such as word limits" [P11, P4, P23]. Some students found that AI instructions were "too vague" or involved "run-on sentences" that could "demotivate students" or lead to "inconsistent responses" [P4, P17, P20]. Additionally, pre-service teachers noted that AI sometimes produced "semantic distortions" and "unsubstantiated causal claims," such as incorrectly attributing a ship's sinking solely to weather without textual evidence [P22, P11]. Without human intervention, students concluded that "AI could not ensure marking reliability" or ensure that a prompt was "fair for everyone" [P11, P7, P26].

This finding matches the research of Navío-Inglés et al. (2025), who emphasize that "human expertise remains essential" to ensure the precision and relevance of assessment feedback. The lack of specific constraints in AI outputs correlates with observations by Busuttil (2025), who identified that "limitations in linguistic accuracy" and "time management" in AI-generated materials require significant human oversight. This supports the argument by Prilop et al. (2025) that teacher education must move toward "reconceptualizing assessment" to establish the clear boundaries and parameters that AI frequently omits. Ultimately, the teacher must act as the "editor-in-chief" to ensure that assessments are logically sound and "scientifically accurate" [P21, P22], a role Bui et al. (2025) describe as the "Human-in-the-Loop" (HITL) paradigm.

Cultural Blind Spots and the Need for Cultural Sensitivity

A critical theme was AI's "lack of cultural sensitivity" and its tendency to produce "generic international contexts" that ignored local realities [P33, P29, P16]. The participants believed that AI failed to understand specific values such as *Adl* (Justice) and *Sejahtera*, or the socio-economic reality that many students rely on Zakat support [P35, P26]. Teachers were required to "purify" items of inappropriate tropes and manually integrate institutional values to make the tests "meaningful for our campus" [P35, P24, P33]. AI also struggled with delicately in value-based subjects, sometimes producing items that "contradict Islamic teaching" [P37, P25].

The struggle to localize content aligns with Naidu and Sevnarayan (2025), who argue that AI often reproduces "Western-centric norms" due to training data being skewed toward homogenized corpora. Pre-service teachers acted as "cultural mediators," a role described by Naidu and Sevnarayan (2025) as vital for ensuring "epistemic justice" such as cultural sensitivity and preventing the marginalization of non-dominant identities. This finding also supports Prilop et al.'s (2025) model of "triadic AI literacy," which requires educators to

evaluate technology within "ethical, cultural, and democratic contexts." The manual injection of local values by students demonstrates that teacher agency is necessary to transform AI from a source of cultural bias into a tool for "cultural scaffolding" (Naidu & Sevnarayan, 2025).

The Irreplaceable Role of the Teacher

The common conclusion across all artifacts was that "AI intelligence can never be compared to the subtlety of human intelligence" [P3]. While AI is a "valuable support tool," it cannot replace the "professional judgment" and "human touch" required to ensure an assessment is fair and aligned with "curriculum standards like the Malaysian DSKP" [P13, P23, P29, P32]. Pre-service teachers asserted that the teacher must remain the "essential architect" and "compass" of the process to ensure "cognitive calibration"—shifting items from simple recall to higher levels of evaluation [P35, P11, P21]. Ultimately, they determined that only a teacher knows their students' "strengths, weaknesses, and overall ability" [P10, P16].

This final theme reinforces the "Human-in-the-Loop" (HITL) paradigm discussed by Bui et al. (2025), where the user is an active participant in augmented cognition rather than a passive consumer. As Oh (2025) concludes, while AI can handle the "heavy lifting," the teacher must "hold the compass" to guide the pedagogical vision and ensure educational rigor. This aligns with Ishmuradova et al. (2025), who found that despite optimism about AI's utility, pre-service teachers maintain a necessary "scepticism" that prioritises the role of the human educator in ensuring "validity and quality." The teacher's role is thus not diminished but augmented, shifting from a content generator to a "critical evaluator and co-designer" (Oh, 2025).

Conclusion

This study demonstrates that while Generative AI is an exceptionally effective efficiency catalyst for the preliminary stages of assessment design, it cannot function as an autonomous creator. The artifacts developed by pre-service teachers reveal a consistent "cognitive reliability gap," where AI-generated items frequently rely on surface-level data extraction rather than the higher-order synthesis required for complex achievement. Human expertise remains irreplaceable for cognitive calibration—the precise process of ensuring that questions challenge students at the appropriate levels of Bloom's Taxonomy.

