

Exploring ChatGPT as a Scaffolding Tool for Posing Mathematical Problems: Insights from Primary Pre-service Teachers

Baharullah

Universitas Muhammadiyah Makassar

Sriyanti Mustafa

Universitas Muhammadiyah Pare-Pare

Sitti Fithriani Saleh

Universitas Muhammadiyah Makassar

Sri Rahayuningsih

Universitas Negeri Malang

Muhammad Gazali

Universitas Negeri Malang

Received: March 2026 || Accepted: June 2026

© 2026 Mathematics Education Research Group of Australasia, Inc.

The study reported in this article investigated ChatGPT's role as a digital scaffolding tool for pre-service primary teachers engaged in mathematical problem posing focused on word problems, an area underexplored in research on generative AI in pre-service teacher education. Grounded in the Zone of Proximal Teacher Development (ZPTD) framework, the research examined how AI-mediated interactions support cognitive, metacognitive, and pedagogical dimensions of instructional design. Twenty-two pre-service mathematics teachers participated in a three-phase qualitative exploratory case study involving independent problem posing (Task A), AI assisted problem posing with ChatGPT (Task B), and subsequent written reflections and semi-structured interviews. Thematic analysis of task outputs, interaction logs, and reflective data revealed that ChatGPT provided multidimensional scaffolding (cognitive, metacognitive, strategic, affective, and pedagogical), which collectively enhanced participants' task quality and depth of reflection. AI-assisted tasks showed notable improvements in contextual relevance, problem structure, cognitive demand, and creativity compared to mathematical problems posed independently. Participants reported substantial shifts in instructional awareness, moving from procedural task design toward student-centred and imaginative problem construction. These findings suggest ChatGPT has potential as a reflective learning partner, supporting development of pedagogical reasoning and instructional creativity in teacher education. The study extends the ZPTD framework to AI-mediated contexts, offering theoretical and practical insights for integrating generative AI into professional learning for pre-service teachers.

Keywords • mathematical problem posing • pre-service teacher education • ChatGPT • instructional scaffolding • artificial intelligence in education • reflective teaching development

Introduction

The integration of generative artificial intelligence (AI) into education has been met with both optimism and caution. Enthusiasts highlight the capacity of large language models such as ChatGPT (www.chatgpt.com) to create responsive, dialogic, and context-aware learning environments (Adiguzel et al., 2023; Rudolph et al., 2024), while critics question whether these tools genuinely foster deep learning or merely simulate pedagogical interaction (Holmes et al., 2022; Selwyn, 2019). Central to this debate is the need to move beyond treating Artificial Intelligence (AI) as a content delivery mechanism and instead examine its potential as a pedagogical partner, a tool that can scaffold not only students' content knowledge but also teachers' instructional reasoning (Msambwa et al., 2025; Zawacki-Richter et al., 2019).



In the study, dialogic learning environments were conceptualised as spaces where meaning is co-constructed through iterative, reflective exchanges (Wells, 1999). Within such environments, cognitive support refers to AI-generated clarifications, reformulations, and examples that reduce task complexity while preserving mathematical integrity (Wood et al., 1976). Pedagogical support encompasses guidance that directs attention to learner characteristics, curricular goals, and instructional strategies. This form of support can be understood more broadly within the literature on scaffolding, which distinguishes conceptual, strategic, metacognitive, and motivational functions of support (Belland et al., 2013). Expert reasoning modelling refers to making aspects of expert decision-making visible to learners (Collins & Michalski, 1989). In AI-mediated learning environments, this type of support may be facilitated through AI-generated suggestions and interactions that support prospective teachers' mathematical reasoning and problem-posing processes (Dilling & Herrmann, 2024; Burgos & Tizón-Escamilla, 2025). Together, cognitive support, pedagogical support, and expert reasoning modelling provide the operational vocabulary for analysing AI-mediated scaffolding in teacher learning.

Problem posing is widely regarded as an integral component of mathematical proficiency and an essential practice for primary teachers (Baharullah, 2009; Cai & Hwang, 2020; Silver, 1994;). It involves not only generating questions but also situating mathematical ideas in meaningful contexts, anticipating student solution strategies, and designing tasks that invite multiple entry points (Singer et al., 2015). For pre-service teachers, the ability to pose rich, cognitively demanding problems is a key indicator of readiness to move from learner of mathematics to designer of mathematical experiences (Cumhur, 2022; Nurita et al., 2024).

Pre-service teachers' spontaneous problem posing tends to be procedural and decontextualised, a pattern that reflects what Warford (2011) terms the self-assistance stage of teacher development, where instructional decisions are guided primarily by personal experience as learners rather than by pedagogical principles. Such tasks limit students to recall and routine application, denying opportunities for the reasoning and problem-solving central to mathematical proficiency (Findell et al., 2001; Stein & Smith, 1998). This reflects Warford's (2011) self-assistance stage: decisions guided by learner experience, not pedagogy. Transformation requires deliberate analysis, feedback, and reflection (Silver et al., 1996); without scaffolding, procedural habits persist.

Recent studies have started to explore AI's role in mathematics teacher education (Buchholtz & Huget, 2024; Cameron et al., 2025; Drushlyak et al., 2025; Pepin et al., 2025), yet none have focused specifically on scaffolding the multidimensional processes of mathematical problem posing. Within mathematics education, studies that do involve teacher-AI interaction often focus on using AI to generate instructional materials for teachers rather than engaging teachers in reflective co-design with AI (Li & Patton, 2025). Far less attention has been paid to how AI might support teachers' professional learning, particularly their pedagogical design capabilities (Cotton et al., 2024; Niraula, 2024). Despite the proliferation of AI applications in education, systematic reviews indicate that research has clustered around three dominant areas: (a) content delivery and personalised tutoring for students, (b) automated assessment and feedback, and (c) language learning support (Athanassopoulos et al., 2023; Zawacki-Richter et al., 2019). Consequently, there is a lack of empirical knowledge about how generative AI can scaffold the cognitive, metacognitive, and affective processes involved in authentic professional tasks of teaching, such as designing mathematical problems for students.

Primary education tasks require child-anchored contexts, accessible language, and dual purposes: they need to support students' engagement with meaningful contexts while also promoting mathematical learning (Singer et al., 2015). Designing suitable mathematical tasks requires pre-service teachers to coordinate mathematical content with pedagogical considerations and knowledge of students, a process that can present significant challenges for novices (Lestari et al., 2026; Suwanti et al., 2024). These demands require pre-service teachers to integrate content, student, and design knowledge when developing mathematical problems for primary school students. Understanding the potential for AI to support pre-service teachers to design mathematical problems warrants attention. To examine this potential, Warford's (2011) Zone of Proximal Teacher Development (ZPTD) provides a useful theoretical lens. Extending Vygotsky's (1978) concept of Zone of Proximal Development to



teacher education, the ZPTD distinguishes between self-assistance and teacher/instructor assistance. Within this framework, development occurs through cycles of mediation in which pre-service teachers appropriate new concepts and ways of reasoning. AI can provide personalised support and scaffolding through interaction, while its mediation differs from human assistance (Msambwa et al., 2025; Yelland, 2026).

The present study addresses the intersecting gaps identified in the literature by investigating the following research questions:

- RQ1: *What scaffolding affordances emerge during pre-service primary teachers' interactions with ChatGPT while posing mathematical problems?*
- RQ2: *How does ChatGPT influence the cognitive and pedagogical quality of the mathematical problems pre-service primary teachers pose?*
- RQ3: *In what ways does engagement with ChatGPT contribute to pre-service primary teachers' reflective thinking and instructional awareness?*

In sum, this study offers three distinct contributions. First, it empirically documents the multidimensional scaffolding affordances of a generative AI tool within an ecologically valid teacher education task. Second, it provides a theory-driven analysis of how AI mediation may support progression through the ZPTD, thereby extending a foundational teacher development framework into the digital age. Third, it generates practical insights for teacher educators seeking to integrate AI not as a shortcut, but as a catalyst for reflective, intentional, and student-centred instructional design. The key constructs—scaffolding affordances, cognitive and pedagogical quality of tasks, and reflective thinking—are defined and operationalised in the Conceptual Framework section.

