

An AI-Based Training Management Model for Enhancing Elementary School Teachers' Ability to Construct Numeracy Literacy Questions.

Siti Ma'rifatin¹, Nurul Ulfatin², Juharyanto³, Raden Bambang Sumarsono⁴

¹ Universitas Negeri Malang, Indonesia: siti.marifatin.2201329@students.um.ac.id

² Universitas Negeri Malang, Indonesia: nurul.ulfatin.fip@um.ac.id

³ Universitas Negeri Malang, Indonesia: juharyanto.fip@um.ac.id

⁴ Universitas Negeri Malang, Indonesia: raden.bambang.fip@um.ac.id

ARTICLE INFO

Keywords:

Training Management Model,
Numeracy Literacy,
Artificial Intelligence,
Teacher Competence.

Article history:

Received 2026-05-01

Revised 2026-08-06

Accepted 2026-09-22

ABSTRACT

Curriculum expectations increasingly require primary school teachers to design reasoning-based numeracy literacy assessments aligned with higher-order thinking skills (HOTS), yet many educators lack the capacity to do so. Although generative artificial intelligence (AI) tools such as ChatGPT are increasingly adopted in education, validated training management models that systematically embed AI support across the assessment-development cycle remain scarce. To address this gap, this study developed and evaluated NULIQFORM, an AI-assisted training management model integrating ChatGPT-supported scaffolding across assessment planning, question development, and evaluation. Utilizing a Research and Development (R&D) approach grounded in the ADDIE framework, the model was implemented and evaluated via a quasi-experimental one-group pretest-posttest design involving 31 primary school teachers purposively selected from Grades 2 and 5 across Islamic and public schools in Waru Subdistrict, Indonesia. Feasibility was triangulated through evaluations by participants and the head of the Teacher Working Group (KKG), while effectiveness was evaluated through pre- and post-training competence scores analyzed descriptively and inferentially. Results demonstrate that NULIQFORM was rated feasible with minor revisions, and its implementation produced a statistically significant improvement in teachers' assessment-development competence ($t = -6.147$, $p < 0.001$; $d = 1.10$), with mean scores rising from 7.03 to 9.35. Given the single-group design and localized sample, these findings provide preliminary evidence of short-term effectiveness and offer a practical framework for embedding generative AI into teacher professional development.

This is an open-access article under the CC BY SA license.



Corresponding Author:

Siti Ma'rifatin

Universitas Negeri Malang; siti.marifatin.2201329@students.um.ac.id

INTRODUCTION

Indonesian students' performance on the Programme for International Student Assessment (PISA) consistently remains significantly below the OECD average, revealing systemic challenges in foundational cognitive domains where only a small proportion of learners reach minimum proficiency in numeracy and almost none demonstrate advanced mathematical reasoning (OECD, 2023). Educational literature underscores that this persistent performance gap is fundamentally linked to daily classroom assessment practices: when teachers rarely construct evaluation items that demand complex reasoning, critical analysis, and real-world problem-solving, students are deprived of essential opportunities to develop, internalize, and demonstrate higher-order thinking skills (HOTS), which in turn severely constrains their overall academic achievement on global reasoning-oriented assessments such as PISA (Blömeke et al., 2022; Febrianti et al., 2025; Hidayanto et al., 2026; Susilawati et al., 2023; Yang & Kaiser, 2022). Consequently, traditional assessment paradigms that prioritize mechanical algorithm execution fail to prepare students for modern analytical challenges (Hariadi et al., 2025; Hidayati et al., 2025; Novrianti & Sari, 2025; Shobihah, 2025). Closing this critical educational gap therefore necessitates a systemic shift toward strengthening teachers' pedagogical competence to design classroom assessments that intentionally cultivate, evaluate, and promote higher-order thinking skills (HOTS), rather than continuing to rely heavily on routine procedural recall and low-level cognitive items that reinforce passive learning patterns (Fitriana et al., 2026; Reimers et al., 2020). Ultimately, reshaping assessment design serves as a pivotal mechanism for elevating nationwide mathematical literacy and educational equity.

Classroom assessment practices across Indonesian primary schools remain predominantly dominated by routine, lower-order questions, with educators rarely constructing contextual numeracy literacy tasks that require cognitive processing beyond basic procedural recall and rote calculation (Jelatu et al., 2019; Safi'i & Amar, 2019; Sukenti et al., 2024; Tiok Wijanarko et al., 2025). This widespread reliance on low-level items stems partly from persistent constraints in teacher pedagogical content knowledge and limited mastery of test construction methodologies. Although targeted professional teacher training has long been recognized as a primary intervention strategy for building and elevating this essential instructional competence (Am et al., 2026; Asiyah et al., 2021; Bahrudin et al., 2026; Perla et al., 2025; Sumarsono et al., 2019), existing professional development models exhibit systemic fragmentation across multiple structural dimensions. Specifically, conventional training initiatives are typically delivered through short-term, one-off workshops that operate in isolation from prior empirical needs analyses, offer minimal guided hands-on practice in constructing complex reasoning-based items, and completely lack a systematic evaluation stage to verify whether the intended assessment competencies have actually been acquired or implemented in real classrooms (Arif et al., 2025; Kariyanto et al., 2025; Nurwahyuni & Za, 2025; Umar et al., 2025). Consequently, training outcomes remain highly superficial and difficult to sustain over time, failing to translate into meaningful, long-term transformations in teachers' actual classroom assessment practices or student learning outcomes.

Recent advances in artificial intelligence (AI) technologies present promising opportunities to reinvent and significantly strengthen this professional training process (Chairawati et al., 2026; Hakim et al., 2026; Lodge et al., 2024; Wale, 2024). Emerging generative AI tools possess the technical capability to actively support educators throughout the complex item-development pipeline by automatically generating context-rich problem scenarios rooted in realistic everyday situations, proposing diverse item variations mapped across distinct cognitive domains, and delivering immediate, formative

diagnostic feedback. This real-time feedback loop assists teachers in critically analyzing, revising, and elevating draft questions toward authentic higher-order reasoning demands and mathematical literacy standards. Rather than serving as a passive repository of static, pre-packaged test items, generative AI functions as a dynamic, interactive, on-demand scaffold that collaborates with teachers during the active cognitive process of item construction (Hanis & Wahyudin, 2024; Pham & Sampson, 2022). By reducing the cognitive load associated with drafting complex narrative contexts and aligning items with rigorous taxonomy levels, AI tools enable educators to focus on pedagogical refinement and cognitive alignment (Humaida et al., 2026; Jomezai et al., 2025; Syukur et al., 2024). Furthermore, this interactive assistance empowers teachers to experiment with innovative problem formats, refine distractor options in multiple-choice questions, and construct rubric criteria tailored to complex mathematical tasks. Ultimately, leveraging AI as a cognitive assistant holds immense potential to bridge skill gaps and enhance teacher self-efficacy in designing sophisticated numeracy assessments.

However, existing scholarly literature exhibits significant limitations in scope, as prior studies have largely examined artificial intelligence as an ad hoc, informal auxiliary tool utilized independently by individual teachers outside of structured, formal professional development programs (Allen et al., 2025; Hanis & Wahyudin, 2024). Rather than conceptualizing AI as an integral operational component embedded within a comprehensive training management system, current research predominantly treats technology adoption as an isolated, self-directed activity. Consequently, rigorous empirical evidence remains notably limited regarding whether embedding AI capabilities within a systematically structured training cycle produces measurable, statistically significant improvements in teachers' professional competence to construct high-quality, reasoning-based numeracy literacy test items. From a theoretical perspective, established teacher education and training frameworks also lack an explicit, cohesive account of how generative AI technologies can be integrated seamlessly across all operational stages, namely planning, organization, implementation, and systematic evaluation (Arora & Mishra, 2025). Without such integrative management frameworks, educators and school administrators lack clear methodological blueprints to guide the purposeful implementation, monitoring, and quality assurance of AI-enhanced professional development initiatives, thereby leaving a substantial void in both empirical research and applied training methodology.

