ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 7 (July 2026)

Exploring Mathematics Teachers' Experiences in Designing and Implementing AI-Generated Assessments

Salamah M. Basher [1*], Nasim B. Sadic [2], Nafisah A. Abdulnasser [3], Sittie Nashebah Guinto [4]
[1]*Institute of Science Education, Mindanao State University, Marawi City, Philippines*
[2]*Ansano Memorial National High School, Taraka, Lanao Del Sur, Philippines*
[3]*Dansalan National High School, Marawi City, Lanao Del Sur, Philippines*
[4]*Bacolod-Kalawi National High School, Bacolod-Kalawi, Lanao del Sur, Philippines*
[*]*salamah.indad@msumain.edu.ph, sadic.nasim@s.msumain.edu.ph,*
abdulnasser.na979@s.msumain.edu.ph, guinto.sittienashebah@s.msumain.edu.ph

Date Submitted:
July 12, 2026

Date Accepted:
July 13, 2026

Date Published:
July 14, 2026

DOI:
10.5281/zenodo.21357913

ABSTRACT

This qualitative phenomenological study explored the lived experiences of mathematics teachers in designing and implementing AI-generated assessments in public and private high schools across Lanao del Sur, Philippines. Grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT) and Pedagogical Content Knowledge (PCK), the research examined teachers' perceptions of value, design processes, encountered challenges and opportunities, and strategies for ensuring alignment with instructional goals. Data were gathered through semi-structured interviews with six purposively selected teachers and analyzed using Moustakas's modified Stevick-Colaizzi-Keen method. Findings revealed that teachers view AI primarily as a supportive tool rather than a replacement for professional expertise, valuing it most for reducing workload and generating diverse assessment materials. They follow a deliberate workflow involving clear prompt specification, rigorous verification, contextual localization, and iterative refinement. Key challenges include mathematical inaccuracies in AI outputs and unstable internet connectivity, while opportunities lie in enhanced efficiency and support for differentiated instruction. Teachers ensure quality by anchoring assessments in curriculum standards and adapting content to the cultural and learning needs of local students. The study concludes that effective AI integration follows a "human-in-the-loop" model, where technology serves as an aid while teachers retain authority over pedagogical quality. Recommendations are offered for teachers, school leaders, policymakers, developers, and future researchers to promote responsible and equitable AI use in mathematics education.

Keywords: *Artificial Intelligence, AI-generated assessments, mathematics education, teachers' experiences, pedagogical content knowledge, technology acceptance, qualitative phenomenology, Philippines*

INTRODUCTION

The educational landscape was experiencing a profound shift, largely driven by the rapid advancements in artificial intelligence (AI). What had once been considered speculative fiction had become an integral part of teaching and learning, promising to transform pedagogical practices and administrative efficiencies across all levels of (Owan et al., 2023). Within the realm of assessment, AI offered remarkable capabilities. It moved beyond static evaluations, allowing for the dynamic generation of personalized and diverse assessment tools, alongside innovative approaches to adaptive feedback and grading (Gorgun et al., 2025; Coskun & Alper, 2024). This evolving scenario held particular significance for mathematics education—a field fundamentally built on precise problem-solving and conceptual clarity. Here, the caliber and appropriateness of assessment directly influenced both student learning outcomes and the accuracy of diagnostic insights (NCTM, 2024).

Despite the widely acknowledged potential of AI-generated assessments to boost efficiency, personalize learning paths, and deliver immediate feedback, a crucial gap persisted in the understanding of their practical integration. This gap was particularly evident from the perspective of the educators responsible for deploying these tools. While AI assessment creation tools were increasingly available, the detailed experiences of mathematics teachers as they designed and implemented these AI-generated assessments in their daily work remained largely unexplored prior to this study. This research, therefore, addressed a fundamental question: What were the challenges, adaptations, and pedagogical considerations mathematics teachers encountered as they strove to move "beyond automation," purposefully weaving AI into their assessment strategies?

Existing academic literature consistently highlighted AI's transformative power across various educational sectors. Studies confirmed AI's capacity to automate routine tasks, tailor learning content, and even provide comprehensive feedback (Owan et al., 2023). In assessment specifically, AI had demonstrated its ability to efficiently generate a range of items, from multiple-choice questions to complex problem scenarios. This capability saved teachers considerable time and allowed for greater diversity in assessment materials, especially with the advancements in large language models (Rane, 2023; Coskun & Alper, 2024). Furthermore, research universally emphasized that a teacher's expertise was indispensable for effective assessment design. While technology could undoubtedly assist, genuine pedagogical insight and deep subject matter knowledge were critical for ensuring assessments were both valid and reliable (NCTM, 2024; Ollington, 2025). The mathematics education community, notably, had always stressed the importance of assessments that measured not only procedural fluency but also conceptual understanding and sophisticated problem-solving skills, requiring careful design and active teacher involvement (NCTM, 2024). Collectively, these developments suggested that while AI offered fertile ground for integration, they also underscored the profound need to understand how teachers navigated this process.

However, despite the growing discussions about AI in education and the recognized importance of teacher expertise in assessment, there was a notable scarcity of empirical data specifically investigating mathematics teachers' lived experiences in designing and implementing AI-generated assessments. Much of the current scholarly work focused on the technical functionalities of AI assessment tools (Rane, 2023), or explored theoretical benefits and broad ethical concerns (Gorgun et al., 2025; Khan et al., 2026). There had been limited attention paid to the practical pedagogical decisions, the specific challenges, and the adaptations teachers made when actively using these tools in authentic classroom environments. Ollington (2025), for instance, pointed out the necessity of exploring how AI influenced teachers' assessment identity, indicating a clear gap in the understanding of teacher perspectives. Crucially, existing research had not yet thoroughly examined how mathematics teachers integrated their specialized pedagogical content knowledge with AI capabilities, nor had it adequately documented the practical implications of such integration. This study filled that gap.

The findings from this study provided an invaluable resource for stakeholders within the educational community. Mathematics teachers gained insights into practical strategies and effective practices shared by colleagues, empowering them to confidently integrate AI into assessment routines. This research also informed professional development initiatives, helping to foster environments where AI was used thoughtfully. Educational technology developers benefited from direct feedback, which helped refine platforms to better suit pedagogical needs. For curriculum designers and policymakers, the insights were instrumental in crafting guidelines and frameworks that integrated AI meaningfully. School administrators gained a clearer understanding of support systems and training requirements. Finally, this research meaningfully contributed to the academic discourse, offering empirical evidence that served as a foundation for future studies on technology in education.