Literature Review

Artificial Intelligence in Education: From Content Delivery to Pedagogical Partnership

The rapid advancement of AI has generated growing scholarly interest in its potential to transform teaching and teacher education. Early applications focused on computer-assisted instruction and intelligent tutoring systems (Baker, 2021). However, generative models like ChatGPT have shifted the discourse toward more interactive and co-constructive learning possibilities (Adiguzel et al., 2023; Rudolph et al., 2024). Despite this shift, critical scholars caution against technological determinism and overstating AI's potential without robust evidence (Cameron & Mesiti, 2024; Holmes et al., 2022; Selwyn, 2019). These critiques underscore the need to examine the affordances of AI in specific educational practices.

Within mathematics education, researchers have begun exploring ChatGPT for problem solving and explanation of mathematical concepts and solution strategies (Athanasopoulos et al., 2023; Li & Patton, 2025; Ming & Mansor, 2023), but the predominant focus remains on student-facing applications. Far less attention addresses how AI might support teachers' pedagogical expertise (Cotton et al., 2024). In this study, AI is positioned not as a replacement for human judgment but as a cognitive and reflective partner, aligning with Msambwa et al.'s (2025) and Yelland's (2026) call to reimagine pedagogies with generative AI.

Problem Posing in Mathematics Teacher Education

Problem posing is a fundamental mathematical activity and essential for primary teachers (Silver, 1994; Cai & Hwang, 2020; Singer et al., 2015). The problems selected reveal teachers' mathematical conceptions and assumptions about worthwhile tasks (Silver et al., 1996). For pre-service teachers, posing rich problems signals readiness to design meaningful mathematical experiences (Cumhur, 2022; Nurita et al., 2024) and preparedness to teach. Yet research consistently shows that novices generate



procedural, routine problems lacking real-world connection (Aktaş, 2022; Kilic, 2015). A typical example: “Ani has 5 apples and buys 3 more. How many apples does Ani have?” Such tasks represent low cognitive demand, falling into memorisation or procedures without connections (Stein & Smith, 1998). Research on moving beyond procedural routines shows that structured variation in prompts stimulates original problem designs (Baumanns & Rott, 2024); opportunities to analyse, revise, and reflect shift novices toward richer tasks (Silver et al., 1996); and systematic problem posing instruction remains rare despite its importance for pedagogical content knowledge (Cai & Hwang, 2020). Without targeted scaffolding (human or AI-mediated), procedural tendencies are unlikely to change spontaneously. By contrast, rich problem-posing tasks invite multiple solution paths and genuine mathematical reasoning. For instance, Singer et al. (2015) described a task in which students design a rectangular garden with a fixed perimeter, exploring various dimensions and justifying which design is “best” for a given purpose—a task that integrates measurement, multiplication, and argumentation. Such tasks, however, are rarely produced by novices without substantial scaffolding. This difficulty may reflect limited pedagogical content knowledge (Shulman, 1987) as well as the influence of the apprenticeship of observation, through which novices’ prior experiences as students shape their developing conceptions of teaching (Lortie, 2020). Warford’s (2011) ZPTD frames this as the self-assistance stage, where novices rely on personal experience and require external mediation.

Scaffolding in Teacher Learning

Scaffolding enables learners to accomplish tasks beyond their independent capability; it is contingent, gradually withdrawn, and transfers responsibility (Van de Pol et al., 2010; Wood et al., 1976). In mathematics education, scaffolding supports conceptual understanding, strategic thinking, and affective engagement (Belland et al., 2013; Kusmaryono et al., 2020). In teacher education, traditional scaffolding (mentor guidance, peer collaboration, reflection) faces limitations in asynchronous contexts (Viberg et al., 2023), prompting interest in technology-mediated scaffolding.

Drawing on Belland et al.’s (2013) framework, scaffolding is conceptualised as potentially operating across multiple dimensions. Cognitive scaffolding supports understanding of content and procedures. Metacognitive scaffolding prompts reflection on one’s own thinking and learning processes. Strategic scaffolding offers heuristics and approaches for tackling complex tasks. Affective scaffolding sustains motivation and reduces anxiety. Pedagogical scaffolding, particularly relevant for teacher learning, directs attention to instructional considerations such as learner characteristics, learning goals, and task design. These categories provide an analytical lens for examining the kinds of support ChatGPT may offer during problem-posing tasks.

Theoretical Framework: The Zone of Proximal Teacher Development

Vygotsky’s (1978) Zone of Proximal Development (ZPD) refers to the distance between a learner’s actual developmental level, determined by independent problem solving, and their potential developmental level, determined through problem solving under the guidance of a more capable other. Development within this zone is mediated through social interaction and cultural tools. Warford (2011) extended this concept specifically to teacher education, proposing the Zone of Proximal Teacher Development (ZPTD). The ZPTD describes the developmental space between what teacher candidates can do based on their spontaneous, experiential knowledge (self-assistance) and what they can potentially achieve with the mediation of more knowledgeable others (teacher assistance). Warford outlined four phases in this developmental trajectory:

1. *Self-assistance*: Novice teachers draw on their own experiences as learners, often resulting in teaching practices that reproduce what they observed.
2. *Teacher assistance*: With guidance from mentors, supervisors, or other knowledgeable sources, novices begin to examine their assumptions and consider alternative approaches.



3. *Internalisation*: Through reflection and practice, novices' appropriate professional concepts and integrate them with their experiential knowledge.
4. *Recurrence*: Novices revisit and reconstruct earlier understandings, now able to articulate pedagogical reasoning and make intentional instructional decisions.

To illustrate in the context of this study:

1. *Self-assistance*: A pre-service teacher draws on their own experience as a learner and produces a simple, single-step word problem (e.g., A farmer has 20 ducks. He sells 5. How many are left?), replicating the type of tasks they repeatedly encountered as students.
2. *Teacher assistance*: With scaffolding from a mentor or, in this study, AI mediation, the novice is prompted to add a second step, introduce a real-life constraint, or consider what students might find confusing. The task is revised collaboratively.
3. *Internalisation*: Over time, the novice begins to independently ask questions such as "Could this problem be more open-ended?" or "What context would engage my students?" appropriating the pedagogical reasoning previously modelled by the mentor or AI.
4. *Recurrence*: In a later practicum or independent design task, the novice spontaneously poses a multi-step, contextually rich problem and can articulate why it is more effective than a routine task, thereby demonstrating reconstructed understanding.

ZPTD was chosen over alternative frameworks of teacher learning, such as Clarke and Hollingsworth's (2002) Interconnected Model, because it foregrounds the role of external mediation in bridging independent and assisted performance. This makes it uniquely suited to examining how an AI tool such as ChatGPT might function as a means of mediation in teacher development, without attributing social agency or intentionality to the technology (Wertsch, 1998).

Conceptual Framework

This study is guided by three interconnected constructs: scaffolding affordances, task quality, and reflective practice.

Scaffolding affordances

Drawing on Belland et al. (2013), we distinguish five forms of support that may emerge during AI-mediated problem posing: cognitive (clarifying mathematical content), metacognitive (prompting reflection on design choices), strategic (offering heuristics for extending problems), affective (sustaining motivation and reducing anxiety), and pedagogical (directing attention to instructional considerations). These are operationalised in the coding scheme presented in Data Analysis: Table 2.

Cognitive and pedagogical quality of tasks

We assess problem-posing quality along four dimensions adapted from Silver (1994) and Singer et al. (2015): contextual relevance, problem structure, cognitive demand, and creativity. Each dimension is defined in the rubric (see Data Analysis: Table 1) and reflects the extent to which problems move from procedural, decontextualised exercises toward richer, student-centred tasks.

Reflective practice and instructional awareness.

Reflective practice refers to participants' capacity to articulate design rationales linked to student thinking and learning goals (Anthony et al., 2015), while instructional awareness encompasses shifts in beliefs about mathematics teaching (Beswick, 2012) and the role of AI in instructional design. Together, these capture the internalisation and recurrence processes posited by the ZPTD.

Synthesis and Research Gap

A search of the literature revealed several important gaps. Research on artificial intelligence in education has primarily focused on student learning rather than teacher professional development.



Although problem posing is a key teacher competency, novice teachers often remain limited to designing procedural tasks (Lestari et al., 2026). Scaffolding offers a valuable framework for understanding teacher learning, yet it faces practical limitations, and the ZPTD framework has not been systematically applied in digital learning contexts.