To directly address these empirical and theoretical gaps, this study develops and evaluates NULIQFORM, an innovative AI-assisted training management model designed to embed artificial intelligence specifically within the implementation stage of a structured, four-stage professional development cycle encompassing planning, organization, implementation, and evaluation. Rather than treating AI as a standalone tool or an unguided auxiliary resource, the NULIQFORM model systematically unifies preliminary needs analysis, evidence-based instructional design, and AI-supported guided practice into a single, coherent management system tailored for primary education contexts. Within this structured framework, AI serves as an interactive cognitive prompt engine during workshop sessions, guiding teachers step-by-step to formulate contextually relevant scenarios, refine mathematical item structures, and align task complexity with target reasoning standards. By embedding continuous scaffolding within a robust management cycle, NULIQFORM bridges the gap between theoretical training principles and practical classroom assessment creation. Accordingly, this study aims to (1) examine the operational feasibility, internal validity, and user acceptability of the NULIQFORM model across primary school settings, and (2) empirically evaluate its overall effectiveness in enhancing primary school teachers' pedagogical competence in designing high-quality,

context-rich, and reasoning-based numeracy literacy assessments that foster higher-order cognitive skills in learners.

METHODS

This study employed a Research and Development (R&D) approach adapting the ADDIE framework (Analysis, Design, Development, Implementation, Evaluation) to design, validate, and evaluate the effectiveness of an Artificial Intelligence (AI)-assisted training management model named NULIQFORM. To measure the model's effectiveness in enhancing primary school teachers' ability to construct numeracy literacy test items, a quasi-experimental one-group pretest–posttest design (O_1 – X – O_2) was applied, analyzed via paired-sample t-tests and Cohen's d . Participants comprised 31 primary school teachers from Waru Sub-district selected through purposive sampling based on four inclusion criteria: (1) actively teaching at the primary level, (2) involved in developing assessment instruments, (3) willing to participate in the training program, and (4) possessing basic digital literacy skills. This sample size was deemed appropriate for exploratory and early-stage developmental research focusing on model feasibility and iterative refinement (Creswell & Poth, 2016). Content validity and practical feasibility were evaluated qualitatively by two doctoral-level experts with over five years of professional experience: Dr. H. Sultoni, M.Pd. (an expert in educational management) validated the training model and implementation tools, whereas Dr. Abdul Haris Rosyidi, M.Pd. (an expert in mathematics education) validated the substantive content, presentation systematics, and academic feasibility. The validation procedure involved direct delivery of initial drafts, oral interviews, open-ended questionnaires, and iterative revisions until final consensus was reached.

The intervention was conducted through a structured training program encompassing three main phases: conceptual exploration, AI-assisted guided practice, and peer presentations with reflective feedback. The training integrated a web-based Generative AI system (ChatGPT GPT-4) as an adaptive cognitive scaffold to guide participants in constructing Higher-Order Thinking Skills (HOTS)-oriented instruments (Bahroun et al., 2023; Owan et al., 2023). This item construction process followed a six-stage operational workflow, as detailed in Table 2. As an illustration regarding Grade 5 2D shape area topics, participants were guided to construct prompts framed around the socio-cultural context "*Taman Sekolah Ramah Anak*" (Child-Friendly School Garden), where students help design a play area combining rectangular and triangular sections. Participants iteratively refined initial AI-generated drafts into contextual story problems requiring students to analyze combined areas and explain their calculation strategies in alignment with the Pusmenjar–PISA numeracy framework and HOTS rubrics.

Table 1. Operational Workflow of AI Integration in Constructing HOTS Numeracy Items

Operational Stage	Main Teacher Activity	Output Produced
Objective Identification	Defining learning outcomes, numeracy indicators, subject matter, and target grade level.	Item specifications criteria
Prompt Formulation	Constructing structured prompts including instructions, role, scenario context, and cognitive constraints.	Standardized prompt text
Draft Generation	Submitting prompts to ChatGPT (GPT-4) to generate initial instrument drafts.	Item draft, stimulus, & answer options
Output Evaluation	Reviewing draft alignment against HOTS indicators and the Pusmenjar–PISA framework.	Cognitive & context revision notes

Iterative Revision	Refining prompt instructions or manually editing drafts if criteria are not yet met.	Revised item draft
Final Validation	Assessing final drafts using performance scoring rubrics prior to adoption as training products.	Final HOTS numeracy item

Research data were collected using a needs analysis questionnaire, intervention observation sheets, and a performance-based item writing assessment. Instrument quality was ensured through content validity via expert judgment, construct validity aligned with theoretical frameworks of numeracy literacy, and internal consistency reliability yielding a Cronbach's Alpha coefficient of 0.82. Teachers' item-writing performance was consistently evaluated on both pre-test and post-test tasks using a 4-point analytical rubric covering three main dimensions: (1) understanding the essence of numeracy items (context relevance and context-concept relationships), (2) ability to construct HOTS-oriented numeracy items (appropriateness of cognitive level, reasoning demands, and context), and (3) ability to construct numeracy items from data or visual representations (data interpretation, representation, and decision-making). The maximum total score of 12 points was converted into four final competence levels: Low (1–3), Fair (4–6), Good (7–9), and Very Good (10–12).

FINDINGS AND DISCUSSION

This study developed the NULIQFORM (Numeracy Literacy Question Formation) training model consisting of four main phases: planning, organization, implementation, and evaluation, which are interconnected within a continuous cycle (Figure 1).