Literature Review

Artificial Intelligence in Educational Assessment

AI is increasingly integrated into education to automate scoring, generate items, and deliver feedback across formative and summative contexts, supporting differentiated instruction (Zawacki-Richter et al., 2019; Holmes & Porayska-Pomsta, 2023; Mekadenaumporn, 2025). Generative AI enables creation of varied, customized content, but raises concerns regarding validity, reliability, fairness, algorithmic bias, and limited capacity to measure higher-order thinking skills (Meylani, 2025; Rane, 2023). The National Council of Teachers of Mathematics (2024) emphasizes that diverse assessment methods must be guided by teacher expertise to measure both procedural fluency and conceptual understanding.

Teachers' Perceptions of AI in Assessment

Educators generally recognize AI's value in reducing workload and streamlining workflows, aligning with constructs from the Unified Theory of Acceptance and Use of Technology (UTAUT) such as performance expectancy and ease of use (Owan et al., 2023; Coskun & Alper, 2024; Alwaqdani, 2025; Venkatesh et al., 2003). While many view AI as useful for generating items and personalizing learning, some express concerns about over-reliance, reduced critical thinking, and ethical risks including academic dishonesty (Mekadenaumporn, 2025; Lin et al., 2025). Both local and foreign literature confirm teachers see AI as a support tool, not a replacement for professional judgment (Ollington, 2025).

Processes in Designing AI-Generated Assessments

Teachers follow an iterative workflow: defining clear prompts, generating content, then reviewing, validating, and customizing outputs—operating on a "human-in-the-loop" model (Meylani, 2025; Gorgun et al., 2025; Khan et al., 2026). They prioritize alignment with learning objectives, curriculum standards, and local context, reflecting the Philippine Department of Education's mandates for localization and contextualization (Zawacki-Richter et al., 2019). Review and verification are universally treated as non-negotiable steps to ensure quality (National Council of Teachers of Mathematics, 2024).

Challenges and Opportunities in AI Integration

Opportunities: AI boosts efficiency, saves time, supports instant feedback, and enables differentiated instruction (Alwaqdani, 2025). Challenges: Mathematical inaccuracies, misalignment with curriculum, and unstable internet/infrastructure gaps—especially in rural areas—limit consistent and equitable use (Meylani, 2025; Zhao, 2024). Philippine studies note these barriers are compounded by lack of training and tool familiarity (Rubio & Pastor, 2026; Calubag, 2026).

Alignment of AI Assessments with Instructional Goals

Teachers validate outputs against standards and adapt them to student needs, drawing on Pedagogical Content Knowledge (PCK) to ensure pedagogical soundness and cultural relevance (Pellegrino et al., 2012; Shulman, 1986; Holmes & Porayska-Pomsta, 2023). Human-centered approaches maintain teacher control over decisions, confirming AI's role as supportive rather than authoritative (Baker & Inventado, 2016; Meylani, 2025).

Related Foreign Studies

Systematic reviews confirm AI enhances feedback and engagement, particularly in mathematics; however, teacher review remains essential for validity and fairness (Holmes & Porayska-Pomsta, 2023; Chiu et al., 2025). While AI improves scoring consistency, transparency and interpretation require human mediation (Zhao, 2024). Confidence in use varies by training, affirming the need for professional development (Meylani, 2025). Research remains limited on lived experiences in subject-specific contexts like Philippine mathematics education (Meylani, 2025).

Related Local (Philippine) Studies

Filipino teachers hold positive views on AI but face barriers: limited training, tool unfamiliarity, and poor connectivity (Calubag, 2026; Rubio & Pastor, 2026; Cabato, 2025). Studies show they treat AI outputs as drafts, conducting rigorous validation and adaptation (Cabato, 2025; Silagan & Tumapon, 2025). Phenomenological inquiries confirm educators prioritize a teacher-centered framework, emphasizing infrastructure upgrades, capacity building, ethical guidelines, and structured AI literacy programs (Jesus & Caumeran, 2026; Dayagbil et al., 2025; Malacapay, 2025; Co, 2025). Educators consistently affirm AI as a supportive tool—never a substitute for professional oversight or contextual judgment.

METHODS

Research Design

This study employed a phenomenological research design to explore the essence of teachers' experiences with AI-generated assessments. This approach was appropriate because the goal was to understand how participants perceived, interpreted, and made meaning of the phenomenon, rather than merely describing events (Creswell & Poth, 2018; Moustakas, 1994).

The design allowed for an in-depth investigation of benefits, challenges, strategies, and perceived impact. It involved bracketing personal biases to focus on participants' subjective accounts, ensuring that authentic perspectives emerged. The findings thus offered a rich, qualitative understanding of how AI shaped modern assessment practices.

Research Locale

The study was conducted in public and private high schools within Lanao del Sur, Philippines. This setting provided a diverse context to examine how different institutional environments influenced experiences.

Participants comprised six mathematics teachers selected through purposive sampling. To qualify, participants had to: (1) be currently teaching high school mathematics in the locale; (2) possess experience specifically in designing AI-generated assessments; and (3) be willing to participate in interviews. This sample size was deemed sufficient to reach data saturation and capture the shared essence of the phenomenon (Moustakas, 1994). All participants received details regarding the study and provided consent prior to involvement.

Research Instrument

The primary instrument was a semi-structured interview protocol, designed to balance structure and flexibility. It guided the conversation toward the study's objectives while allowing participants to elaborate on their unique experiences and narratives (Creswell & Poth, 2018).

The questions elicited insights into initial usage, perceived advantages, encountered difficulties, changes in practice, and professional growth. Key questions included:

1. In your opinion, what value and purpose do you see in using AI tools when creating tests or assessments for your students?
2. When you use assessments created or generated by AI, what step do you take? How do you design, adjust, or modify them to fit your needs, and use them in your teaching?
3. As you use AI tools to create assessments, what difficulties or problems do you encounter? And on the other hands what positive opportunities or advantages does this bring you to your work?
4. How do you make sure that assessments created by AI match what you aim to teach and what your students actually need to learn and practice?

To ensure clarity and relevance, the protocol underwent a pilot test with two qualified teachers not included in the final sample. Revisions were made based on feedback to enhance the instrument's ability to capture detailed narratives.

Data Collection Procedure

The data collection phase followed strictly to ethical guidelines and established qualitative research practices, ensuring the integrity and richness of the information gathered. After securing all necessary ethical approvals, the following systematic steps were undertaken:

Participant Identification and Informed Consent. Eligible teachers were identified via professional networks and outreach. They received comprehensive information about the study's purpose and their rights. Voluntary participation was emphasized, and written consent was obtained before any data gathering began.