The intersection of AI supported scaffolding and teacher problem posing in primary mathematics is nascent. This study addresses this gap by examining how ChatGPT can function as a multidimensional scaffolding tool that supports preservice teachers in generating mathematical problems and strengthening pedagogical reasoning. By grounding the analysis in the ZPTD framework, the study contributes both empirical evidence and theoretical insight into digitally mediated teacher development.

Methods

Research Design

This study employed a qualitative exploratory case study design (Yin, 2018). This approach is appropriate for investigating a contemporary phenomenon—AI-mediated scaffolding in teacher problem posing—within its real-world context, where the boundaries between phenomenon and context are not clear. The exploratory nature of the design allows for in-depth examination of the processes, perceptions, and outcomes of pre-service teachers' engagement with ChatGPT, which remains an under-researched area. The study was structured in three phases aligned with Warford's (2011) ZPTD framework: self-assistance (independent problem posing), AI-mediated (AI-assisted problem posing), and internalisation/recurrence (reflection and interview).

Participants

Twenty-two pre-service primary mathematics teachers (15 females, 7 males; aged 19–22) enrolled in the fifth semester of a teacher education program at a public university in Indonesia participated in the study. Fifth-semester students were purposefully selected because they had completed foundational coursework in mathematics education and had completed previously a primary classroom practicum, positioning them at a transitional stage from learner to prospective practitioner. This developmental phase is particularly relevant for examining how external scaffolding (in this case, AI) might support pedagogical choices and reasoning.

Purposive sampling (Patton, 2015) was employed with the following inclusion criteria: (a) enrolment in the primary mathematics teacher education program, (b) completion of at least one school-based practicum, and (c) basic digital literacy sufficient to interact with ChatGPT. Exclusion criteria included prior extensive use of AI for instructional design, which might have confounded the scaffolding effects. This was determined through a screening question during recruitment asking whether participants had previously used ChatGPT or similar AI tools to design lesson materials or mathematical tasks. Those who reported more than two such experiences were excluded. Variation in academic achievement and teaching experience was intentionally sought to capture a range of responses, though all participants were novice in terms of formal problem-posing instruction. The sample size of 22 is consistent with recommendations for exploratory qualitative case study research, where information power (Malterud et al., 2016) rather than statistical power guides sample adequacy.

Participants' prior teaching experience ranged from one to two semesters of practicum, involving 4–6 weeks of classroom observation and limited teaching responsibilities in local primary schools. This experience familiarised them with classroom realities but did not include systematic training in problem posing, making the cohort ideal for examining the impact of AI scaffolding.



Procedure

The study was conducted over three structured phases across approximately three weeks. The duration was chosen to allow sufficient time for orientation, independent task completion, AI interaction, and reflective engagement without causing participant fatigue or attrition. Each phase was deliberately aligned with a stage of the ZPTD framework.

Phase 1: Orientation and baseline task (self-assistance stage)

Participants attended a 90-minute orientation session that introduced the research objectives, explained ethical considerations (including voluntary participation, confidentiality, and the right to withdraw), and provided technical training on using ChatGPT for educational purposes. During this session, participants were given examples of pedagogically relevant prompts (e.g., “Create a mathematics word problem for Grade 4 students based on a daily-life situation”) and strategies for critically evaluating AI-generated content. A summary of the evaluation strategies taught is provided in the *Appendix*. The orientation also addressed ethical boundaries, emphasising academic integrity, avoiding over-reliance on AI, and maintaining critical judgment. After the orientation, each participant independently designed one primary level mathematics problem (Task A) without any AI assistance. Task A was designed to capture baseline problem-posing capability. Participants were instructed to create a problem suitable for Grade 2–5 students, specifying the topic, grade level, and learning objective. The word-problem format was intentionally chosen to capture baseline procedural tendencies. No further guidance was provided, allowing their spontaneous problem-posing approaches to emerge. The minimal instruction specified a word problem format to capture unassisted baseline capability. Consistent with ZPTD’s self-assistance stage (Warford, 2011), it was anticipated most would produce procedural, single-step tasks (Aktaş, 2022; Kilic, 2017). The aim was a reference point for comparison with AI-assisted Task B, not rich open-ended problems.

Phase 2: AI-supported problem posing (teacher assistance stage)

Approximately one week later, participants engaged individually with ChatGPT (GPT-4 version) for 45–60 minutes. To ensure consistency in AI engagement while allowing authentic interaction, all participants began with a standardised prompt: “I am a pre-service elementary teacher. Please help me design a mathematics word problem for Grade [X] students focusing on [topic].” Participants were free to choose the grade level (2–5) and mathematical topic based on their interest or practicum experience. They were instructed to interact iteratively with ChatGPT for at least three conversational turns, saving the complete dialogue log for analysis. ChatGPT was used in its standard consumer configuration without additional prompting to ask reflective questions; instances where the AI responded with questions or evaluative prompts were emergent behaviours of the model rather than pre-programmed features. Using ChatGPT’s standard configuration, all interactions occurred within one week to minimise update-related variability. Participants treated ChatGPT as a design partner, asking questions and requesting revisions. No prescribed questioning framework was provided. Participants drew on orientation evaluation strategies and their own pedagogical knowledge to decide what to revise. As novices, their ability to diagnose problem weaknesses varied, likely influencing the depth and focus of their ChatGPT interactions.

Phase 3: Reflection and interview (internalisation and recurrence stages)

Immediately following the AI interaction, participants wrote a guided reflection of 400–500 words addressing: (a) how ChatGPT influenced their problem design, (b) what they learned about mathematical problem posing, and (c) how they perceived the pedagogical value of AI. The reflection prompt was developed based on a framework of reflective practice (Schon, 1983), and pilot-tested with two non-participant pre-service teachers to ensure clarity and depth. Within one-week, semi-structured interviews were conducted online via Zoom or Google Meet, each lasting 20–30 minutes. Interview questions were designed to probe reflective thinking and pedagogical awareness, aligned with Research Question 3. Examples include: “How did ChatGPT influence the way you think about



students?” and “What changed in your beliefs about creating mathematics problems?” Interviews were audio-recorded, transcribed verbatim, and returned to participants for member checking to enhance credibility.

Data Sources

Four complementary datasets were collected:

1. Task Outputs: Problems posed independently (Task A) and with AI assistance (Task B). These provided direct evidence of changes in problem quality, structure, and complexity.
2. Interaction Logs: Full transcripts of participants’ dialogues with ChatGPT, documenting iterative reasoning, scaffolding patterns, and decision-making processes.
3. Written Reflections: Participants’ reflective accounts of their experiences, perceived challenges, learning gains, and evolving views on problem posing and AI.
4. Interview Transcripts: Narrative accounts offering deeper insight into pedagogical reasoning, reflective development, and shifts in instructional beliefs.

Data Analysis

Data analysis followed Braun and Clarke (2006) thematic analysis, adapted for the multiple data sources. Data saturation was monitored throughout the iterative coding process. By the 18th participant, no new themes emerged, and analysis of the remaining four participants confirmed thematic saturation. The analysis proceeded in four stages:

Stage 1: Analysis of task outputs

Task A and B problems were evaluated using a deductively developed rubric. Contextual relevance and creativity drew from Silver (1994) and Singer et al. (2015); problem structure and cognitive demand from Stein and Smith’s (1998) Task Analysis Guide. Four-level descriptors were refined through pilot coding (two researchers independently scoring sample problems). Minor wording adjustments ensured the rubric captured pilot responses, finalised as presented in Table 1. In qualitative data analysis, descriptive statistics (means and standard deviations) are used to summarise and illustrate patterns in these scores. They are presented for descriptive purposes only and are not intended for statistical generalisation or hypothesis testing (Subhaktiyasa et al., 2023).