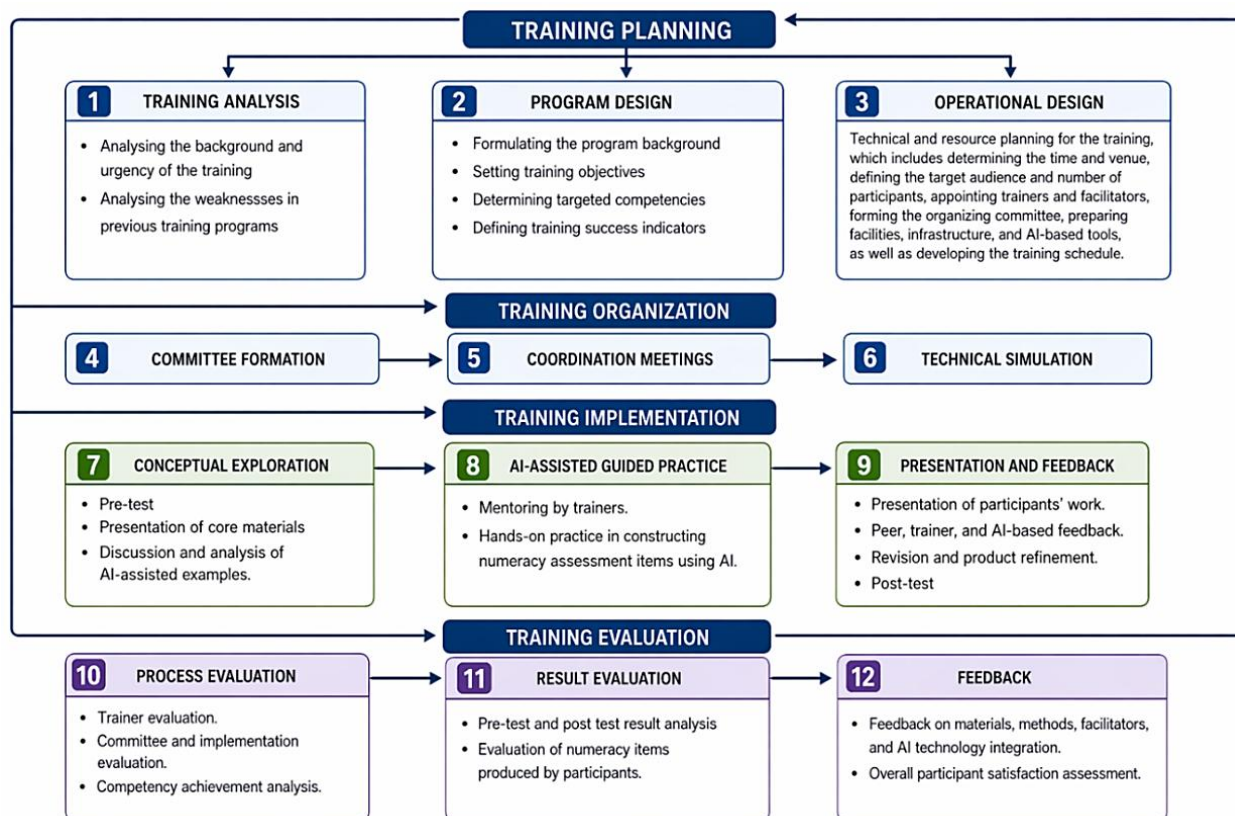


Figure 1. NULIQFORM Teacher Training Model

Training Planning: Empirical Needs Analysis and Emerging Demand for AI-Based Training

The initial training needs analysis reveals an overwhelming demand for a structured artificial intelligence (AI)-assisted training management model among primary school teachers. Specifically, 42.2% of surveyed respondents reported that such a model is highly needed, while 55.6% indicated it is needed, establishing a cumulative demand of 97.8%. Regarding pedagogical preferences, 77.8% of participants prioritized workshop-based or hands-on training approaches, supplemented by group discussions (13.3%) and case studies (8.9%), whereas traditional lecture-based methods received zero selection (0%). In terms of delivery format, face-to-face instruction emerged as the primary preferred mode (57.8%), followed by hybrid delivery formats (35.6%) and fully online approaches (6.7%). Furthermore, participating teachers identified comprehensive training modules as their main facility requirement (51.1%), followed by professional certification (26.7%) and direct consultation with facilitators (20.2%). Qualitative survey responses further confirmed that educators experience substantial difficulties in constructing numeracy literacy items and have not previously utilized artificial intelligence tools within classroom assessment development.

Beyond the quantitative survey results, preliminary qualitative investigations involving semi-structured interviews and document reviews with primary school teachers in Gugus 3, Waru Subdistrict, surfaced critical insights further explaining this high demand. The qualitative analysis revealed two major systemic challenges that substantiate the quantitative findings. First, teachers emphasized that their difficulty in constructing numeracy literacy test items extends beyond simple technical formatting; many educators struggle to design questions that are authentic, contextualized, aligned with learners' characteristics, and capable of measuring higher-order thinking skills. Second, teachers noted that existing professional development activities were rarely governed by an integrated management system. Prior training programs typically operated in fragmentation rather than systematically connecting needs analysis, planning, organization, guided practice, and evaluation into a single, coherent cycle. Consequently, teachers lacked the continuous managerial framework necessary to translate workshop training into lasting classroom assessment practices.

Third, qualitative findings indicated that no dedicated training tool had previously guided teachers through a structured process of AI-assisted item development. Educators lacked practical protocols for designing realistic problem contexts, formulating precise prompts, evaluating AI outputs, and iteratively revising or validating generated assessment items. Fourth, teachers underscored that usable item banks of numeracy literacy questions derived from authentic classroom practice remained extremely scarce, limiting available reference resources for test construction. Together, these qualitative themes substantiate and extend the quantitative survey results by providing deeper pedagogical context regarding teacher needs. Most importantly, these empirical findings directly informed the architectural framework of the NULIQFORM model. By unifying needs analysis, instructional design, AI-supported guided practice, and evaluation across a four-stage professional development cycle, the model directly addresses identified managerial, pedagogical, and technological gaps, providing primary school teachers with both an interactive item-writing scaffold and a sustainable item-generation workflow.

Training Organization Establishing an Operational Framework for AI-Assisted Delivery

The organization stage of the NULIQFORM training management model commenced with formal committee formation, structured planning, and role allocation to establish a robust operational framework for professional development. Empirical findings demonstrate that the training was

supported by a clearly defined, multi-tiered organizational structure designed to streamline logistical and instructional execution. Within this management architecture, explicit operational roles were assigned to both lead facilitators and dedicated technical support teams. Facilitators were tasked with managing pedagogical delivery, guiding participant interactions, and scaffolding the item-writing process, while technical personnel were assigned to oversee digital infrastructure, troubleshoot platform accessibility, and maintain seamless network connectivity. This deliberate division of responsibilities ensured that every instructional and administrative dimension of the training program was systematically managed. Consequently, this structured organizational foundation provided the operational clarity necessary to support primary school teachers effectively, minimizing administrative ambiguity and creating a highly coordinated environment conducive to intensive professional learning and collaborative AI-assisted item design.

Following committee formation, pre-training coordination meetings were systematically conducted to align all key stakeholders' understanding regarding core training objectives, implementation procedures, and the pedagogical application of artificial intelligence (AI) tools. These preliminary briefing sessions brought together primary school administrators, workshop facilitators, and technical support staff to establish unified operational expectations prior to program rollout. During these meetings, participants thoroughly reviewed the standardized training syllabus, clarified workshop schedules, and established explicit guidelines for integrating generative AI tools into numeracy item construction. Particular emphasis was placed on defining the exact boundaries of AI utilization, ensuring that facilitators were prepared to guide teachers in using AI as an interactive cognitive scaffold rather than a substitute for professional judgment. By establishing shared instructional goals, standardized protocols, and transparent communication channels among all team members, these coordination meetings ensured that all parties strictly adhered to consistent guidelines throughout the subsequent implementation phase, thereby safeguarding the methodological integrity of the training model.

Finally, comprehensive technical simulations were carried out prior to live execution to rigorously assess the operational readiness of the training infrastructure and AI-based systems. These simulation trials specifically evaluated the stability of web interfaces, server response times, and prompt processing speeds when accessing large language model platforms like ChatGPT within a workshop environment. By simulating actual training conditions, technical teams successfully identified potential operational bottlenecks, network latency issues, and software interface challenges before the actual sessions took place, allowing for proactive adjustments to be made to both digital tools and facilitation scripts. As a direct result of these thorough preparatory simulations, no major technical disruptions or system failures were recorded throughout the implementation phase. Training activities were delivered with high consistency, strict adherence to planned procedures, and uninterrupted access to AI tools, enabling participating teachers to engage deeply in guided practice and iterative item construction without operational friction.

Training Implementation: AI-Assisted Guided Practice and Competence Gains

The implementation stage represents the operational core of the NULIQFORM model, where conceptual understanding is systematically transformed into practical competency through three structured phases: conceptual exploration, AI-assisted guided practice, and presentation and feedback. During the initial conceptual exploration phase, participants completed a pre-test to establish a baseline measurement of their entry-level ability in constructing numeracy literacy questions. Prior to further

inferential statistical testing, the collected pre-test and post-test dataset was examined using a normality test to evaluate parametric assumptions, with the empirical probability distribution results presented in Figure 2.