Scheduling and Conducting Semi-Structured Interviews. Interviews were scheduled at the participants' preferred time and venue, conducted either in-person or via secure online platforms. Each session lasted approximately 60–90 minutes and was audio-recorded with permission. Detailed field notes were also taken to capture non-verbal cues and contextual observations.

Ensuring Data Quality and Saturation. Data collection continued concurrently with analysis until saturation was reached—when no new themes or insights emerged from further interviews. All data were stored securely and treated with strict confidentiality.

Data Analysis

The analytical phase of this phenomenological study systematically processed the rich, qualitative data gathered from the semi-structured interviews to uncover the essence of mathematics teachers' experiences with AI-generated assessments. Following the collection of data and transcription of all audio recordings, the analysis proceeded through a series of rigorous steps, primarily informed by Moustakas's (1994) modified Stevick-Colaizzi-Keen method of phenomenological analysis (Creswell & Poth, 2oth18). This approach was designed to transform participants' lived experiences into a textual description that captured the commonalities of the phenomenon.

The process involved systematic steps:

Horizontalization and Significant Statements. Initially, each transcribed interview was read multiple times to achieve a holistic understanding and to immerse the researcher in the participants' narratives. Following this, a process of horizontalization was undertaken, wherein every statement relevant to the participants' experiences with AI-generated assessments was extracted. These "significant statements" were treated as having equal initial value. Redundant or irrelevant statements were set aside.

Clustering into Themes. The identified significant statements were grouped into meaning units or preliminary themes. This involved reflecting on the underlying meaning of each statement and categorizing them based on shared concepts or ideas.

Textural and Structural Descriptions. From these meaning units, textural descriptions were developed for each participant, detailing what they experienced regarding AI-generated assessments. This involved articulating the verbatim and paraphrased accounts of their feelings, perceptions, and concrete actions. Concurrently, structural descriptions were formulated, describing how the phenomenon was experienced. This involved exploring the context and conditions that influenced their experiences, such as their interaction with the technology, the environment, or their prior beliefs.

Essence of the Phenomenon (Composite Description). Finally, the individual textural and structural descriptions were synthesized to construct a composite description (also known as the "essence" of the phenomenon). This comprehensive narrative integrated the common themes and unique perspectives, illustrating the overarching structure and shared meaning of mathematics teachers' experiences in designing and implementing AI-generated assessments. It represented the core understanding of the phenomenon, common to all participants.

To ensure the trustworthiness and credibility of the findings, member checking was employed. This involved returning the composite description or key themes to participants to verify that the interpretations accurately reflected their lived experiences. Additionally, peer debriefing was utilized, involving discussions with an experienced qualitative researcher to challenge assumptions and validate interpretations (Creswell & Poth, 2018). These strategies enhanced the rigor and transferability of the research outcomes.

The designation RMT stands for Research Participant. Mathematics Teacher. Each teacher possesses varying years of teaching experience, handles different grade levels, and has varying levels of exposure to digital tools and AI applications. Despite these differences, all participants regularly use AI tools in their instructional preparation, specifically for developing assessment materials. - RMT1 is a senior teacher with over 10 years of experience, handling junior high school mathematics. She often uses AI to generate supplementary test items and practice exercises to support students with diverse learning needs. - RMT2 has 7 years of teaching experience and teaches both junior and senior high school levels. He views AI primarily as a time-saving tool to manage his heavy workload and multiple teaching assignments. - RMT3 is a mid-career teacher with 5 years of experience, specializing in senior high school mathematics. He is one of the more technically proficient participants, often uploading curriculum guides and modules directly into AI tools for more accurate output. - RMT4 has 4 years of teaching experience and works in a school with limited technological resources. He relies on free, accessible AI tools to create varied assessment formats that he would otherwise spend hours developing manually. - RMT5 is a young teacher with 3 years of experience, teaching in a rural setting. He values the mobility of AI tools, allowing him to prepare assessments anytime using his mobile device, despite occasional connectivity challenges. - RMT6 has 6 years of experience and emphasizes pedagogical rigor. She is highly cautious when using AI, treating all generated content as a draft that requires extensive review and alignment with DepEd standards before use. Collectively, these teachers represent the range of experiences, challenges, and practices found among mathematics educators in the region. Their narratives provide the basis for understanding how AI is being integrated into assessment practices in local contexts.

RESULTS AND DISCUSSION

Using a phenomenological approach, interview responses from seven mathematics teachers were analyzed through thematic analysis. Significant statements were extracted from the participants' narratives, coded, and clustered into themes and subthemes. The findings are presented according to the research questions that guided the study.

Consistent with the study's theoretical framework, findings are interpreted through the lens of the Unified Theory of Acceptance and Use of Technology (UTAUT) by Venkatesh et al. (2003) and Pedagogical Content Knowledge (PCK) by Shulman (1986). UTAUT illuminates the factors that shape teachers' readiness and engagement with AI tools, while PCK highlights how teachers draw on their specialized knowledge of mathematics content and student learning to evaluate, adapt, and contextualize AI-generated outputs. Each research question is preceded by a summary table presenting the themes and their corresponding codes.

Research Question 1: Perceived Value and Purpose of AI Tools in Assessment

The first research question explored how mathematics teachers perceive the value and purpose of using AI tools to create assessments for their students. Across all six participants, a consistent and strongly held view emerged: AI tools are perceived as practical enablers of efficiency and creativity in assessment preparation, rather than as replacements for teacher judgment and expertise.

Table 1. RQ1: Perceived Value and Purpose

RQ1: Perceived Value and Purpose	***Theme 1.1: Time-Saving and Workload Reduction***	AI reduces time spent on writing test items
		Teachers redirect saved time to instruction and student support
		Mobile and anytime access increases convenience
		AI is perceived as free and accessible
	Theme 1.2: Generation of Varied and Creative Assessment Items	AI produces diverse question types teachers would not normally generate
		AI helps cover different levels of difficulty for diverse learners
		AI broadens teachers' repertoire of assessment strategies
		AI serves as a creative supplement, not a primary author
	Theme 1.3: AI as a Support Tool, not a Replacement	Teachers frame AI as a guide or assistant, not the decision-maker
		Teacher remains the final authority on assessment quality
		AI supports organization and alignment with competencies
		Professional judgment is indispensable even when using AI

Theme 1.1: Time-Saving and Workload Reduction

The most frequently and emphatically cited value of AI tools among all participants was the significant reduction in the time required to prepare assessments. Teachers in Lanao del Sur II manage large and diverse student populations, often with limited instructional support, making the ability to generate assessment items quickly a substantial professional benefit.