Furthermore, the research highlights the critical importance of contextual and ethical sensitivity. AI often fails to grasp the "lived experience" of specific communities, such as the socio-economic specifics of students receiving Zakat support or the integration of Islamic values. The pre-service teachers' transition from passive consumers to critical co-designers reinforces the "Human-in-the-Loop" (HITL) paradigm, where technology augments human capability without diminishing the teacher's professional identity.

Ultimately, GenAI is a powerful "brainstorming partner" that helps educators overcome the initial "blank page" hurdle, but the teacher remains the "editor-in-chief" and the "pedagogical compass". To ensure fairness, relevance and assessment validity, future teacher education must prioritize AI literacy—specifically in prompt engineering and evaluative judgment—to empower educators to refine AI outputs into meaningful, fair, and rigorous educational measurements. While AI can do the "heavy lifting" of formatting and drafting, it is the human teacher who ensures that the final assessment is pedagogically sound, inclusive, and scientifically accurate.

Acknowledgements: The authors would like to express their sincere gratitude to International Islamic University Malaysia (IIUM) for providing the necessary support throughout the course of this research.

Funding Statement: No Funding

Conflict of Interest Statement: The author declares that there is no conflict of interest regarding the publication of this paper. The author has contributed to this work and approved the final version of the manuscript for submission to the International Journal of Education, Psychology and Counseling. (IJEPC).

Ethics Statement: This project was undertaken as a Scholarship of Teaching and Learning (SoTL) study within a regular course context. According to institutional guidelines, research of this nature does not require prior ethics committee approval. Nonetheless, ethical safeguards were implemented: participants were informed about the purpose of the study, participation was voluntary, and all data were anonymized to protect confidentiality. No identifying information was collected, and participation had no impact on students' grades or academic standing.

Author Contribution Statement: The author was solely responsible for all aspects of this study, including conceptualization, study design, classroom implementation, data collection, data analysis, interpretation of findings, and manuscript preparation. The author affirms full responsibility for the integrity and accuracy of the work.

References

Addison, L., Ramdial, P. T., & Hosein, P. (2024). *An investigation into the use of large language models for adaptive grading of course projects*. In 2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD 2024). https://www.scopus.com/pages/publications/86000016073?origin=scopusAI

Banawan, M., Kukas, E. J., Tano, J. R., & Villegas, R. (2024). *Transformative approach to fairness and transparency in classroom participation assessment*. In L@S 2024: Proceedings of the 11th ACM Conference on Learning @ Scale. https://www.scopus.com/pages/publications/85199860877?origin=scopusAI

Bui, P., Korhonen, T., Kontkanen, S., Karme, S., Piispa-Hakala, S., & Veermans, M. (2025). *Exploring pre-service teachers' generative AI readiness and behavioral intentions: A pilot study*. *LUMAT: International Journal on Math, Science and Technology Education, 13*(1), 8. https://doi.org/10.31129/LUMAT.13.1.2755

Busuttil, L. (2025). *Leveraging generative AI in pre-service teacher training: Insights into assessment as learning*. In Communications in Computer and Information Science. https://doi.org/10.1007/978-3-031-94002-6_4

Chen, X., & Pi, X. (2025). *Human-in-the-loop: A conceptual framework for Business English teachers' AI-empowered assessment*. In Lecture Notes in Computer Science. https://www.scopus.com/pages/publications/105003621373?origin=scopusAI

Dayagbil, F. T., Boholano, H. B., & Sumalinog, G. G. (2025). *Are they in or out? Exploring pre-service teachers' knowledge, perceptions, and experiences regarding artificial intelligence (AI) in teaching and learning*. *Frontiers in Education*. https://www.scopus.com/pages/publications/105018501353?origin=scopusAI

Dockens, A. L., & Shelton, K. (2025). *AI for formative and summative assessment: A balanced approach*. In Emerging Trends, Global Perspectives, and Systemic Transformation in AI. https://www.scopus.com/pages/publications/105017950051?origin=scopusAI

Erdem Coşgun, G. (2025). *Artificial intelligence literacy in assessment: Empowering pre-service teachers to design effective exam questions for language learning*. *British Educational Research Journal*. https://www.scopus.com/pages/publications/105002217074?origin=scopusAI

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). *Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance*. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Ishmuradova, I. I., Zhdanov, S. P., Kondrashev, S. V., Erokhova, N. S., Grishnova, E. E., & Volosova, N. Y. (2025). *Pre-service science teachers' perception on using generative artificial intelligence in science education*. *Contemporary Educational Technology, 17*(3), ep579. https://doi.org/10.30935/cedtech/16207

Jho, H., & Lee, B. (2025). *Analysis of pre-service science teachers' inquiry design through interaction with an AI inquiry assistant*. *New Physics: Sae Mulli*. https://www.scopus.com/pages/publications/105017055644?origin=scopusAI