The resulting probability plots for both pre-test and post-test scores demonstrate that the observed data points closely follow the diagonal reference line, indicating a normal distribution pattern without substantial deviations or extreme outliers. This alignment confirms a symmetrical and consistent distribution of competency scores across participants. Statistically, both datasets yield p-values greater than 0.100, formally verifying that the assumption of normality is fully met. Furthermore, the spread of data points in the post-test is slightly more concentrated around central values compared to the pre-test, indicating enhanced uniformity and consistency in teachers' question-construction performance following the training intervention.

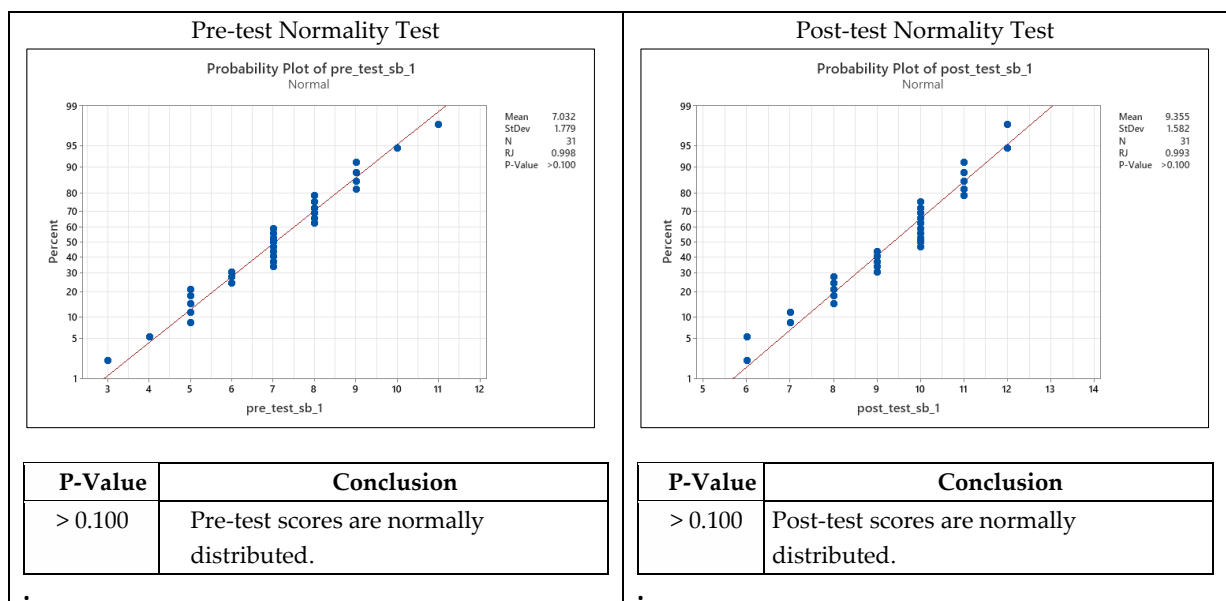


Figure 2: Normality Test

To examine the relationship between participants' initial and final performance, a correlation test using Spearman's rho was conducted. The correlation test results are presented in Figure 3. The analysis using Spearman's rho shows a correlation coefficient of 0.211 with a significance value of 0.254 ($p > 0.05$). This result indicates that there is no statistically significant relationship between pre-test and post-test scores. The distribution pattern shows that higher or lower initial scores are not consistently associated with corresponding final scores.

Correlations				
			pre_test_sb	post_test_sb
Spearman's rho	pre_test_sb	Correlation Coefficient	1.000	.211
		Sig. (2-tailed)	.	.254
		N	31	31
	post_test_sb	Correlation Coefficient	.211	1.000
		Sig. (2-tailed)	.254	.
		N	31	31

Figure 3: Correlation Test

The core learning process occurs within the AI-assisted guided practice phase, where participants actively construct numeracy literacy items through scaffolding provided by facilitators and generative AI tools. Operationalizing the multi-stage integration workflow outlined in Table 2, participants engage in iterative prompt formulation, draft generation, and output evaluation. After generating each initial item, participants present their work to receive formative feedback from peers and facilitators, which is subsequently applied to refine and validate the question prior to final adoption. To concretely illustrate how artificial intelligence functions as an adaptive cognitive scaffold during this guided-practice phase, Box 1 demonstrates a representative participant's prompt sequence alongside the item transformation produced before and after critical output evaluation and iterative refinement (as contextualized in Table 2).

Table 2. Example of AI-assisted item refinement during guided practice

Teacher / school	Wahyuni Tri Astuti - SD Al Muslim
Grade / topic	Grade 2 Measurement (non-standard units); Context: Personal
Prompt sequence	1. "You are a professional elementary numeracy item developer... Suggest alternative contexts (personal, socio-cultural, or scientific)..." 2. "Give me 5 more items on the above." 3. "Make items appropriate for Grade 2 age." 4. "Use non-standard units." 5. "For the item using footsteps, illustrate with the father's and the younger sibling's footsteps." 6. "Review it again so it better fits Grade 2 children." 7. "Give me just 1 item." 8. "Revise again the father and sibling walk from home to the tahu-petis seller, not to the gate."
Item before revision (AI draft)	Father needs 8 steps and the sibling needs 16 steps to reach the gate; then Father needs 12 more steps to reach the tahu-petis stall. Question: total steps Father took. Flaw identified by the teacher: the two people's step counts were being compared over different, unstated distances, the comparison had no valid mathematical basis.
Item after revision (teacher-corrected, final)	Father and the younger sibling walk directly from their front door in Buduran to their favourite tahu-petis stall, counting their own steps for the same distance: Father needs 20 steps, the sibling needs 40 steps. (1) What is the difference in steps between Father and sibling? (2) Why might the number of steps differ?
Why this example matters	The teacher, not the AI, caught and corrected a conceptual flaw in the AI-generated item (an invalid comparison of unequal distances), illustrating human evaluation and revision within the AI-assisted workflow rather than passive adoption of AI output.

Note. Prompts are translated from the original Indonesian and lightly condensed for presentation.

The effectiveness of the training is confirmed through a paired samples t-test, which reveals a statistically significant difference between pre-test and post-test scores ($t = -6.147$, $p < 0.001$). The mean increase of 2.32 points indicates a statistically and practically significant improvement in participants' competency. The results of the t-test are presented in Figure 4.

Paired Samples Test								
		Paired Differences						
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df
					Lower	Upper		Sig. (2-tailed)
Pair 1	pre_test_sb - post_test_sb	-2.32258	2.10376	.37785	-3.09425	-1.55092	-6.147	30
								<.001

Figure 4. Paired Samples Test

The effect size analysis using Cohen's *d* showed a large effect of the intervention on teachers' competency improvement ($d = 1.10$), indicating a large magnitude of change between pre-test and post-test competence scores.

Figure 5 visualizes this change at the sample level, showing the mean pre-test and post-test scores together with their standard deviations.