RMT1 described how AI helps teachers generate varied questions and performance tasks that serve as a guide, saving time that can be redirected toward direct instruction and student engagement. Similarly, Bagul A. articulated this value in practical terms:

> *"As a teacher in Marawi, I have a lot on my plate — big classes, multiple subjects, limited time. AI helps me come up with test questions faster... the biggest value is that it saves time so I can focus more on actually teaching my students rather than spending all night writing test items."*

RMT5 extended this perception by noting not only time savings but also mobility and accessibility, emphasizing that AI tools allow him to create assessments anytime and anywhere using any device. This perception aligns with UTAUT's construct of performance expectancy—the belief that using technology improves job performance.

This finding is strongly supported by related literature, which consistently identifies efficiency and workload reduction as the primary drivers for teacher adoption of AI in assessment. According to Owan et al. (2023) and Coskun & Alper (2024). AI automates routine tasks and significantly streamlines assessment workflows, allowing educators to focus more on instructional delivery and student support. In the Philippine context, local studies such as that conducted in Bacolod City confirmed that AI tools contributed greatly to improved teacher efficiency, especially in content generation and instructional preparation. Teachers' recognition of AI's ability to make work easier and more convenient also mirrors findings from Batac City, where educators reported positive perceptions regarding AI's usefulness, though actual use remained limited due to other factors. These results collectively validate that performance expectancy is the most influential factor shaping teachers' acceptance of AI, as predicted by the UTAUT framework.

Theme 1.2: Generation of Varied and Creative Assessment Items

Beyond time-saving, teachers recognized AI's capacity to produce diverse and creative question types that they might not independently generate. Haniya S. observed that AI introduces new ways to frame topics and helps ensure comprehensive coverage of key lessons. Harley S. similarly highlighted AI's utility in creating assessments with different levels of difficulty, which is especially useful for addressing the learning diversity present in her classroom.

RMT7 echoed this view, noting that AI gives teachers different types of questions to choose from. RMT2 reinforced this, stating that AI provides varied ideas for questions and activities that make assessment preparation easier and more comprehensive. These perceptions reflect an appreciation for AI as a creative supplement—a tool that broadens the repertoire of assessment strategies available to mathematics teachers.

This theme connects directly to the concept of Pedagogical Content Knowledge (PCK), specifically the component related to assessment strategies suitable for specific content areas (Shulman, 1986). As noted in the review of related literature, generative AI enables the design of varied, context-specific assessment formats that support differentiated instruction. This capability aligns with the recommendations of the National Council of Teachers of Mathematics (NCTM, 2024), which advocates for diverse assessment methods to measure both procedural fluency and conceptual understanding. Findings from a systematic review by Rane (2023) further confirm that AI's strength lies in its ability to generate multiple test versions and varied item types, which prevents rote learning and addresses diverse learning needs.

Local studies in the Philippines also support this finding. In Northern Samar, research revealed that mathematics teachers value AI for expanding their assessment strategies, though their ability to fully utilize these features is influenced by their training and experience. Teachers in the present study similarly reported that AI helps them create assessments for different levels of difficulty—a feature highly valued in multigrade or mixed-ability classrooms common in many Philippine schools.

Theme 1.3: AI as a Support Tool, not a Replacement

Importantly, all participants were explicit and consistent in framing AI as a support tool that assists rather than replaces the teacher. RMT2 stated directly that the purpose of using AI is not to replace the teacher, but to support teachers in preparing assessments that are organized, updated, and aligned with learning competencies. RMT6 reinforced this position plainly: At the end of the day AI is just a tool. The teacher still makes the final call.

This perspective is deeply consistent with the PCK framework, which foregrounds the indispensable role of teacher expertise in ensuring that assessments are pedagogically sound and responsive to students' specific needs.

This finding is among the most significant in this study and resonates strongly with both foreign and local literature. Ollington (2025) emphasized that while technology assists, genuine pedagogical insight and subject matter expertise remain critical to ensuring assessment validity and reliability. This is precisely what participants meant when they insisted that *"the teacher still makes the final decision."*

In the Philippine setting, this view is even more pronounced. A phenomenological study of Filipino educators explicitly concluded that while AI supports teaching tasks such as assessment, teachers remain essential for contextualization, ethical use, and alignment with learning goals. Another study involving ESL teachers similarly found that Filipino educators are cautious users who treat AI as a draft generator—always requiring validation, modification, and teacher input before actual use. This practice reflects the principle of *"human-in-the-loop"* assessment design widely discussed in literature, which posits that AI should function as a support, never the authority, in educational decision-making.

From a theoretical perspective, this reinforces that PCK—the specialized knowledge of how to teach specific subjects—is not replaceable by technology; rather, teachers integrate AI tools *within* their existing professional knowledge structures.

Research Question 2: Processes in Designing, Customizing, and Implementing AI-Generated Assessments

The second research question examined the specific steps and processes mathematics teachers follow when designing, customizing, and implementing AI-generated assessments. The findings reveal a consistent multi-stage workflow that reflects a deliberate, teacher-mediated approach rather than direct or uncritical use of AI output.

Table 2. RQ2: Design and Implementation Process

RQ2: Design and Implementation Process	***Theme 2.1: Prompt Specification and Purposeful Generation***	Teachers specify topic, level, and goals before generating content
		Number and type of questions are explicitly defined
		Teaching modules uploaded as reference for AI generation
		Purposeful prompting embeds pedagogical goals into AI output
	Theme 2.2: Critical Review and Verification of AI Output	All participants review AI output before classroom use
		Questions checked for accuracy and lesson alignment
		Errors, mismatches, and unclear items are identified and removed
		Review is treated as a non-negotiable professional responsibility
	Theme 2.3: Contextual Adaptation and Localization	AI content adapted to local culture and context of Marawi City
		Language and examples adjusted to fit students' backgrounds
		Difficulty levels modified to match students' actual proficiency
		Teachers add their own contextualized examples and items
	Theme 2.4: Iterative Finalization and Deployment	Assessments deployed as quizzes, formative tests, or practice tasks
		Teachers reflect on what worked and improve for the next use
		Final verification of alignment performed before administration
		Iterative refinement develops over time with continued AI use

Theme 2.1: Prompt Specification and Purposeful Generation

Several participants described beginning the assessment design process by providing the AI with clear parameters and specific details before generating content. RMT3 elaborated: I start by giving clear details about the topic, level, and goals before generating any content. RMT4 similarly noted that he specifies the types of exam, the number of questions, and even the structure of items. He also described uploading his teaching module directly into the AI tool so that it can use the module as a reference when generating questions.

This stage reflects a level of technological fluency and intentionality that goes beyond passive consumption of AI output. Teachers who specify prompts carefully are functioning as designers of the generation process itself, embedding their pedagogical goals into the parameters they submit to the AI system.