Table 1

Rubric for Assessing Problem Posing Quality

Dimension	Score 1	Score 2	Score 3	Score 4
Contextual Relevance	Problem is abstract, disconnected from students’ experiences	Problem has few connections to real life situations	Problem uses familiar context but does not use the context to solve the problem.	Problem is richly embedded in students lived experiences with situational detail
Problem Structure	Single-step, single correct answer	Single step with multiple possible answers	Multi-step leading to single answer	Multi-step, open-ended, multiple entry points
Cognitive Demand	Memorisation or recall of facts/procedures	Application of routine procedures	Application of procedures, making connections to mathematical ideas and relationships	Doing mathematics: using strategies, reasoning, and representations
Creativity	Routine, textbook-like problem	Problem with minor variation from routine	Problem with engaging narrative or context	Original, novel problem inviting multiple strategies



Prior to scoring, both researchers participated in a two-hour moderation session using sample problems to calibrate their understanding of the rubric. Two researchers independently scored a random sample of 20% of the tasks (9 problems) to establish inter-rater reliability. Initial agreement was 82%; Cohen's kappa (κ) was calculated at 0.84, indicating substantial agreement (Landis & Koch, 1977). Discrepancies were resolved through discussion, and the remaining tasks were scored by one researcher with periodic checks on every fifth task to prevent coder drift.

Stage 2: Analysis of interaction logs

Interaction logs were coded using a scaffolding typology developed deductively from Belland et al. (2013) and refined inductively through initial coding. The final typology included five categories, as shown in Table 2.

Table 2
Scaffolding Typology for Analysing ChatGPT Interactions

Scaffolding Type	Definition	Indicators	Unit of Analysis
Cognitive	Supports understanding of mathematical content or problem structure	Clarifications, reformulations, error corrections, examples	Turn or episode
Metacognitive	Prompts reflection on design choices or student thinking	Questions about clarity, appropriateness, student interpretation	Turn
Strategic	Offers heuristics or approaches for extending or varying problems	Suggestions for multi-step problems, alternative contexts	Episode
Affective	Sustains motivation, reduces anxiety	Encouragement, affirmations, positive feedback	Turn
Pedagogical	Directs attention to instructional considerations	Prompts about grade level, student needs, learning goals	Episode

The unit of analysis was a conversational turn (a single participant or AI message) for cognitive, metacognitive, and affective scaffolding, and an episode (a sequence of turns focused on a particular design move) for strategic and pedagogical scaffolding. Two researchers independently coded 20% of the interaction logs (five participants' logs). Inter-rater reliability reached $\kappa = 0.84$, indicating substantial agreement. The remaining logs were coded by one researcher.

Stage 3: Analysis of reflections and interviews

Written reflections and interview transcripts were analysed thematically. Initial coding was open and inductive (Braun & Clarke, 2006), identifying meaningful units related to participants' experiences, learning, and beliefs. Codes were then grouped into themes through iterative discussion among the research team. Themes were reviewed and refined, and negative cases (participants whose reflections did not align with dominant themes) were explicitly sought and analysed to ensure balanced interpretation. Saturation was monitored: after analysing 18 participants, no new themes emerged, confirming adequacy.

Stage 4: Integration and interpretation

Findings from all data sources were synthesised to address the research questions. Patterns across tasks, interactions, and reflections were examined to identify how scaffolding affordances related to changes in task quality and reflective development. The ZPTD framework was used to interpret participants' developmental trajectories, from which claims about internalisation are made cautiously given the short duration of the study.



Trustworthiness and Rigour

As recommended by Lincoln and Guba (1985), several methodological strategies were employed to enhance the trustworthiness of the findings. Credibility was established through triangulation of multiple data sources (tasks, logs, reflections, interviews), member checking of interview transcripts, and peer debriefing among the research team. Transferability was supported by thick description of the context, participants, and procedures, enabling readers to assess applicability to other settings. Dependability and confirmability were addressed through an audit trail documenting analytical decisions, coding schemes, and reflexive notes.

Results

Observed Changes in Problem Posing Quality

The comparative analysis of problems generated in Task A (independent posing) and Task B (AI-assisted posing) revealed shifts in the cognitive and pedagogical quality of participants' outputs. These shifts were examined using a four-dimensional rubric adapted from Silver (1994) and Singer et al. (2015): contextual relevance, problem structure, cognitive demand, and creativity. In qualitative data analysis, descriptive statistics (means and standard deviations) are used to summarise and illustrate patterns in these scores. They are presented for descriptive purposes only and are not intended for statistical generalisation or hypothesis testing (Subhaktiyasa et al., 2023). As shown in Table 3, scores shifted across all four dimensions from Task A to Task B. Most problems moved from single step to multi-step (procedural to connected). Very few reached open-ended "doing mathematics" (Stein & Smith, 1998). One AI session enriched structure but not radically open-ended expert design.

Table 3

Comparative Dimensions of Problem Posing Quality

Dimension	Task A Mean (SD)	Task B Mean (SD)	Change
Contextual Relevance	2.15 (0.64)	3.41 (0.58)	+1.26
Problem Structure	2.09 (0.71)	3.22 (0.67)	+1.13
Cognitive Demand	1.95 (0.59)	3.05 (0.62)	+1.10
Creativity	2.00 (0.68)	3.18 (0.65)	+1.18

Note: $N = 22$ paired tasks

Problems posed in Task A were frequently disconnected from students' lived experiences, often framed as decontextualised arithmetic computations. For example, Participant 7 wrote in Task A: "Ali has 7 pencils, and he buys 5 more. How many pencils does he have now?" While mathematically valid, such tasks offer limited opportunity for students to engage meaningfully with everyday contexts. In Task B, with ChatGPT's mediation, problems became more situated. The same participant revised the problem to:

Ali wants to share 12 pencils equally among his three friends. If one pencil is broken, how does this change the sharing? What if he buys an additional pack of 6 pencils?

This version introduces distributive reasoning, accounts for real-life contingencies (a broken pencil), and connects to familiar classroom experiences. The mean score for contextual relevance shifted from 2.15 to 3.41, indicating a trend toward more authentic task design.

The structure of problems posed independently was typically closed, demanding single-step answers. Participant 12's Task A problem exemplifies this pattern: "A box contains 25 chocolates. If 5 are eaten, how many remain?" This question requires only subtraction and yields a single solution. After interacting with ChatGPT, the same participant generated a multi-step problem:



A box contains 25 chocolates. If 5 are eaten, how many remain? If the remaining chocolates are to be shared equally among 4 children, how many will each receive? How many are left undistributed?

This layering of steps introduces division with remainders, reflecting ChatGPT's scaffolding prompts (e.g., "Can you extend the problem to involve sharing or grouping?"). The mean score for problem structure shifted from 2.09 to 3.22.

Task A problems were predominantly procedural, focusing on straightforward arithmetic. The mean score for cognitive demand was below 2.0, indicating that most tasks fell into the categories of memorisation or procedures without connections (Stein & Smith, 1998). Participant 3's Task A problem illustrates this: "There are 4 rows of chairs with 6 chairs each. How many chairs are there in total?"

In Task B, the same participant developed a problem with different cognitive demands:

A hall has 4 rows of 6 chairs each. If 5 new chairs are added to each row, how many chairs are there now? If 15 students are absent, how many seats remain empty?

This problem requires multiple steps, integration of multiplication and subtraction, and understanding of real-life classroom constraints. The mean score for cognitive demand shifted to 3.05.

Creativity in Task A was limited; most problems followed conventional school-textbook phrasing. Participant 18's Task A problem stated: "A farmer has 20 ducks. He sells 5. How many are left?" In Task B, with ChatGPT's scaffolding, the same participant generated:

A farmer has 20 ducks. He sells 5 at the market and uses the money to buy 3 bags of rice. Each bag of rice costs the same amount. How much does each bag cost if he sold each duck for Rp10,000?

This revised problem intertwines arithmetic with proportional reasoning, monetary context, and narrative flow. The mean score for creativity shifted from 2.00 to 3.18.

While the overall pattern showed shifts across all four dimensions, variation was observed among participants. Not all participants demonstrated substantial gains. For example, Participant 9's Task B problem remained largely procedural: "A student has 15 candies. He gives 3 to his friend. How many are left?" Although the interaction log showed that ChatGPT suggested multi-step variations, the participant chose not to incorporate them, indicating that AI mediation alone does not guarantee transformation. Participant 14 showed improvement in contextual relevance but scored lower on cognitive demand in Task B compared to Task A, suggesting that shifts in one dimension may not always align with shifts in others.