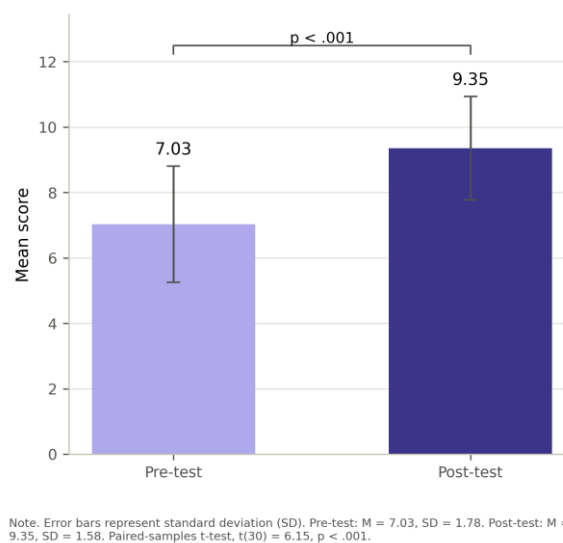


Figure 5. Mean Pre-test and Post-test Scores

Training Evaluation: Feasibility and Effectiveness of the NULIQFORM Model

The evaluation stage consists of process evaluation, outcome evaluation, and participant feedback, as outlined in the NULIQFORM model. This stage aims to assess both the quality of implementation and the effectiveness of the training in improving teachers' competencies.

The evaluation stage consists of process evaluation, outcome evaluation, and participant feedback. The results of participant feedback are presented in Table 3.

Table 3 Training Participant Feedback

Variables	Mean Score	Category	Interpretation
Training Implementation	4.7	Very High	Participants perceived the training implementation as highly effective, with clear facilitation and well-structured learning activities.
Teacher Competence Development	5.0	Very High	Participants reported substantial improvement in their competence, particularly in designing numeracy-based and HOTS-oriented assessment tasks.

Impact and Sustainability	4.8	Very High	Participants perceived strong positive impacts, including increased confidence, motivation, and commitment to continuous professional development.
Participant Satisfaction	5.0	Very High	Participants expressed very high satisfaction and strongly endorsed the broader adoption of the training model.

All evaluated variables fall within the “very high” category. The mean score for training implementation is 4.7, teacher competence development is 5.0, impact and sustainability is 4.8, and participant satisfaction is 5.0. The scores across all variables show a consistently high pattern, with minimal variation between indicators. The highest scores are observed in teacher competence development and participant satisfaction, both reaching the maximum mean value.

In addition to the closed self-report survey summarised in Table 3, model feasibility was further triangulated through a structured observation instrument completed by the head of the Teacher Working Group (*Ketua KKG*), who directly observed training delivery across all four NULIQFORM stages. The instrument comprised 22 observable indicators (6 for planning, 4 for organization, 8 for implementation, and 4 for evaluation), each rated on whether the activity was carried out (yes/no) and on a four-point implementation-quality scale (1 = strongly disagree to 4 = strongly agree). All 22 indicators were recorded as carried out, and no planned activity was reported as unimplemented. Twenty-one of the 22 indicators received the highest quality rating (4 = agree); the remaining indicator (“determining a relevant resource person/facilitator”) received a rating of 3. In the accompanying open-ended comment, the KKG head noted that the training material and resource person were excellent and that the material was ultimately delivered well (translated from the original Indonesian). These observation results converge with the high self-reported scores for Training Implementation ($M = 4.7$) and Participant Satisfaction ($M = 5.0$) in Table 3, providing an independent, external source of evidence for the feasibility of the NULIQFORM model.

DISCUSSION

The findings of this study indicate that the effectiveness of the NULIQFORM model cannot be reduced to improvements in training delivery alone, but rather represents a direct consequence of the structural alignment between needs analysis, instructional design, and artificial intelligence integration. In this context, alignment is no longer understood merely as a technical principle of instructional design, but as an active generative mechanism that shapes how learning processes are constructed, mediated, and ultimately translated into measurable competence transformation (Octavio & Gonzales, 2025). This perspective aligns closely with recent empirical literature emphasizing that educational effectiveness within digital learning environments is fundamentally determined by systemic coherence and intentional alignment across all instructional components (Hailikari, 2022; Salinas-navarro et al., 2024). When training elements operate in isolation, skill acquisition remains fragmented; however, by establishing structural congruence between diagnostic assessments and technological interventions, NULIQFORM creates a unified learning ecosystem. Consequently, this holistic alignment ensures that technological capabilities directly support pedagogical objectives, enabling primary school teachers to meaningfully bridge theoretical concepts with classroom assessment practices.

Theoretical and Practical Implications of the NULIQFORM Stages

At the planning stage, needs analysis functions as an epistemic foundation that not only identifies initial performance gaps but also actively shapes the underlying logic of the entire intervention design. Recent studies highlight that effective professional development programs must be firmly grounded in authentic classroom needs and contextual teacher challenges (Ajani & Africa, 2023; Sanusi et al., 2022). However, the findings of this study further demonstrate that a rigorous needs analysis determines how technology can be meaningfully integrated into pedagogical practice rather than applied superficially. The identified gap, namely teachers' limited ability to construct higher-order numeracy questions, reflects a severe disruption between national curriculum expectations and daily classroom assessment practices. Therefore, the integration of artificial intelligence in this study is not merely additive but truly transformative, as it directly targets both cognitive constraints and pedagogical limitations experienced by educators. This perspective aligns with recent research positioning artificial intelligence as a pedagogically embedded tool rather than a mere technological supplement in teacher education (Bahroun et al., 2023).

This conceptual grounding positions NULIQFORM's contribution as distinctly different from prior studies on artificial intelligence-assisted assessment training. Whereas earlier scholarly work primarily examined artificial intelligence as an ad hoc, informal auxiliary tool utilized independently by individual teachers outside structured professional development programs (Hanis & Wahyudin, 2024;), the present findings indicate that embedding artificial intelligence within a structured four-stage training cycle linking needs analysis, organization, guided practice, and evaluation is associated with a statistically significant and practically large improvement in teacher competence ($d = 1.10$). This substantial outcome suggests that the training management structure itself, rather than simple access to artificial intelligence technology alone, serves as a critical contributor to the observed professional growth. By embedding technological tools within a coherent management cycle, NULIQFORM provides a meaningful benchmark and point of comparison for future empirical studies evaluating structured versus unstructured approaches to artificial intelligence integration in teacher professional development programs.

The empirical insights gathered from the organization stage reveal that the effectiveness of any instructional design cannot be evaluated in isolation from the real-world conditions of its implementation. Recent literature suggests that failures in professional teacher development programs are frequently caused by inconsistencies and logistical breakdowns during implementation rather than inherent weaknesses in original model design (Ockerman & Bagui, 2024). The findings of this study strongly reinforce this argument by demonstrating that a well-defined organizational structure functions as an essential regulatory mechanism that actively maintains operational consistency between theoretical design and actual training practice. Thus, organization within the NULIQFORM framework is not merely a passive administrative component or logistical checklist, but rather a vital structural condition that ensures both the long-term sustainability and pedagogical integrity of instructional design. By establishing clear roles, standardized coordination, and technical simulations, the organizational framework shields the intervention from operational friction, ensuring that planned pedagogical activities translate effectively into meaningful learning experiences.

The primary theoretical contribution of this study emerges during the implementation stage, where observed improvements in teacher competence are explained through the dynamic interaction between experiential learning and artificial intelligence-mediated cognitive support. While recent studies reaffirm the fundamental importance of experiential learning in professional development

(Salinas-navarro et al., 2024), the large effect size observed ($d = 1.10$), alongside participants' self-reported gains in Teacher Competence Development ($M = 5.0$), suggests that experience alone may be insufficient to produce such substantial competence transformation. Because this study did not directly trace cognitive processes, this pattern is best understood as a conceptual interpretation regarding how AI scaffolding operates. Interpreted this way, artificial intelligence introduces an adaptive scaffolding function through repeated feedback and revision cycles, consistent with emerging perspectives on human-AI collaboration in education where technology acts as a cognitive partner extending human thinking capacity (Tugarin & Husen, 2025). Future research should collect process-level data, such as interaction logs, to empirically substantiate this mechanism.