This finding aligns with emerging frameworks such as the AI Assessment Scale (AIAS), which advocates structured prompting and clear definition of objectives before using AI tools. According to Gorgun et al. (2025), effective AI assessment design begins with the teacher providing clear guidance, standards, and reference materials—essentially *teaching the AI what to produce*.

In the Philippines, this practice mirrors findings from studies on teachers' knowledge, attitudes, and practices (KAP) regarding AI. For instance, Filipino teachers reported that they never generate content without first defining their lesson objectives and standards, often uploading modules or curriculum guides to ensure alignment. This demonstrates that teachers are not merely users but *designers* of AI use, applying their PCK to guide the technology toward their instructional needs.

This stage also reflects effort expectancy in UTAUT: while teachers invest effort in specifying prompts, they perceive this effort as worthwhile because it improves the quality and relevance of the output.

Theme 2.2: Critical Review and Verification of AI Output

Without exception, all six participants described reviewing all AI-generated content before use in the classroom. RMT4 described reviewing the questions carefully to ensure alignment with lesson objectives and appropriateness for students' level of understanding. RMT7 articulated a structured review process: she first identifies the lesson objectives and competencies that need to be assessed, then asks the AI to generate sample questions, and subsequently reviews, edits, and simplifies the questions before use.

RMT6 summarized her process concisely: First I check all the questions carefully. Second I edited them to match our lesson. This universal practice of critical review strongly reflects the PCK construct—teachers applying their deep knowledge of the mathematics curriculum to evaluate and filter AI output.

This practice is perhaps the most documented in related literature. Multiple studies confirm that teacher review and validation are essential steps; AI output is treated as a draft or suggestion, never a final product. A systematic review by Khan et al. (2026) highlighted that the *"human-in-the-loop"* model—where teachers actively review, verify, and validate—is the dominant and most effective approach in real-world implementation.

Local studies in the Philippines reinforce this strongly. In Batac City, teachers reported that they always check for errors, alignment, and appropriateness, viewing AI as a starting point only. Similarly, research in Cebu showed that pre-service teachers, while enthusiastic about AI, always verify content accuracy before use. This practice ensures reliability and validity, which are core concerns in assessment as emphasized by NCTM (2024).

Theme 2.3: Contextual Adaptation and Localization

A particularly notable finding was the emphasis all participants placed on adapting AI-generated content to fit the local cultural, linguistic, and social context of their students in Marawi City, Lanao del Sur I, and Lanao del Sur II. RMT1 described editing difficult terms, adjusting examples to fit the local context, and simplifying instructions when necessary. RMT3 similarly described adjusting the language and examples to fit our local setting here in Marawi.

RMT5 likewise described adjusting difficulty, wording, and examples to fit local context and correcting any errors found. This consistent pattern of localization highlights that AI-generated assessments are treated as drafts requiring substantial teacher input before they are pedagogically appropriate for the specific student population being served.

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 7 (July 2026)

This theme is highly significant and directly addresses a gap identified in international literature: generic AI content often lacks cultural relevance, local examples, or familiarity with students' lived experiences. Teachers in this study demonstrated a sophisticated application of PCK—specifically knowledge of their students' cultural background, language proficiency, and community context—to make assessments meaningful and relevant.

This aligns with the Philippine context where localization and contextualization of curriculum are mandated by the Department of Education (DepEd). Local studies have shown that Filipino teachers prioritize adapting materials to fit local culture and experiences, and this study confirms that they apply the same rigorous standards to AI-generated assessments. This step ensures that assessments are not only valid technically but also culturally responsive and accessible, a principle emphasized in both local and international educational standards.

Theme 2.4: Iterative Finalization and Deployment

Following review and adaptation, teachers finalize the assessment and deploy it in varied instructional contexts. Participants described using AI-generated and adapted assessments as quizzes, formative assessments, practice activities, and summative tests. RMT4 noted using adapted materials as quizzes, exams, practice activities, or tools to provide targeted feedback and instruction. RMT1 mentioned taking note of what worked and what didn't for next time, indicating a reflective and iterative relationship with AI-assisted assessment design. This iterative practice resonates with UTAUT's concept of facilitating conditions.

This reflective and iterative use is consistent with recommendations in literature that assessment design is an ongoing process, not a one-time task. Studies show that effective AI integration improves over time as teachers learn what works, refine their prompts, and build a library of adapted items.

In the Philippines, this aligns with the cyclical nature of assessment in the K-12 curriculum, where reflection and improvement are built into teaching practice. Teachers' practice of *"taking note of what worked"* also supports professional growth, which is linked to facilitating conditions—as teachers gain more experience and resources, their use becomes more effective and sophisticated.

Research Question 3: Challenges and Opportunities in AI Integration

The third research question examined what challenges and opportunities mathematics teachers identify in integrating AI into their assessment creation workflows. Participants articulated a nuanced picture in which significant practical difficulties coexist with substantial professional benefits.

Table 3. Themes and Codes — RQ3: Challenges and Opportunities

RQ3: Challenges and Opportunities	***Theme 3.1: Inaccuracy and Misalignment of AI Content***	AI generates factually incorrect or unreliable questions
		Generated items sometimes too difficult, vague, or generic
		Content does not always match local curriculum or student context
		Teachers must spend extra time correcting and revising AI output
	Theme 3.2: Connectivity and Access Barriers	Unstable internet connection limits access to AI tools
		Rural settings face greater connectivity disadvantages
		Dependence on internet creates unpredictability in tool availability
		Infrastructure gaps constrain equitable use of AI in education

	Theme 3.3: Enhanced Efficiency and Reduced Workload (Opportunity)	AI significantly reduces time and effort in assessment preparation
		Efficiency frees teachers for higher-order instructional activities
		AI manages heavy workload more sustainably
		Teachers express consistent appreciation for time-saving benefits
	Theme 3.4: Differentiation and Creative Assessment Design (Opportunity)	AI enables creation of multi-level assessments for diverse learners
		Teachers produce more creative and varied question formats
		AI introduces new ways to frame mathematical topics
		Technology encourages pedagogical innovation among teachers

Theme 3.1: Inaccuracy and Misalignment of AI-Generated Content

The most consistently cited challenge across participants was the problem of inaccurate, inappropriate, or contextually misaligned questions generated by AI tools. RMT1 noted that some generated questions are too advanced, unclear, or sometimes not fully accurate, and that the content does not always match the cultural background or learning situation of students. RMT6 described AI as sometimes generating inaccurate, repetitive, or overly difficult questions that are too general and do not fully match the local context or the actual learning competency.