These negative cases highlighted that the effectiveness of AI scaffolding may be mediated by individual factors such as prior knowledge, engagement level, or receptiveness to AI suggestions. They also underscore that the changes reported above represent group-level trends rather than universal outcomes. Variations in participants' baseline mathematical pedagogical content knowledge (MPCK; Shulman, 1987) may also have influenced their responsiveness to AI scaffolding, though this was not measured directly. Participants' baseline MPCK may explain variation: stronger MPCK enabled critical evaluation of ChatGPT's suggestions and targeted prompts; weaker MPCK led to passive acceptance, explaining marginal improvement in some Task B problems. As MPCK was not directly measured, this interaction could not be disentangled. Future research should incorporate MPCK assessments to examine differential benefits.

Task B changes suggested ChatGPT expanded participants' problem-posing repertoires. Through ZPTD, the Task A→B shift indicated movement from self-assistance (existing knowledge) to teacher assistance (AI-mediated design). However, these changes occurred in a single session; without longitudinal data, claims of internalisation or lasting transformation remain provisional as negative cases caution against overgeneralisation.

ChatGPT as a Multi-dimensional Scaffolding Tool

The analysis of participants' interaction logs revealed that ChatGPT functioned as a multifaceted scaffolding agent, providing support across cognitive, metacognitive, strategic, affective, and pedagogical dimensions. This finding underscores that the role of ChatGPT extended beyond simply



supplying problem examples; it actively mediated participants' thinking, guided their reflection, and expanded their pedagogical awareness. In the ZPTD framework, ChatGPT's mediation reflects the teacher assistance stage, in which external guidance helps novice teachers progress beyond their independent capabilities.

As shown in Table 4, the largest proportion of scaffolding instances was cognitive ($n = 153$), followed by metacognitive ($n = 115$), strategic ($n = 96$), affective ($n = 57$), and pedagogical ($n = 57$). The balance across these categories indicates that ChatGPT was not merely a cognitive tool but a multi-dimensional scaffolding partner. Each form of scaffolding played a distinct role in supporting teacher growth, as discussed below with illustrative examples from interaction logs.

Table 4
Distribution of Scaffolding Affordances in ChatGPT Interactions

Scaffolding Type	Frequency (n)	Definition
Cognitive	153	Supports understanding of mathematical content or problem structure
Metacognitive	115	Prompts reflection on design choices or student thinking
Strategic	96	Offers heuristics or approaches for extending or varying problems
Affective	57	Sustains motivation, reduces anxiety
Pedagogical	57	Directs attention to instructional considerations

Note: $N = 478$ coded segments from 22 participants

Cognitive scaffolding was evident when ChatGPT directly supported participants in improving the mathematical accuracy and coherence of their problems. For example, Participant 5 initially posed: "There are 30 candies. If divided by 5 students, how many each?" ChatGPT responded:

Good start! To make it clearer for Grade 3 students, you might write: A teacher has 30 candies. She wants to share them equally among 5 students. How many candies will each student get?

The reformulation preserved the mathematical essence while situating it in a Variation and Contradictory Cases scenario. This aligns with Wood et al.'s (1976) original conceptualisation of scaffolding as reducing complexity while maintaining the integrity of the learning goal.

Metacognitive scaffolding occurred when ChatGPT asked participants to evaluate or justify their design choices. For instance, after Participant 12 drafted a problem involving fractions, ChatGPT asked: "How do you think a Grade 4 student might interpret this? Would they need visual support to solve it?" This prompted the participant to reconsider the wording and eventually add a diagram to accompany the problem. By asking reflective questions, ChatGPT pushed participants toward self-monitoring and critical evaluation of their work, skills central to effective teaching practice. The process resonates with Schoenfeld (1992) emphasis on metacognitive regulation as essential for mathematical problem solving and instruction. Future applications could deliberately configure AI tools as reflective partners—for instance, by instructing the model to ask probing questions rather than provide direct suggestions—to further strengthen the metacognitive dimension of scaffolding.

Strategic scaffolding was particularly influential in helping participants broaden the scope of their problems. For example, Participant 8 initially posed: "A farmer has 20 chickens. He sells 5. How many left?" ChatGPT suggested:

You could extend this problem by adding another step: If he buys 10 more chickens, how many does he have now? Or, if he wants to sell the remaining chickens equally among 3 buyers, how many does each buyer get?

Participant 8 subsequently integrated these ideas into a multi-step, open-ended problem. This demonstrates how ChatGPT supported participants in moving from procedural to richer problem designs.

Affective scaffolding played an important motivational role. Several participants expressed initial hesitation. For instance, Participant 3 noted in the log: "I don't know if my problem is good enough."



ChatGPT responded with affirmations such as: “That’s a good start—let’s try to make it even more engaging for students.” Similarly, Participant 15 received encouragement when ChatGPT replied: “You’re on the right track! What if we add a twist to make it more interesting?” These encouraging responses helped reduce anxiety and sustain engagement. For novice teachers, affective support is critical in building confidence, as highlighted by Belland et al. (2013), who argued that scaffolding is not only cognitive but also socio-emotional in nature. Much of this affective scaffolding, however, was generic praise (e.g., “Good start!”). Such affirmations reduce anxiety but lack diagnostic specificity. Effective feedback, as Williams (2024) argued, focuses on improving the work, not emotive reactions. Thus, ChatGPT’s scaffolding, while useful, does not match the nuanced feedback of a skilled human mentor.

Pedagogical scaffolding was observed when ChatGPT oriented participants toward instructional considerations. For example, Participant 19 designed a problem involving decimals: “A student buys 2.5 litres of milk and drinks 0.5 litres. How much remains?” ChatGPT replied: “This is appropriate for upper primary. For Grade 4, you may want to simplify decimals or connect it to fractions, such as: A student buys 2 $\frac{1}{2}$ litres of milk and drinks $\frac{1}{2}$ litre. How much remains?” ChatGPT’s grade-level recommendations were generic (e.g., “appropriate for upper primary”), based on broad training data patterns—not the Indonesian curriculum or classroom variation. A task deemed suitable for Grade 4 may be too challenging for some students and too simplistic for others. Thus, AI-generated advice requires teacher evaluation and adaptation, as teachers hold the contextual knowledge for sound instructional decisions. By encouraging participants to tailor tasks to students’ developmental levels, ChatGPT supported pedagogical reasoning. This reflected the teacher assistance function of scaffolding within ZPTD, helping novices consider learners’ readiness and potential misconceptions.

From ZPTD, ChatGPT’s scaffolding aligned with teacher assistance, enabling progress beyond independent abilities. Cognitive/strategic scaffolding expanded problem spaces; metacognitive/pedagogical fostered instructional awareness; affective provided socio-emotional support. Together, these illustrate AI as symbolic mediator in teacher development. Findings align with Wood et al.’s (1976) dynamic scaffolding, extend AI research beyond cognitive correction (Holmes et al., 2022), and provided empirical evidence for Warford’s ZPTD—showing AI bridges pre-service teachers from self-assistance toward internalisation.

Reflective Growth and Pedagogical Insights

The analysis of written reflections and interview transcripts provided insight into how participants perceived their engagement with ChatGPT and what they learned about mathematical problem posing. Through thematic analysis following Braun and Clarke (2006), four main themes emerged, reflecting shifts in participants’ thinking about problem design, students, and the role of AI in teaching. Table 5 presents the distribution of these themes across participants, including frequency counts to illustrate the prevalence of each theme while acknowledging variation in individual experiences. The most frequently occurring theme was reframing mathematics tasks, followed by recognising student thinking, and confidence in task construction. These themes suggested that participants’ engagement with ChatGPT prompted many, though not all, to reconsider their assumptions about what constitutes a good mathematical problem and to adopt a more student-centred perspective. The following sections elaborate on each theme with illustrative examples from participants’ reflections and interviews.