Furthermore, the empirical findings of this study contribute significantly to the ongoing reconceptualization of the Technological Pedagogical Content Knowledge (TPACK) framework. Recent educational literature suggests that TPACK is inherently dynamic, multifaceted, and context-dependent when applied within digital learning environments (Mario et al., 2023; Tschönhens et al., 2024). This study extends that theoretical perspective by demonstrating that the successful integration of technology, pedagogy, and content knowledge is the direct result of intentional, structured instructional design. Consequently, TPACK should no longer be viewed solely as an isolated individual attribute or static trait possessed by a teacher, but rather as an evolving competence that is mediated and nurtured by the deliberate design of structured learning experiences. By embedding artificial intelligence tools into specific steps of item development, NULIQFORM provides the instructional scaffolding necessary for teachers to operationalize TPACK in practice. This structural mediation enables educators to synthesize technological affordances with mathematical content knowledge, thereby transforming theoretical knowledge frameworks into practical test construction capabilities.

Another important theoretical and practical implication of this study relates directly to educational equity among primary school teachers. The absence of a statistically significant relationship between participants' initial entry-level ability and their final post-intervention outcomes indicates that the NULIQFORM model operates effectively across diverse baseline competence levels. This finding aligns with recent empirical studies emphasizing the crucial role of structured instructional design in reducing skill disparities and promoting equitable professional growth among educators (Blömeke et al., 2022; Fathi et al., 2025; Sunandar et al., 2022). Therefore, the NULIQFORM training model can be conceptualized as a powerful compensatory mechanism that supports a more equitable distribution of professional learning outcomes through adaptive, artificial intelligence-based cognitive scaffolding. By providing individualized feedback and iterative prompts, the model enables teachers with lower initial assessment skills to achieve proficiency comparable to their peers, thereby narrowing competence gaps across school districts and fostering broader instructional quality.

From a conceptual perspective, this study proposes three main theoretical contributions to the field of teacher professional development. First, needs analysis functions as a generative mechanism in instructional design rather than serving merely as a static diagnostic tool for identifying skill deficits (Blegur et al., 2024). Second, artificial intelligence acts as an active co-cognitive agent that directly mediates and elevates human thinking processes during task execution, rather than simply functioning as a passive technical aid or automated repository (Salas-Pilco, 2023). Third, teacher competence within the TPACK framework is shown to be the direct outcome of structured, scaffolded instructional design, and should therefore be re-conceptualized as a design-mediated competence (Dilling et al., 2024). Together, these three conceptual pillars reframe how technological tools and managerial structures interact in teacher training. By shifting focus from isolated skill acquisition to design-mediated

cognitive collaboration, these contributions provide a clear conceptual foundation for future research exploring technology-enhanced professional development systems.

Synthesis: Structural Alignment, Experiential Learning, and AI Scaffolding as Interconnected Drivers of NULIQFORM Effectiveness

Overall, the empirical and theoretical findings suggest that the effectiveness of the NULIQFORM model is not determined by any single isolated factor, but rather by the synergistic interaction between structured instructional design, practice-oriented experiential learning, and artificial intelligence-supported cognitive scaffolding. These three interconnected elements collectively create a coherent learning system that yields statistically significant, highly consistent, and equitable improvements in primary school teachers' assessment competence. Structured management ensures procedural fidelity, experiential practice fosters meaningful application, and artificial intelligence provides continuous cognitive reinforcement. Without the structured management cycle, technology adoption risks becoming fragmented; without artificial intelligence scaffolding, experiential practice may fail to reach higher cognitive demands. By orchestrating these components into a unified framework, NULIQFORM demonstrates how modern professional development can transcend traditional workshop limitations. Ultimately, this integrated system bridges the gap between theoretical assessment standards and daily classroom practices, ensuring that primary teachers acquire sustainable skills in constructing higher-order numeracy items.

This study advances the field of teacher professional development by demonstrating that competence formation in the era of artificial intelligence is not merely an isolated outcome of individual learning processes, but an emergent property of systematically designed instructional ecosystems. By integrating needs analysis, experiential learning, and artificial intelligence-mediated cognitive scaffolding into a unified cycle, the NULIQFORM model redefines the operational role of artificial intelligence from a supportive technological tool to an active co-cognitive agent. This paradigm shift re-conceptualizes teacher competence from an intrinsic individual attribute into a design-mediated construct, while simultaneously providing robust empirical evidence that such an integrated model enhances learning outcomes in an equitable manner across diverse educator profiles. Therefore, this study makes a significant contribution to the growing global discourse on human-AI collaboration in education by offering a theoretically grounded and empirically validated framework for designing effective, adaptive, and inclusive professional development systems for primary education.

Limitations

Despite the positive empirical outcomes demonstrated by the NULIQFORM model, several methodological limitations must be acknowledged to properly contextualize the findings within teacher education literature. This study applied a quasi-experimental one-group pretest-posttest design involving a localized sample of primary school teachers as part of a phased model development and initial empirical validation. While this design is appropriate for early-stage intervention research aimed at identifying preliminary effectiveness prior to large-scale testing, the absence of a dedicated control group limits the extent to which observed competence gains can be exclusively attributed to the intervention. In particular, because participants completed the same performance-assessment task at both pre-test and post-test, a portion of the observed improvement may reflect a testing or practice effect rather than the training program alone. Furthermore, the evaluation focused primarily on short-term gains in teachers' item-construction competence without assessing long-term skill retention or

direct transferability to daily classroom assessment practices. Finally, individual heterogeneity in teachers' entry-level digital literacy was not formally modeled as a moderating variable, despite its potential influence on how effectively educators engage with AI-assisted training environments.

Directions for Future Research

To address these limitations and build upon the foundational insights of this study, several critical directions for future research are recommended. First, future empirical investigations should advance toward large-scale quasi-experimental or randomized controlled trial designs incorporating control groups to establish stronger causal inferences regarding model efficacy. Second, longitudinal studies are necessary to examine the long-term durability of acquired competence and evaluate how effectively teachers translate AI-assisted item construction skills into sustained classroom assessment practices over time. Third, future research should incorporate key moderating variables, such as individual digital literacy and institutional readiness, to explain differential training effects across diverse teacher demographics. Furthermore, researchers should conduct comparative evaluations between NULIQFORM and alternative professional development models to isolate the specific value added by structured AI scaffolding. Ultimately, capturing process-level interaction logs will offer deeper empirical insights into how educators interact with generative AI tools, informing adaptive teacher education system design.

CONCLUSION

This study examined the feasibility and effectiveness of the NULIQFORM model in enhancing primary school teachers' competence to construct reasoning-based numeracy literacy questions. The empirical results demonstrate substantial and consistent post-intervention improvements with a large effect size, alongside high feasibility ratings from both training participants and KKG coordinators. Crucially, the absence of a meaningful relationship between participants' initial and final competence levels indicates that these gains occurred independently of prior baseline ability, establishing the model as an equitable professional development tool. Conceptually, these findings suggest that competence development was driven less by simple training exposure than by the systemic alignment of needs analysis, instructional design, and AI-supported guided practice. Although the proposed role of generative artificial intelligence as an adaptive cognitive scaffold extending the TPACK framework into a design-mediated construct represents a conceptual interpretation rather than a directly traced empirical mechanism, it provides a coherent theoretical account for the observed skill transformation.