RMT3 observed that answers are not always 100% correct, while RMT2 identified common issues including factual mistakes, generic content, and unclear questions. RMT5 remarked that AI is sometimes unreliable and gives out questions that cannot be found in the module. These experiences reinforce the importance of the teacher's role as a critical reviewer, consistent with both PCK and the broader literature on human-in-the-loop approaches to AI in education.

This challenge is widely documented in research. Rane (2023) and Khan et al. (2026) both warned that AI tools often produce errors, logical inconsistencies, or mathematically incorrect solutions—risks that are especially serious in mathematics where precision is essential. Studies in higher education similarly noted that while AI improves consistency, it lacks contextual understanding and may produce content that is technically correct but educationally inappropriate.

In the Philippine setting, this challenge is amplified by the need for strict alignment with DepEd competencies. Local studies from Batac City and Bacolod confirmed that teachers frequently encounter AI content that is too advanced, too generic, or simply not aligned with the Philippine curriculum. This finding underscores that while AI is powerful, it lacks the deep contextualized knowledge that teachers possess—a core distinction emphasized in the PCK framework.

Theme 3.2: Connectivity and Access Barriers

A second recurring challenge was limited or unstable internet connectivity, which directly constrains the practical usability of AI tools. RMT2 cited internet connection and limited access to technology as challenges. She described internet connection as not always reliable which makes it frustrating sometimes. Rmt7., who resides in a rural area, noted that unstable internet connection makes him unable to use AI at times.

This finding is particularly significant in the context of Lanao del Sur II, where infrastructure limitations represent a concrete barrier to equitable access to educational technologies. Within the UTAUT framework, this challenge relates directly to facilitating conditions—where infrastructure is inadequate, even motivated and capable teachers face systemic barriers to meaningful AI integration.

This aligns perfectly with local studies. In Northern Samar, research showed that while teachers had positive attitudes, actual usage was limited by lack of reliable internet and limited access to devices. Similar findings emerged in studies of pre-service teachers in Cebu, where infrastructure gaps were cited as major obstacles. These findings confirm that technology adoption is not only about attitude or skill but also about the *conditions* that enable use—one of the four key constructs in UTAUT.

Theme 3.3: Enhanced Efficiency and Reduced Workload (Opportunity)

Despite the challenges described above, participants expressed strong and consistent appreciation for the professional opportunities that AI tools create. RMT1 noted that AI reduces workload, helps generate ideas quickly, and allows teachers to create more varied and interactive assessments. RMT3 highlighted that AI lessens workload and gives more time for lesson planning and student support.

RMT2 framed this efficiency as an enabler of better teaching: it saves me a lot of time and it pushes me to use more creative question types that challenge my students to think critically. RMT5 described AI as offering huge time savings and making heavy workloads much more manageable. These perceptions directly reflect UTAUT's performance expectancy construct.

This opportunity is the strongest driver identified in both foreign and local literature. A systematic review of 159 studies confirmed that the most consistent benefit reported is time efficiency, which allows teachers to shift energy from administrative tasks to instruction and student interaction. In the Philippines, where class sizes are often large and workloads heavy, this benefit is even more valued. Research in Bacolod City found that teachers specifically appreciated AI for reducing burnout and making workload sustainable.

This also connects to the theoretical concept of PCK enhancement: by reducing time spent on routine creation, teachers have more mental space to apply their specialized knowledge to improving assessment quality and instruction.

Theme 3.4: Differentiation and Creative Assessment Design (Opportunity)

Participants also identified AI as a valuable resource for creating assessments that address diverse learner needs. RMT4 specifically cited AI's usefulness in creating activities for students with varying abilities. RMT2 highlighted AI's capacity to make it easy to create assessments for different learning levels. RMT1 noted that AI encourages teachers to become more innovative and explore different teaching strategies that can improve student learning.

This finding aligns with literature highlighting AI's potential to support personalized and differentiated assessment—key goals in modern education. According to Coskun & Alper (2024), AI can rapidly generate multiple versions, varied difficulty levels, and different formats, making differentiation feasible even in large classes.

In the Philippine context, this is especially valuable because of the inclusive education policy that requires teachers to address diverse learning needs. Local studies showed that Filipino teachers struggle to create differentiated assessments manually, making AI's capacity a highly valued resource. This benefit supports the idea that AI can help teachers implement best practices that would otherwise be too time-consuming.

Research Question 4: Ensuring Alignment of AI-Generated Assessments with Instructional Goals and Student Needs

The fourth research question explored how teachers ensure that AI-generated assessments align with their instructional goals and students' learning needs. The findings reveal a multi-layered alignment process grounded in professional knowledge, curriculum standards, and direct knowledge of students.

Table 4. Themes and Codes — RQ4: Ensuring Alignment

RQ4: Ensuring Alignment	***Theme 4.1: Curriculum and Competency Verification***	AI output compared against DepEd curriculum and lesson competencies
		Learning objectives used as the primary filter for item selection
		Items outside lesson scope are removed or rewritten
		Curriculum alignment is a systematic, non-optional step for all participants
	Theme 4.2: Student-Centered Adaptation Based on Contextual Knowledge	Assessments tailored to students' proficiency level and background
		Language and phrasing adjusted for comprehension and relevance
		Teacher's intimate knowledge of students guides final item selection
		AI cannot replicate teacher's contextualized knowledge of students
	Theme 4.3: Peer Consultation and Collaborative Review	Colleagues consulted to review AI-generated assessments
		Peer review adds an additional quality-assurance layer
		Collaborative approach reflects established professional norms
		AI output treated with same scrutiny as externally sourced materials

Theme 4.1: Curriculum and Competency Verification

All participants described checking AI-generated assessments against official curriculum standards and lesson competencies as a fundamental step in the alignment process. RMT1 stated that she always compares generated questions with the curriculum competencies and lesson outcomes, checking whether questions measure the skills and knowledge that students are expected to learn. RMT3 described using his learning objectives as my filter when reviewing the generated questions, matching all items against the DepEd curriculum.

RMT6 articulated a systematic approach: I always start by defining clear learning objectives and share these directly when generating content. I check all items against my lesson plans, teaching materials, and official curriculum standards to ensure everything is relevant. RMT2 described comparing AI output with her lesson plan and adjusting the questions to match the needs of her students. These practices reflect the direct application of PCK—teachers drawing on their subject-matter expertise to evaluate the suitability of AI-generated content.

This rigorous verification is consistent with literature emphasizing that alignment with standards is non-negotiable. According to the AI Assessment Scale framework, alignment with learning outcomes is the first and most critical principle. NCTM (2024) similarly stresses that assessments must directly reflect curriculum goals and competencies.