Table 5
Themes Emerging from Reflective Data (N = 22 participants)

Theme	Participants (n/N)	Description
Reframing mathematics tasks	19/22	Recognition that problems can be creative, narrative-driven, and open-ended rather than routine
Recognising student thinking	17/22	Increased awareness of how students might interpret problems and the need to anticipate misconceptions
Confidence in task construction	16/22	Expressed greater self-efficacy in designing problems after AI interaction
Role of technology in pedagogy	15/22	View of AI as a collaborative tool for instructional design, not a shortcut or replacement
Designing for engagement	14/22	Deliberate effort to make problems interesting and contextually relevant for students

Reframing mathematics tasks

Many participants reported a shift in their understanding of what makes a problem valuable. Initially, they viewed problems primarily as tools for practicing procedures and obtaining correct answers. After interacting with ChatGPT, they began to see problems as opportunities for storytelling, creativity, and multiple solution paths. For example, Participant 7 reflected:

I used to think a good math problem is one that gets a quick correct answer. Now I see that an open problem with a story or dilemma can actually be more valuable. It makes students think, not just compute.

Similarly, Participant 12 noted in her interview: “ChatGPT showed me that problems don’t have to be routine. They can tell stories. I started thinking about how to make math feel like a puzzle, not just a task.” This theme reflects a movement away from procedural task design toward more imaginative and pedagogically rich problem construction. Through the lens of ZPTD, such shifts could be interpreted as movement from the self-assistance stage toward the teacher-assistance stage, where external mediation contributes to expanded pedagogical awareness.

Recognising student thinking

A second theme involved participants’ growing awareness of the need to consider students’ perspectives when designing problems. Several participants mentioned that ChatGPT’s questions prompted them to think about how students might interpret the problem, what difficulties they might encounter, and whether the language was age appropriate. Participant 4 wrote in her reflection: “ChatGPT asked me, ‘How do you think a Grade 3 student would understand this?’ I hadn’t thought about that before. I realised I was writing for myself, not for the students.” Participant 9 expressed a similar insight during his interview: “I started thinking: how will students interpret this? Will they get confused? Maybe I need to add a picture or simplify the words. That was new for me.” This theme indicates that AI mediation encouraged participants to engage in perspective-taking and anticipatory planning, key components of pedagogical reasoning. However, it is important to note that such awareness, while promising, does not necessarily guarantee that participants will consistently apply this thinking in future practice without continued support.

Confidence in task construction

Many participants reported increased confidence in their ability to design mathematical problems after working with ChatGPT. They described feeling less anxious about where to start and more capable of generating ideas. Participant 2 reflected: “Before, I was afraid my problem wouldn’t be good enough. ChatGPT gave me ideas when I got stuck. It made me feel more capable, like I could actually do this.”



Participant 16 shared a similar sentiment: “I used to just copy problems from textbooks. Now I feel like I can create my own. ChatGPT helped me see that I have ideas too.” This theme highlights the affective dimension of scaffolding, consistent with Belland et al. (2013) emphasis on motivational support in learning. For novice teachers, such confidence may be an important precursor to developing greater autonomy in instructional design.

Role of technology in pedagogy

A fourth theme captured participants’ evolving views on the role of AI in teaching. Rather than seeing ChatGPT as a replacement for teacher judgment, most participants framed it as a collaborative partner that could support their thinking. Participant 11 stated: “AI won’t replace me, but it can help me think better. It’s like having a co-teacher who gives suggestions.” Participant 20 echoed this view in her reflection: “I used to think technology was just for finding information. Now I see it can help me design lessons. But I still have to decide what’s best for my students.”

The responses in this theme suggested that participants developed a balanced understanding of AI’s pedagogical potential, recognising both its affordances and the necessity of human oversight. Such views may reflect an emerging professional agency, though claims about lasting change would require longitudinal investigation.

Designing for engagement

Finally, several participants explicitly mentioned that they now consider student engagement when designing problems. They described efforts to make problems interesting, contextually relevant, and connected to students’ lives. Participant 13 wrote: “Making problems fun made me enjoy teaching more. I want my students to enjoy math too. So, I tried to add stories they would like.” Participant 6 reflected: “I thought about what kids in my practicum school like—snacks, games, animals. So, I made problems about those things. ChatGPT helped me think of contexts I hadn’t considered.” This theme indicates a shift toward engagement-oriented thinking about teaching, where the goal is not only correct answers but also meaningful learning experiences. However, it is important to note that such intentions, while valuable, have not yet been observed in actual classroom practice.

While most participants reported positive shifts, not all reflections indicated deep transformation. A few participants provided relatively superficial reflections that largely repeated ChatGPT’s suggestions without evidence of personal appropriation. For example, Participant 10 wrote: “ChatGPT helped me make a better problem. It suggested adding more steps, so I did.” When probed in the interview about what he learned, he responded: “I learned that problems can have more steps. That’s about it.” This case suggests that for some participants, AI mediation led to surface-level changes in task design but not necessarily to deeper pedagogical reflection.

Similarly, Participant 14 expressed uncertainty about how to adapt AI-generated ideas to her classroom: “ChatGPT gave me lots of ideas, but I’m not sure which ones would actually work with my students. I felt a bit overwhelmed.” These contradictory cases highlight that the impact of AI scaffolding is not uniform and may depend on individual factors such as prior knowledge, reflective capacity, or receptiveness to feedback. They also underscore that while AI can provoke reflection, internalisation of pedagogical principles likely requires time, repeated practice, and additional support structures such as peer discussion or mentor guidance.

Examining reflections chronologically—from immediate post-task reflections to interview responses one week later—suggested a pattern of deepening reflection for many, though not all, participants. Initially, reflections tended to focus on the immediate task and the novelty of AI. For instance, Participant 1 wrote immediately after the task: “ChatGPT helped me fix my problem. It made it longer.” During an interview a week later, the same participant offered a more developed reflection: “I’ve been thinking about why ChatGPT suggested those changes. It made me realise that problems should make students think, not just calculate. I tried to apply that idea when I helped my friend with her lesson plan.” This trajectory—from task-focused to conceptually oriented reflection—was observed in approximately two-thirds of participants. It is, however, important to note that without longitudinal follow-up beyond one week, claims about lasting change or genuine internalisation remain provisional.



The presence of participants whose reflections remained superficial further cautions against overgeneralising these patterns to all individuals.

Through ZPTD, reflective shifts suggest movement from self-assistance (task completion) toward teacher-assistance (pedagogical considerations). Themes like reframing tasks and recognising student thinking indicate active construction of new understandings, not mere AI reproduction. Confidence and engagement reflect affective dimensions of learning; technology views show emerging professional agency. Variation in reflective depth and superficial responses, however, show such shifts are neither automatic nor universal. Without evidence of classroom transfer or stability over time, it is premature to claim full internalisation. Rather, ChatGPT serves as a catalyst for reflective thinking, creating professional learning opportunities that require continued practice and support. Future longitudinal research must examine whether such reflection translates into lasting instructional change.

A ZPTD-oriented Perspective

The findings can be tentatively interpreted through Warford (2011) ZPTD, which describes pre-service teachers' progression from self-assistance (relying on personal experience) through teacher assistance (external mediation) toward internalisation and recurrence (autonomous, principled practice). Table 6 summarises how the evidence aligns with each ZPTD stage.

Table 6
Mapping of Findings to ZPTD Stages

ZPTD Stage	Description	Evidence from Findings
Self-assistance	Reliance on personal experience as learners	Task A problems were predominantly procedural and decontextualised (e.g., "Ani has 5 apples"). Initial reflections focused on task completion rather than pedagogical reasoning.
Teacher Assistance	External mediation enables tasks beyond independent capability	ChatGPT provided multi-dimensional scaffolding (cognitive, metacognitive, strategic, affective, pedagogical). Task B problems showed shifts toward greater contextual relevance, problem structure, cognitive demand, and creativity.
Internalisation	Beginning to appropriate professional concepts	Reflective themes such as <i>reframing mathematics tasks</i> and <i>recognising student thinking</i> suggest emerging pedagogical awareness. However, depth of reflection varied considerably across participants.
Recurrence	Independent application and reconstruction of understandings	Limited evidence within the study's timeframe. A few participants (e.g., Participant 1) described attempting to apply ideas beyond the immediate task, but these were self-reported rather than observed.

Participants began in the self-assistance stage, drawing on learner experiences. ChatGPT functioned as teacher assistance, enabling problems beyond independent capability. Some reflective comments suggested early internalisation (e.g., articulating open-ended problems). However, these interpretations are provisional: short duration precludes claims of full internalisation or lasting change. Variation in reflective depth cautions against overgeneralisation. Rather than a human mentor analogue, ChatGPT is more accurately a symbolic mediator that expands the developmental zone when users engage critically.