Practically, NULIQFORM offers a replicable four-stage management framework covering needs analysis, organization, AI-supported guided practice, and evaluation that training providers can adapt to other reasoning-based competencies, while demonstrating to school leaders and KKG coordinators that structured, feedback-oriented AI integration is far more effective than ad hoc individual usage. For policymakers, the model highlights a viable pathway for narrowing competence gaps across diverse teacher populations to foster educational equity. However, given the exploratory one-group pretest–posttest design and localized sample size, these conclusions remain preliminary. Future research should implement quasi-experimental or randomized controlled designs with comparison groups to rigorously test causal assertions, conduct longitudinal tracking to evaluate long-term competence retention and classroom transfer, and collect process-level data such as interaction logs or think-aloud protocols to empirically substantiate the cognitive scaffolding mechanisms during AI-assisted guided practice.

Acknowledgments: The author would like to express sincere gratitude to the Indonesia Endowment Fund for Education (LPDP) for the financial support provided for this study. The author also appreciates all parties who contributed administrative and technical assistance throughout the research process.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

- Ajani, O. A., & Africa, S. (2023). *Journal of Integrated Elementary Education*. 3(2), 120–136.
- Allen, K. A., Costner-Lark, A., Serratt, T., & Gordon, J. A. (2025). Artificial intelligence in advanced practice nursing education: Opportunities and challenges. *Journal of the American Association of Nurse Practitioners*. Scopus. <https://doi.org/10.1097/JXX.0000000000001169>
- Am, S. A., Am, S. A., Am, S. A., & Saiful, S. (2026). The impact of the professional education program for in-service teachers on pedagogical competence in Indonesia. *Jurnal Cakrawala Pendidikan*, 45(1). <https://doi.org/10.21831/cp.v45i1.65044>
- Arif, M., Aziz, M. K. N. A., & Arif, M. A. M. (2025). A Recent Study on Islamic Religious Education Teachers' Competencies in the Digital Age: A Systematic Literature Review. *Journal of Education and Learning (EduLearn)*, 19(2), 587–596.
- Arora, D. S., & Mishra, P. S. (2025). Innovation And Technology In Strategic Management. In *IIP Series* (pp. 74–84). Iterative International Publishers (IIP). <https://doi.org/10.58532/nbennurbpsm7>
- Asiyah, S., Wiyono, B. B., Hidayah, N., & Supriyanto, A. (2021). The Effect of Professional Development, Innovative Work and Work Commitment on Quality of Teacher Learning in Elementary Schools of Indonesia. *Eurasian Journal of Educational Research*. <https://eric.ed.gov/?id=EJ1321959>
- Bahrour, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability (Switzerland)*, 15(17). <https://doi.org/10.3390/su151712983>
- Bahrudin, N., Zamroni, A., & Kausar, S. (2026). Strengthening Teachers' Professional Competence through Strategic Leadership of Madrasah Principals to Improve the Quality of Madrasah Aliyah Education. *Journal of Education and Learning Innovation*, 3(1), 15–30. <https://doi.org/10.59373/jelin.v3i1.215>
- Blegur, J., Ma, A., Mahendra, A., & Layao, M. M. (2024). *Need analysis for innovation in integrated learning models for micro- teaching course: Explanatory sequential design*. 12(1), 50–66.
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79(February). <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Chairawati, F., Ganiyu, O. A., Hajar, I., Walidin, W., Mahmud, S., & Fakri, F. (2026). Opportunities and Challenges of Using Artificial Intelligence (AI) in Learning Tahfidz Qur'an. *Dirasah International Journal of Islamic Studies*, 4(1), 131–144. <https://doi.org/10.59373/drs.v4i1.93>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.
- Dilling, F., Schneider, R., Weigand, H.-G., & Witzke, I. (2024). Describing the digital competencies of mathematics teachers: Theoretical and empirical considerations on the importance of experience and reflection. *ZDM - Mathematics Education*, 56(4), 639–650. <https://doi.org/10.1007/s11858-024-01560-4>
- Fathi, E., Fathi, T. El, Saad, A., Larhzil, H., Lamri, D., Mehdi, E., & Ibrahmi, A. (2025). *Integrating generative AI into STEM education: Enhancing conceptual understanding , addressing misconceptions , and assessing student acceptance*.

- Febrianti, E. F., Arifudin, M., & al-Islam, M. (2025). Fostering Religious Character Through Safari Istighotsah: Insights from Islamic Education Practice in Indonesia. *Interdisciplinary Journal of Social Sciences*, 2(1), 23–34. <https://doi.org/10.59373/ijoss.v2i1.135>
- Fitriana, F., Azzukhrufi, J. R., Trihantoyo, S., & Kristanto, A. (2026). Digital School Safety Governance and School Climate: A Single Case Study in an Indonesian Public Elementary School. *Journal of Elementary Education Research and Practice*, 2(3), 158–172. <https://doi.org/10.70376/ahvy7m83>
- Hailikari, T. (2022). *Student perspectives on how different elements of constructive alignment support active learning*. 9(Siltavuorenpenger 10). <https://doi.org/10.1177/1469787421989160>
- Hakim, F., Syam, N., Ch, M., & Ibrahim, A. M. (2026). Intelligence in Islamic Education: Knowledge Integration or Ethical Alienation. *Attaqwa: Jurnal Ilmu Pendidikan Islam*, 22(1), 111–127. <https://doi.org/10.54069/attaqwa.v22i1.1149>
- Hanis, M., & Wahyudin, D. (2024). Pemanfaatan Artificial Intelligence (AI) Dalam Penyusunan Asesmen Pembelajaran Bagi Guru Sekolah Dasar. *Jurnal Ilmiah Profesi Pendidikan*, 9(2), 1199–1207. <https://doi.org/10.29303/jipp.v9i2.2252>
- Hariadi, Valianto, B., Akhmad, I., & Bin Syed Ali, S. K. (2025). Implementation study of authentic assessment in physical education in Indonesia and Malaysia. *Cakrawala Pendidikan*, 44(1), 102–115. <https://doi.org/10.21831/cp.v44i1.71485>
- Hidayanto, H. N., Sumardjoko, B., & Sutopo, A. (2026). Innovating critical reading through HOTS-based digital books: A mixed-methods study of the FIVES strategy among EFL university students. *Jurnal Cakrawala Pendidikan*, 45(1). <https://doi.org/10.21831/cp.v45i1.91424>
- Hidayati, N., Aripin, S., & Hina, S. (2025). Gamified Formative Assessment with Quizizz to Enhance Learning Outcomes at Madrasa. *Journal of Islamic Education Research*, 6(3), 303–314. <https://doi.org/10.35719/jier.v6i3.492>
- Humaida, H., Wati, I. K., Mukmin, M., & Kausar, S. (2026). Digitalizing Islamic Education: Developing a Methodological Model for Generation Z Learners. *Attaqwa: Jurnal Ilmu Pendidikan Islam*, 22(1), 75–95. <https://doi.org/10.54069/attaqwa.v22i1.1044>
- Jelatu, S., Mandur, K., Makur, A. P., Nendi, F., & Gunur, B. (2019). Konstruksi Tes High Order Thinking Skills (HOTS) bagi Guru-Guru Matematika SMP di Manggarai Timur. *E-Dimas: Jurnal Pengabdian Kepada Masyarakat*, 10(2), 214. <https://doi.org/10.26877/e-dimas.v10i2.3070>
- Jogezai, N. A., Koroleva, D., & Ivanov, I. (2025). Generative artificial intelligence in higher education: Challenges, opportunities and future course of actions to achieve sustainable development goals. *International Journal of Educational Management*, 1–25. <https://doi.org/10.1108/IJEM-12-2024-0786>
- Kariyanto, H., Hanipah, S., Dimas, M., Hakim, R. X., & Saputra, R. (2025). Digital Transformation of Islamic Education Learning: A Study of iPad-Based Innovation in Enhancing Student Learning Engagement. *Dirasah International Journal of Islamic Studies*, 3(2), 172–185. <https://doi.org/10.59373/drs.v3i2.85>
- Lodge, J. M., Ellerton, P., Zaphir, L., & Brown, D. (2024). Assessing in the Age of AI: Is Critical Thinking the Answer? In *Artificial Intelligence Applications in K-12*. Routledge.
- Mario, J., Iglesias, O., Cabero-almenara, J., & Palacios-rodríguez, A. (2023). *Development of the teacher ' s technological pedagogical content knowledge (TPACK) from the Lesson Study: A systematic review*. (February). <https://doi.org/10.3389/educ.2023.1078913>
- Novrianti, N., & Sari, L. C. (2025). Competence-Adaptive Assessment Using IRT for Junior Photographer Qualification. *Online Learning In Educational Research (OLER)*, 5(2), 305–317. <https://doi.org/10.58524/oler.v5i2.858>
- Nurwahyuni, N., & Za, T. (2025). Enhancing Teacher Professionalism through the Development of a PositiveSelf-Concept Based on Communication Psychology. *Jurnal Nispatti*, 10(3), 265–284. <https://doi.org/10.26811/nispatti.v10i3.154>