In the Philippines, where curriculum alignment is strictly monitored, this practice is even more pronounced. Local studies consistently report that Filipino teachers treat alignment as the primary criterion for accepting or rejecting AI content, often using curriculum guides as their reference. This confirms that teachers are applying their PCK—specifically knowledge of curriculum and standards—to ensure assessment validity.

Theme 4.2: Student-Centered Adaptation Based on Contextual Knowledge

Beyond curriculum alignment, participants consistently described tailoring assessments to the specific characteristics, abilities, and backgrounds of their students. RMT4 noted that she considers the students' learning level, experiences, and classroom performance before finalizing the assessment. RMT2 described relying on his personal knowledge of students: I know their level, their background, what they struggle with. If a question doesn't feel right for them, I rewrite it or throw it out entirely.

This theme reflects the essence of PCK as applied to AI integration: teachers are not simply aligning assessments with abstract curriculum standards but are calibrating them against their intimate and contextualized understanding of their specific students—a form of knowledge that AI tools fundamentally cannot replicate.

This finding directly supports the theoretical distinction between technological knowledge and pedagogical content knowledge. As Shulman (1986) defined it, PCK includes understanding what makes topics easy or difficult, knowing common misconceptions, and understanding student needs—knowledge that is deeply personal and contextual. Literature confirms that AI lacks this kind of deep, situated understanding, making teacher adaptation essential.

Local studies in the Philippines mirror this finding. Teachers consistently report that their knowledge of their students—their culture, language, and learning challenges—is the most important filter they apply to any material, including AI-generated content. This practice ensures that assessments are not only valid but also *fair and accessible*.

Theme 4.3: Peer Consultation and Collaborative Review

Some participants described engaging colleagues in the review of AI-generated assessments as an additional quality assurance measure. RMT5 noted that she sometimes asks colleagues to review assessments before administering them. This practice suggests an emerging collaborative dimension to AI-assisted assessment design, wherein professional consultation serves as an additional check on the quality and appropriateness of AI output. Its application to AI-generated assessments indicates that teachers are treating AI output with the same critical scrutiny they would apply to any externally sourced instructional material.

This collaborative practice aligns with literature advocating for professional dialogue and shared standards in assessment design. It also reflects social influence from the UTAUT framework—where teachers' practices are shaped by consultation with peers and professional norms.

In the Philippine educational culture, where collaboration and community are highly valued, this practice is common and effective. Local studies have shown that peer review is a standard way to ensure quality and alignment, and teachers extend this practice to AI-generated materials as well

Synthesis: Composite Description of the Phenomenon

Taken together, the findings across all four research questions reveal a coherent and nuanced portrait of mathematics teachers' lived experiences with AI-generated assessments in the divisions of Marawi City, Lanao del Sur I, and Lanao del Sur II , Philippines. Teachers are neither uncritical adopters nor resistant skeptics; rather, they are pragmatic, reflective practitioners who engage with AI tools selectively and deliberately, always filtering AI output through the lens of their professional knowledge and their deep understanding of their students and context.

The essence of the phenomenon can be described as follows: Mathematics teachers in this context experience AI as a valuable but imperfect ally in the demanding work of assessment design. They value AI primarily for its capacity to save time and generate varied assessment ideas, but they consistently position themselves as the authoritative judges of what is educationally appropriate. Their engagement with AI is characterized by active mediation—reviewing, editing, contextualizing, and verifying—rather than passive acceptance. The challenges they encounter, including AI inaccuracies, contextual misalignment, and connectivity barriers, reinforce rather than diminish their sense of professional agency, as these challenges necessitate the exercise of the very expertise that defines their role as educators.

This composite description resonates strongly with both the UTAUT and PCK frameworks. UTAUT helps explain why teachers adopt AI tools in the first place—performance expectancy and effort expectancy create motivation to engage, while facilitating conditions shape the practical parameters of engagement. PCK explains how teachers engage: they bring their specialized knowledge of mathematics content, curriculum standards, and student learning to bear on the evaluation and transformation of AI output, ensuring that technological convenience never comes at the cost of pedagogical quality.

The findings also align with existing literature on human-in-the-loop approaches to AI in education (Ollington, 2025; Gorgun et al., 2025), which emphasizes that AI should support rather than supplant teacher judgment. This study provides empirical evidence from a specific and underrepresented context—Muslim-majority, under-resourced communities in Mindanao—that this principle is not merely theoretical but is actively practiced by teachers navigating the real-world complexities of AI integration.

Furthermore, this study confirms and extends local research in the Philippines. It supports findings that Filipino teachers are cautious, deliberate users who view AI as a support, not a replacement; who prioritize alignment and contextualization; and whose use is shaped by infrastructure, training, and professional values. It also highlights that for mathematics teachers, assessment design remains fundamentally a pedagogical act—one that technology can assist but never fully automate.

In conclusion, the lived experience of mathematics teachers in Lanao del Sur II demonstrates that effective AI integration is not about adopting technology, but about integrating technology *into* professional practice. It is a practice defined by the intersection of efficiency and expertise, innovation and tradition, and technological capability and pedagogical wisdom.

CONCLUSION

Based on the findings and synthesis of the study, the following conclusions are drawn:

1. ***Mathematics teachers in Lanao del Sur are pragmatic, deliberate, and critical users of AI.*** They do not view AI as a substitute for their professional role but as a powerful assistant that enhances their work. Their acceptance is driven primarily by *Performance Expectancy*—the proven ability of AI to save time and improve efficiency—while their sustained use is regulated by their *Pedagogical Content Knowledge*, which ensures educational quality remains paramount. This confirms that technology adoption in education is not merely about access or skill, but about how well the tool serves the pedagogical goals defined by the teacher.
2. ***Effective use of AI in assessment is defined by a "Human-in-the-Loop" model.*** The study concludes that there is no direct or uncritical use of AI-generated content. The essence of the practice is *mediation*: teachers act as designers, reviewers, editors, and validators. The workflow is not automatic but iterative and reflective, proving that assessment design remains fundamentally a pedagogical act rather than a technical one. The effort invested in reviewing and adapting AI output is viewed as essential professional responsibility, aligning with the principles of validity and reliability emphasized by the National Council of Teachers of Mathematics (NCTM) and the Department of Education (DepEd).
3. ***Contextualization and localization are critical filters for AI integration.*** The study reveals that generic AI content is insufficient for the specific cultural and educational landscape of Marawi City and Lanao del Sur. Teachers' ability to adapt materials to local context, language, and student experiences is a unique dimension of their PCK that AI cannot replicate. Consequently, AI tools are most valuable when they provide a draft or framework, while the teacher infuses it with local relevance and cultural responsiveness—mandated standards in the Philippine K-12 curriculum.