Kohnke, L., Zou, D., Ou, A. W., & Gu, M. M. (2025). *Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration*. *Computers and Education: Artificial Intelligence*. https://www.scopus.com/pages/publications/105000935937?origin=scopusAI

Lampropoulos, G., & Papadakis, S. (2025). The educational value of artificial intelligence and social robots. In *The educational value of artificial intelligence and social robots*. (pp 3-15). Springer. https://doi.org/10.1007/978-3-031-82915-4_1

May, T. A., Provinzano, K., Koskey, K. L. K., & Rimkunas, N. (2025). *Expanding the team: Integrating generative artificial intelligence into the assessment development process*. *Applied Sciences (Switzerland)*. https://www.scopus.com/pages/publications/105017225799?origin=scopusAI

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

Naidu, K., & Sevnarayan, K. (2025). *Culturally relevant assessment and generative AI: A co-creation framework for pre-service teachers*. *Journal of Applied Learning & Teaching, 8*(2). https://doi.org/10.37074/jalt.2025.8.2.20

Navío-Inglés, M., Guzmán Mora, J., O'Connor-Jiménez, P., & García González, A. (2025). *What's next for feedback in writing instruction? Pre-service teachers' perceptions of assessment practices and the role of generative AI*. *Education Sciences, 15*(11), 1534. https://doi.org/10.3390/educsci15111534

Nguyen, A., Duong, A. T., Nguyen, D. T. B., Lai, V. T. T., & Dang, B. (2025). *Guidelines for learning design and assessment for generative artificial intelligence-integrated education: A unified view*. *Information and Learning Sciences, 126*(7–8), 491–512. https://doi.org/10.1108/ILS-11-2024-0148

O'Dea, X. (2024). Generative AI: is it a paradigm shift for higher education? *Studies in Higher Education*, *49*(5), 811–816. https://doi.org/10.1080/03075079.2024.2332944

OECD. (2023). *Generative AI in the classroom: From hype to reality?* OECD Publishing.

Oh, S. (2025). *Integration of MATH41 and generative AI in pre-service mathematics teacher education: An empirical study on lesson design competency*. *IEEE Access, 13*. https://doi.org/10.1109/ACCESS.2025.3586593

Oktay, O. S., & Kilinç, H. (2025). *Pre-service teachers' perspectives on the use of artificial intelligence in learning processes: Biases, advantages, and disadvantages*. In Policy and Practice for Ethical AI Integration in Education. https://www.scopus.com/pages/publications/105024037881?origin=scopusAI

Öz, B., Wang, R., Chandler, C., & Léger, P.-M. (2022). *Responsible artificial intelligence in knowledge work: User experience design problems and implications*. In Lecture Notes in Computer Science. https://www.scopus.com/pages/publications/85144404160?origin=scopusAI

Prilop, C. N., Mah, D.-K., Jacobsen, L. J., Hansen, R. R., Weber, K. E., & Hoya, F. (2025). *Generative AI in teacher education: Educators' perceptions of transformative potentials and the triadic nature of AI literacy explored through AI-enhanced methods*. *Computers and Education: Artificial Intelligence, 9*, 100471. https://doi.org/10.1016/j.caeai.2025.100471

Puerta-Beldarrain, M., Gómez-Carmona, O., Sánchez-Corcuera, R., Casado-Mansilla, D., López-de-Ipiña, D., & Chen, L. (2025). *A multifaceted vision of the human-AI collaboration: A comprehensive review*. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3536095

Rojas-Contreras, M., Orjuela Duarte, A., & Marina Santos Jaimes, L. (2025). *Human-in-the-loop (HITL) as a verification and validation strategy for knowledge generated by generative artificial intelligence*. In Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology. https://www.scopus.com/pages/publications/105019318086?origin=scopusAI

Tanik Önal, N., Önal, N., & Ilhan, A. G. (2026). *Artificial intelligence in early childhood classrooms: Insights from in-service and pre-service teachers*. *European Early Childhood Education Research Journal*. https://www.scopus.com/pages/publications/105026582759?origin=scopusAI

Tocchetti, A., Bianchi, M., Campi, R., & Brambilla, M. (2025). *On the principles and effectiveness of gamification in bidirectional artificial intelligence and explainable AI*. In Bi-directionality in Human-AI Collaborative Systems. https://www.scopus.com/pages/publications/105019729263?origin=scopusAI

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). Innovative early childhood STEM education with ChatGPT: Teacher perspectives. *Technology, Knowledge and Learning, 30*(2), 809–831. https://doi.org/10.1007/s10758-024-09804-8