- Ockerman, L. S., & Bagui, S. (2024). *Critical Reflection Sessions: Teacher ' s Perspectives During Professional Development*. 00(October), 1–10. <https://doi.org/10.47852/bonviewIJCE42023792>
- Octavio, M., & Gonzales, F. (2025). *Artificial Intelligence as a Pedagogical Resource in Initial Teacher Training La inteligencia artificial como recurso de enseñanza en la formación inicial docente*. <https://doi.org/10.62486/latia2025367>
- OECD. (2023). Pisa 2022 results volume v: Learning strategies and attitudes for life. In *Factsheets: I*.
- 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*. 19(8).
- Perla, L., Agrati, L. S., & Beri, A. (2025). Post-teaching and professional learning: An investigation on teachers attitudes towards AI. *Professional Development in Education*, 51(3), 466–477. <https://doi.org/10.1080/19415257.2025.2465970>
- Pham, S. T. H., & Sampson, P. M. (2022). The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning*, 38(5), 1408–1421. <https://doi.org/10.1111/jcal.12687>
- Reimers, F., Schleicher, A., Saavedra, J., & Tuominen, S. (2020). Supporting the continuation of teaching and learning during the COVID-19 Pandemic. In *Oecd*. globaled.gse.harvard.edu.
- Safi'i, I., & Amar, F. (2019). Pelatihan Penyusunan Instrumen Evaluasi Berstandar HOTS bagi Guru-Guru SD di Wilayah Banyudono. *Abdimas Dewantara*, 2(2), 149. <https://doi.org/10.30738/ad.v2i2.4176>
- Salas-Pilco, S. Z. (2023). Correction to: Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review (Education Sciences, (2022), 12, 8, (569), 10.3390/educsci12080569). In *Education Sciences* (Vol. 13, Issue 9). <https://doi.org/10.3390/educsci13090897>
- Salinas-navarro, D. E., Vilalta-perdomo, E., & Michel-villarreal, R. (2024). *Education sciences Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment*.
- Sanusi, A., Khoeriah, N. D., Arifin, D., Studi, P., Pendidikan, I., Pascasarjana, S., & Nusantara, U. I. (2022). *Training Needs Assessment On Blended Learning Model Management To Improve Teachers ' .* 01(9), 1222–1230.
- Shobihah, I. F. (2025). School Well-Being Among Children and Adolescents: Assessing Strategies for Bullying Prevention in Madrasahs. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, 8(3), 566–574. <https://doi.org/10.54069/attadrib.v8i3.1023>
- Sukenti, D., Agustina, & Fauzan, A. (2024). Self-esteem, Islamic Behavior, and Writing Assessment for Indonesian Language Madrasa Teachers: An Empirical Phenomenological Investigation. *Nazhruna: Jurnal Pendidikan Islam*, 7(3), 570–589. <https://doi.org/10.31538/nzh.v7i3.31>
- Sumarsono, R. B., Kusumaningrum, D. E., Gunawan, I., Alfarina, M., Romady, M., Ariyanti, N. S., & Budiarti, E. M. (2019). *Training on the Implementation of Cooperative Learning Models as an Effort to Improve Teacher's Performance*. 259–263. <https://doi.org/10.2991/coema-19.2019.53>
- Sunandar, A., Mustiningsih, Sunarni, Efendi, M., Ediyanto, Nafi'a, M. Z. I., Praherdhiono, H., Setiawati, L., & Casmal, M. (2022). Development of Computer Teaching Materials on Education Management Applications Through e-Learning Module. *2022 2nd International Conference on Information Technology and Education (ICIT&E)*, 127–131. <https://doi.org/10.1109/ICITE54466.2022.9759872>
- Susilawati, S., Yaqin, M. Z. N., Wahidmurni, W., Ulfa, F. A., & Krismoneta, K. (2023). Training on Preparation of Islamic Integrated HOTS-Based Questions and Their Application in Online Learning in Elementary Schools. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 4(4), 867–877. <https://doi.org/10.31538/munaddhomah.v4i4.459>

- Syukur, F., Maghfurin, A., Marhamah, U., & Jehwae, P. (2024). Integration of Artificial Intelligence in Islamic Higher Education: Comparative Responses between Indonesia and Thailand. *Nazhruna: Jurnal Pendidikan Islam*, 7(3), 531–553. <https://doi.org/10.31538/nzh.v7i3.13>
- Tiok Wijanarko, Fitria, Y., Ladiva, H. B., Rifa'i, M., Melati, R., Adrimal, A., Zuryanty, Z., & Susanto, N. (2025). Innovative Training on HOTS Learning and Digital-Based Literacy and Numeracy Question Development. *GANDRUNG: Jurnal Pengabdian Kepada Masyarakat*, 6(1), 1877–1888. <https://doi.org/10.36526/gandrung.v6i1.4421>
- Tschönhens, F., Backfisch, I., Fütterer, T., & Lachner, A. (2024). TPACK in action: Contextual effects of pre-service and in-service teachers' knowledge structures for technology integration. *Computers and Education Open*, 7, 100219. <https://doi.org/10.1016/j.caeo.2024.100219>
- Tugarin, N., & Husen, C. Van. (2025). International Journal of Information Development and integration of human-AI interactions in service applications: Conceptual framework and review. *International Journal of Information Management Data Insights*, 5(2), 100357. <https://doi.org/10.1016/j.jjime.2025.100357>
- Umar, M., S, W. Q., Ma'arif, M. A., Ismail, F., & Rahman, R. (2025). Revitalization of Islamic Character Values in Local Folklore and its Implication on Character Education. *Ulumuna*, 29(1), 188–220. <https://doi.org/10.20414/ujis.v29i1.1375>
- Wale, B. D. (2024). Artificial intelligence in education: Effects of using integrative automated writing evaluation programs on honing academic writing instruction. *Cakrawala Pendidikan*, 43(1), 273–287. <https://doi.org/10.21831/cp.v43i1.67715>
- Yang, X., & Kaiser, G. (2022). The impact of mathematics teachers' professional competence on instructional quality and students' mathematics learning outcomes. *Current Opinion in Behavioral Sciences*, 48, 101225. <https://doi.org/10.1016/j.cobeha.2022.101225>