4. ***Integration is significantly constrained by infrastructure and content limitations.*** While the perceived benefits are high, the full potential of AI is limited by two major factors: (a) the unreliability of internet connectivity and access to devices, which relates to *Facilitating Conditions* in the UTAUT framework, and (b) the tendency of AI to generate inaccurate, misaligned, or overly complex mathematical content. These challenges reinforce the teacher's role but also create inefficiencies that can reduce the perceived value of the technology when support systems are lacking.
5. ***Alignment is achieved through the intersection of curriculum knowledge and student knowledge.*** Ensuring that assessments measure what they intend to measure and meet student needs is not a technical setting but a professional judgment. Teachers ensure alignment by anchoring AI outputs in official curriculum guides and, more importantly, in their personal understanding of their students' capabilities and cultural backgrounds. This confirms that PCK is the dominant framework guiding AI integration, with technology serving only as a medium for content generation.
6. ***The experience of teachers in this region validates that AI integration is a process of professional growth.*** Participants demonstrated that engagement with AI tools actually deepens their pedagogical reflection. By constantly evaluating, editing, and adapting AI materials, teachers are forced to clarify their learning objectives and analyze their students' needs more explicitly. Thus, AI integration does not replace expertise; it activates and expands it.

The journey of mathematics teachers in Lanao del Sur I and II in navigating AI-generated assessments reflects a broader global shift in education: the intersection of human wisdom and technological capability. This study confirms that while AI brings efficiency and innovation, the heart of assessment remains the teacher—their expertise, their judgment, and their deep commitment to the unique needs of their learners. As technology continues to evolve, the partnership between teacher and tool will remain the defining factor in educational quality.

Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

References

Alwaqdani, M. (2025). Investigating teachers' perceptions of artificial intelligence tools in education: Potential and difficulties. *Education and Information Technologies*, *30*, 2737–2755. https://doi.org/10.1007/s10639-024-12903-9

Baker, R. S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Applications to educational software. *Journal of Educational Data Mining*, *6*(1), 6–12. https://doi.org/10.15215/aupress/9781771991490.01

Cabato, J. U. (2025). From awareness to practice: Exploring the knowledge, attitudes, and practices of secondary ESL teachers in the Philippines toward ChatGPT in education. *LatIA*, *3*, Article 60. https://doi.org/10.62486/latia2025360

Calubag, L. V. (2026). Professional correlates of AI-TPACK levels among secondary mathematics teachers in Northern Samar, Philippines. *Asian Research Journal of Mathematics*, *22*(3), 47–62. https://doi.org/10.9734/arjom/2026/v22i31055

Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2025). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Education and Information Technologies*, *30*, 6667–6685. https://doi.org/10.1007/s10639-024-13094-z

Co, S. J. (2025). Artificial intelligence in Philippine education: A narrative review of applications, perceptions, and challenges. *International Research Journal of Science, Technology, Education & Management*, *5*(2), 25. https://doi.org/10.5281/zenodo.15902950

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

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*, *10*, Article 1665205. https://doi.org/10.3389/feduc.2025.1665205

Gorgun, G., Bulut, O., & Tan, B. (2025). Designing classroom assessments with generative AI: A teacher-in-the-loop framework. *Taylor & Francis Group*, *73*(1), 234–252.

Holmes, W., & Porayska-Pomsta, K. (2023). *The ethics of artificial intelligence in education*. Routledge. https://doi.org/10.4324/9780429329067

Jesus, J. B., & Caumeran, R. R. (2026). Human-centered AI in education: Educators' perspectives on teacher roles, ethics, and pedagogical value. *International Journal on Studies in Education*, *8*(1), 142–154. https://doi.org/10.46328/ijonse.5748

Khan, W., Topham, L., Jones, N., Atherton, P., Al-Shabandar, R., Kolivand, H., et al. (2026). Auto-assessment of assessment: A human-in-the-loop AI framework addressing policy gaps in academic assessment. *PLOS ONE*, *21*(4), Article e0346815. https://doi.org/10.1371/journal.pone.0346815

Lin, D., Shuang, X., & Zongben, X. (2025). Teacher-student-AI collaborative classroom: The platform and practice of AI-empowered university mathematics education. *Frontiers of Education in China*, *20*(4), 477. http://doi.org/10.3868/s110-020-025-0025-8

Malacapay, M. (2025). AI literacy, preparedness, and use motives of elementary education pre service teachers of a state university in the Philippines. *Edelweiss Applied Science and Technology*, *9*(9), 1120–1136. https://doi.org/10.55214/2576-8484.v9i9.10059

Mekadenaumporn, N. (2025). *Can AI universally help students and teachers?: A cross-cultural investigation of perceptions of AI's potential in education* (Research Programme Report). Nanyang Technological University. https://doi.org/10.37985/jer.v6i4.2846

Meylani, R. (2025). AI-powered assessments in mathematics education: A systematic review of contemporary research literature. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, *66*, 3642–3674. https://doi.org/10.53444/deubefd.1527781

National Council of Teachers of Mathematics. (2024). *Principles to actions: Ensuring mathematical success for all* (2nd ed.). NCTM.

Ollington, N. (2025). Challenging assumptions: AI and teacher assessment identity formation. *Australian Journal of Teacher Education*, *50*(2), 56–73. https://search.informit.org/doi/10.3316/informit.T2025083100014001596448297

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(8), Article em2307. https://doi.org/10.29333/ejmste/13428

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2012). *Knowing what students know: The science and design of educational assessment*. National Academies Press. https://doi.org/10.17226/10019

Rane, N. (2023). *Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: Roles and challenges in solving mathematical problems* (SSRN Scholarly Paper No. 4603237). http://dx.doi.org/10.2139/ssrn.4603237

Rubio, R. B. T., & Pastor, M. J. D. (2026). School leaders' and teachers' levels of awareness and use, and perspectives on AI integration in education. *American Journal of Education and Technology*, *5*(1), 62–67. https://doi.org/10.54536/ajet.v5i1.6768

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Silagan, B. L. D., & Tumapon, T. T. (2025). Technological competence, training and support, attitude towards AI, and teachers' acceptance. *Psychology and Education: Multidisciplinary Journal*, *36*(8), 941–964. https://doi.org/10.70838/pemj.360809

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Zawacki-Richter, O., Marín, V., Bond, M., & Gouverneur, D. (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*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhao, C. (2024). AI-assisted assessment in higher education: A systematic review. *Journal of Educational Technology and Innovation*, *6*(4). https://doi.org/10.61414/jeti.v6i4.209