Discussion

ChatGPT provided multidimensional scaffolding (cognitive, metacognitive, strategic, affective, pedagogical), helping pre-service teachers move beyond procedural problems toward more authentic,



cognitively demanding tasks. Reflective data showed shifts in thinking about problem design and student perspectives, though depth varied. Interpreted through ZPTD, AI mediation supported transition from self-assistance toward teacher assistance, with early internalisation indicators. Cognitive scaffolding ($n = 153$) helped clarify mathematical ideas. Metacognitive ($n = 115$) prompted reflection on design choices. Strategic ($n = 96$) offered heuristics. Affective ($n = 57$) sustained engagement. Pedagogical ($n = 57$) directed attention to instructional considerations

Cognitive scaffolding aligns with research on pre-service teachers' struggles to formulate coherent problems due to developing content knowledge (Kilic, 2017). Strategic scaffolding resonates with Singer et al. (2015) that novices benefit from explicit prompts to transform closed tasks into more open ones. Pedagogical scaffolding, though less frequent, is notable because anticipating student thinking and aligning tasks with goals are PCK hallmarks (Shulman, 1987) that novices rarely activate spontaneously. Response variation further suggests individual differences in MCPK may have moderated AI mediation effectiveness, though not directly measured.

The presence of affective and pedagogical scaffolding is noteworthy. Novice teachers often experience anxiety about instructional design; ChatGPT's supportive tone may have reduced this anxiety, creating a safe space for experimentation. However, scaffolding distribution reflects frequency, not quality or impact. Effectiveness likely depends on individual participant characteristics, as suggested by variation in reflective depth.

Task B improvements (increased contextual relevance, complex problem structures, higher cognitive demand, greater creativity) resonate with calls for richer mathematical tasks (Silver, 1994; Cai & Hwang, 2020). Consistent with the ZPTD framework, these shifts represent early movement from the self-assistance stage toward the teacher-assistance stage. The predominance of multi-step but still closed tasks suggests that participants had begun to appropriate some pedagogical strategies modelled by the AI but had not yet internalised them sufficiently to generate fully open-ended, non-routine problems. Reaching the recurrence stage, where such tasks are designed independently and flexibly, likely requires repeated practice and sustained scaffolding beyond a single session. AI mediation can disrupt novices' routine problem tendencies (Aktaş, 2022; Kilic, 2017). Reflective themes suggest that the pre-service teachers were beginning to appropriate pedagogical principles through reflection on their practice (Beswick, 2012; Schon, 1983). However, variation in depth of reflective evident in Participant 10's responses, suggest that AI support does not necessarily result in deep learning and that its potential mediating impact may depend on individual factors.

Viewed through ZPTD, ChatGPT functions as teacher assistance enabling tasks beyond independent capability (Task A→B), with emerging pedagogical awareness indicating early internalisation, though short duration precludes claims of lasting change. Rather than a human mentor analogue, AI can provide learning support through AI-mediated interaction, while the quality of mediation depends on user engagement, including prompting, critical evaluation, and reflection (Msambwa et al., 2025; Yelland, 2026). Three implications follow. First, AI-supported problem posing offers low-stakes experimentation, with real-time feedback scaffolding task development when human mentoring is limited. Second, response variation necessitates structured guidance: teacher educators should help pre-service teachers critically evaluate AI content and maintain pedagogical judgment using reflection activities. Third, affective scaffolding builds novice confidence, though educators must discuss AI limitations to prevent overreliance.

The findings also carry implications for assessment design in teacher education. AI-supported problem-posing tasks could be embedded as formative assessments, where pre-service teachers submit not only their final problem designs but also their interaction logs with ChatGPT. This would allow teacher educators to evaluate both the product, the posed problem, and the process, the quality of critical engagement with AI suggestions. Such assessments could capture evidence of pedagogical reasoning, reflective capacity, and professional judgment in ways that traditional task submissions cannot. Over time, repeated engagement with AI-mediated design tasks could serve as a developmental portfolio, documenting pre-service teachers' progression from procedural to more sophisticated instructional design.



Limitations

Several methodological limitations should be considered when interpreting the findings. The study's three-week timeframe precludes claims about the persistence or internalisation of observed gains; longitudinal research is needed to examine whether improvements in problem posing are sustained and transfer to classroom practice. Without a control group, observed improvements cannot be definitively attributed to ChatGPT alone; repeated practice or reflection may have contributed. Future studies should employ a quasi-experimental design to isolate AI scaffolding effects. The lack of a control group confounds ChatGPT's effect with practice/reflection. Self-report data also warrant caution. Future research should use observational measures, lesson plans, or student outcomes.

Contextual factors also bound the study. The sample, drawn from a single Indonesian teacher education institution, limits cross-cultural and cross-institutional generalisability. Moreover, the orientation session, while necessary for meaningful ChatGPT engagement, may have introduced priming effects; future research could include a control condition to isolate this influence. Moreover, the problem-posing tasks were limited to arithmetic word problems, and findings may not extend to more open-ended or advanced problem types (e.g., decimals, ratios, geometry).

Conceptually, the scaffolding typology, though literature-grounded, is not exhaustive, and the study was confined to primary arithmetic word problems; findings may not extend to more open-ended or advanced problem types. The use of a single AI tool and a consumer web interface, rather than a controlled application programming interface (API), introduced variability inherent to large language models and limited replicability. Future studies employing API-based designs with controlled parameters and multiple AI platforms would strengthen the evidence base.

Conclusion

This study provides empirical evidence that ChatGPT can serve as a multi-dimensional scaffolding tool, supporting pre-service primary teachers in developing more authentic, cognitively demanding, and pedagogically reflective mathematical problems. The findings tentatively suggest that AI mediation may help novice teachers move from self-assistance toward teacher assistance within the ZPTD framework, with some participants showing early indicators of internalisation. However, these interpretations are provisional, and the variation in participant responses underscores that AI scaffolding is most effective when integrated into broader pedagogical structures that include critical reflection, peer collaboration, and human mentoring. With thoughtful integration, generative AI holds promise as a catalyst for instructional creativity and professional growth in teacher education.

Corresponding author

Baharullah

Universitas Muhammadiyah Makassar

[Jl. Sultan Alauddin No. 259, Makassar, Sulawesi Selatan 90221], Indonesia

baharullah@unismuh.ac.id

Ethical approval

Ethical approval for this research was granted by Universitas Muhammadiyah Makassar, Indonesia. Informed consent was obtained from all participants for their participation and for the publication of their data.



Competing interests

The authors declare there are no competing interests.

References

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), Article ep429. <https://doi.org/10.30935/cedtech/13152>
- Aktaş, M. C. (2022). Problem-posing research in mathematics education: A bibliometric analysis. *Journal of Pedagogical Research*, 6(4), 217–233. <https://doi.org/10.33902/JPR.202217414>
- Anthony, G., Hunter, J., & Hunter, R. (2015). Supporting prospective teachers to notice students' mathematical thinking through rehearsal activities. *Mathematics Teacher Education and Development*, 17(2), 7–24.
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/amler.2023.02.009>
- Baharullah. (2009). Penerapan pendekatan problem posing berlatar pembelajaran kooperatif di sekolah dasar [The implementation of a problem posing approach in a cooperative learning setting in primary schools]. *Sigma: Suara Intelektual Gaya Matematika*, 1(1), 1–9. <https://journal.unismuh.ac.id/index.php/sigma/en/article/view/7339>
- Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. In A. Schleicher (Ed.), *OECD Digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain, and robots* (pp. 43–54). OECD Publishing. <https://doi.org/10.1787/f54ea644-en>
- Baumanns, L., & Rott, B. (2024). Problem-posing tasks and their influence on pre-service teachers' creative problem-posing performance and self-efficacy. *The Journal of Mathematical Behavior*, 73, 101130. <https://doi.org/10.1016/j.jmathb.2024.101130>
- Belland, B. R., Kim, C. M., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational psychologist*, 48(4), 243–270. <https://doi.org/10.1080/00461520.2013.838920>
- Beswick, K. (2012). Teachers' beliefs about school mathematics and mathematicians' mathematics and their relationship to practice. *Educational Studies in Mathematics*, 79(1), 127–147. <https://doi.org/10.1007/s10649-011-9333-2>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Buchholtz, N., & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. In E. Faggiano, A. Clark-Wilson, M. Tabach & H.-G. Weigand (Eds.), *Proceedings of the 17th ERME Topic Conference MEDA 4* (pp. 129–136). University of Bari Aldo Moro.
- Burgos, M., & Tizón-Escamilla, N. (2026). ChatGPT as a tool for posing of mathematical problems by prospective teachers. *LUMAT: International Journal on Math, Science and Technology Education*, 14(1), Article 3. <https://doi.org/10.31129/LUMAT.14.1.2942>
- Cai, J., & Hwang, S. (2020). Learning to teach through mathematical problem posing: Theoretical considerations, methodology, and directions for future research. *International Journal of Educational Research*, 102, Article 101391. <https://doi.org/10.1016/j.ijer.2019.01.001>
- Cameron, S., & Mesiti, C. (2024). What kind of mathematics teacher is ChatGPT? In J. Višňovská, E. Ross, & S. Getenet (Eds.), *Surfing the waves of mathematics education*. Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia (pp. 135–142). MERGA.
- Cameron, S., Mesiti, C., & Fijn, P. (2025). *How knowledgeable is ChatGPT 4o? Assessing the pedagogical content knowledge of a generative artificial intelligence tool*. In S. M. Patahuddin, L. Gaunt, D. Harris & K. Tripet (Eds.), *Unlocking minds in mathematics education*. Proceedings of the 47th Annual Conference of the Mathematics Education Research Group of Australasia (pp. 93–100). MERGA.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967. [https://doi.org/10.1016/S0742-051X\(02\)00053-7](https://doi.org/10.1016/S0742-051X(02)00053-7)



- Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science*, 13(1), 1–49. [https://doi.org/10.1016/0364-0213\(89\)90010-4](https://doi.org/10.1016/0364-0213(89)90010-4)
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Cumhur, F. (2022). Pre-service teachers' approaches to guiding students in problem solving process. *Problems of education in the 21st century*, 80(1), 144–161. <https://doi.org/10.33225/pec/22.80.144>
- Dilling, F., & Herrmann, M. (2024). Using large language models to support pre-service teachers mathematical reasoning—an exploratory study on ChatGPT as an instrument for creating mathematical proofs in geometry. *Frontiers in Artificial Intelligence*, 7, 1460337. <https://doi.org/10.3389/frai.2024.1460337>
- Drushlyak, M., Lukashova, T., Shamonia, V., & Semenikhina, O. (2025). ChatGPT-based simulation helps to develop the pre-service mathematics teachers' critical thinking. *International Journal of Instruction*, 18(1), 153–172. <https://doi.org/10.29333/iji.2025.1819a>
- Findell, B., Swafford, J., & Kilpatrick, J. (2001). *Adding it up: Helping children learn mathematics*. National Academies.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Kilic, C. (2015). Analyzing pre-service primary teachers' fraction knowledge structures through problem posing. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(6), 1603–1619. <https://doi.org/10.12973/eurasia.2015.1425a>
- Kusmaryono, I., Gufron, A. M., & Rusdiantoro, A. (2020). Effectiveness of scaffolding strategies in learning against decrease in mathematics anxiety level. *Numerical: Jurnal Matematika dan Pendidikan Matematika*, 4, 13–22. <https://doi.org/10.25217/numerical.v4i1.770>
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363–374. <https://doi.org/10.2307/2529786>
- Lestari, N. D. S., Suwarno, S., Pambudi, D. S., Kurniati, D., & Murtafiah, W. (2026). Decision-making skills of prospective mathematics teachers in designing numeracy tasks: A study of medium level pedagogical content knowledge. *Mathematics Teacher Education and Development*, 28(1), Article 2.
- Li, L., & Patton, W. (2025). Streamlining online astronomy education: Integrating ChatGPT and Stellarium Planetarium software to serve as a virtual teaching assistant. *Journal of Educational Technology Systems*, 53(3), 205–218. <https://doi.org/10.1177/00472395241309599>
- Lincoln, Y. S., Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.
- Lortie, D. C. (2020). *Schoolteacher: A sociological study*. University of Chicago Press.
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753–1760. <https://doi.org/10.1177/1049732315617444>
- Ming, G. K., & Mansor, M. (2023). Exploring the impact of chat-GPT on teacher professional development: opportunities, challenges, and implications. *Asian Journal of Research in Education and Social Sciences*, 5(4), 54–67.
- Msambwa, M., Wen, Z., & Kangwa, D. (2025). The impact of AI on the personal and collaborative learning environments in higher education. *European Journal of Education*, 60(1). <https://doi.org/10.1111/ejed.12909>
- Niraula, S. (2024). The impact of ChatGPT on academia: A comprehensive analysis of AI policies across UT system academic institutions. *Advances in Mobile Learning Educational Research*, 4(1), 973–982. <https://doi.org/10.25082/amler.2024.01.009>
- Nurita, T., Yuliati, L., Mahdiannur, M. A., Ilhami, F. B., Fauziah, A. N. M., Hendratmoko, A. F., & Puspitarini, S. (2024). Increasing pre-service science teacher creativity through STEM problem-solving. *Jurnal Penelitian Pendidikan IPA*, 10(1), 72–79. <https://doi.org/10.29303/jppipa.v10i1.6335>
- Patton, M. Q. (2015). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.
- Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11(1), 9–41. <https://doi.org/10.1007/s40751-025-00172-1>
- Rudolph, J., Ismail, M. F. B. M., & Popenici, S. (2024). Higher education's generative artificial intelligence paradox: The meaning of chatbot mania. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/54fs5e77>
- Schoenfeld, A. H. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of videotapes. *The Journal of the Learning Sciences*, 2(2), 179–214. https://doi.org/10.1207/s15327809jls0202_3



- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics*, 14(1), 19–28.
- Silver, E. A., Mamona-Downs, J., Leung, S. S., & Kerney, P. A. (1996). Posing mathematical problems: An exploratory study. *Journal for Research in Mathematics Education*, 27(3), 293–309. <https://doi.org/10.2307/749366>
- Singer, F. M., Ellerton, N. F., & Cai, J. (Eds.) (2015). *Mathematical problem posing: From research to effective practice*. <https://doi.org/10.1007/978-1-4614-6258-3>
- Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School*, 3(4), 268–275. <https://doi.org/10.5951/MTMS.3.4.0268>
- Subhaktiyasa, P. G., Ayu, S., Candrawati, K., Sumaryani, N. P., Sunita, W., & Syakur, A. (2023). Penerapan statistik deskriptif: Perspektif kuantitatif dan kualitatif. [The implementation of descriptive statistics: A quantitative and qualitative perspective]. *Jurnal Edukasi Matematika dan Sains*, 14(1), 96–104. <https://doi.org/10.59672/emasains.v14i1.4450>
- Suwanti, V., Sa’dijah, C., & Muksar, M. (2024). Prospective Indonesian secondary mathematics teachers’ epistemological obstacles in designing project-based numeracy tasks: A case study. *Mathematics Teacher Education and Development*, 26(2), Article 2. <https://mtd.merga.net.au/index.php/mtd/article/view/1131>
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>
- Viberg, O., Grönlund, Å., & Andersson, A. (2023). Integrating digital technology in mathematics education: a Swedish case study. *Interactive Learning Environments*, 31(1), 232–243. <https://doi.org/10.1080/10494820.2020.1770801>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Warford, M. K. (2011). The zone of proximal teacher development. *Teaching and Teacher Education*, 27(2), 252–258. <https://doi.org/10.1016/j.tate.2010.08.008>
- Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge Press.
- Wertsch, J. V. (1998). *Mind as action*. Oxford University Press.
- Williams, R. T. (2024). The ethical implications of using generative chatbots in higher education. *Frontiers in Education*, 8, Article 1331607. <https://doi.org/10.3389/feduc.2023.1331607>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yelland, N. (2026). Reimagining multiliteracies: The transformative influence of generative AI on pedagogies. *Pedagogies*, 21(2), 375–384. <https://doi.org/10.1080/1554480X.2025.2525270>
- Yin, R. K. (2018). *Case study research and applications* (Vol. 6). SAGE Publications.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1). <https://doi.org/10.1186/s41239-019-0171-0>

