ISSN. 2621-9832
JURNAL MathEdu (Mathematic Education Journal)
http://journal.ipts.ac.id/index.php/MathEdu
Vol. 8 . No. 3 November 2025

WHO IS DOING THE THINKING? AI, METACOGNITION, AND DEEP LEARNING IN MATHEMATICS EDUCATION

Oleh :

Muhammad Rizal Usman[1)*], Nur Farida[2)], Erry Hidayanto[3)]
[1,2,3]FMIPA, Pendidikan Matematika, Universitas Negeri Malang
muhammad.rizalrb.2303119@students.um.ac.id

Abstrak

Penggunaan kecerdasan artifisial (artificial intelligence/AI) dalam pendidikan matematika memungkinkan peserta didik memperoleh strategi, penjelasan, dan solusi secara cepat. Kemudahan ini memunculkan pertanyaan tentang siapa yang merencanakan, memantau, dan mengevaluasi proses berpikir ketika tugas diselesaikan dengan bantuan AI. Artikel ini bertujuan mengkaji peran metakognisi dalam penggunaan AI untuk mendukung pembelajaran mendalam dalam pendidikan matematika. Pembahasan menggunakan pendekatan kajian literatur dengan sintesis konseptual yang menghubungkan teori metakognisi dan scaffolding dengan penelitian empiris tentang pembelajaran berbantuan AI. Pembelajaran mendalam dimaknai secara pedagogis sebagai pengembangan pemahaman konseptual, keterkaitan antargagasan, kemampuan memberikan justifikasi, dan transfer pengetahuan. Argumen utama yang dikembangkan adalah bahwa kualitas pembelajaran berbantuan AI perlu ditinjau melalui keterlibatan peserta didik dalam mengatur proses berpikirnya serta kemampuan yang dapat ditunjukkan secara mandiri. Desain prompt dan scaffolding dibahas sebagai sarana untuk mendorong perencanaan strategi, monitoring pemahaman, evaluasi keluaran AI, dan revisi penyelesaian berdasarkan alasan matematis. Risiko yang ditelaah meliputi penerimaan informasi tanpa verifikasi, pengalihan regulasi berpikir kepada AI, dan ketergantungan terhadap bantuan. Artikel ini mengusulkan kerangka konseptual yang menempatkan aktivitas metakognitif sebagai mekanisme penghubung potensial antara desain bantuan AI dan pembelajaran mendalam. Implikasinya mencakup pemberian bantuan sesuai kebutuhan, pengurangan dukungan secara bertahap, serta asesmen yang memeriksa penjelasan, keputusan, revisi, dan transfer secara mandiri. Kerangka tersebut memerlukan pengujian empiris untuk memperjelas kondisi ketika AI mendukung perkembangan regulasi belajar dan pemahaman matematis.

Kata kunci*—kecerdasan artifisial, metakognisi, pembelajaran mendalam, pendidikan matematika, scaffolding*

Abstract

The use of artificial intelligence (AI) in mathematics education enables learners to obtain strategies, explanations, and solutions rapidly. This ease of access raises questions about who plans, monitors, and evaluates the thinking process when tasks are completed with AI assistance. This article aims to examine the role of metacognition in AI use to support deep learning in mathematics education. It adopts a literature-based conceptual synthesis that connects theories of metacognition and scaffolding with empirical research on AI-assisted learning. Deep learning is understood pedagogically as the development of conceptual understanding, connections among ideas, mathematical justification, and knowledge transfer. The central argument is that the quality of AI-assisted learning should be examined through learners' engagement in regulating their thinking and the capabilities they can demonstrate independently. Prompt design and scaffolding are discussed as means of encouraging strategic planning, comprehension monitoring, evaluation of AI outputs, and revision of solutions based on mathematical reasoning. The risks examined include accepting information without verification, delegating the regulation of thinking to AI, and dependence on assistance. The article proposes a conceptual framework positioning metacognitive activity as a potential mechanism linking the design of AI assistance with deep learning. Implications include providing support responsive to learners' needs, gradually fading assistance, and designing assessments that examine explanations, decisions, revisions, and independent transfer. The framework requires empirical testing to clarify the conditions under which AI supports the development of self-regulated learning and mathematical understanding.

Keywords— *artificial intelligence, metacognition, deep learning, mathematics education, scaffolding*

ISSN. 2621-9832
JURNAL MathEdu (Mathematic Education Journal)
http://journal.ipts.ac.id/index.php/MathEdu
Vol. 8 . No. 3 November 2025

1. INTRODUCTION

The emergence of artificial intelligence (AI), particularly generative AI, has expanded the forms of support available to students in mathematics learning. Through natural language conversations, students can request conceptual explanations, explore alternative strategies, and receive feedback on their proposed solutions (Zhou et al., 2025; Yoon et al., 2024). Such support provides opportunities to revisit difficulties encountered during problem solving and to review and revise their work, highlighting the need to consider the timing of assistance and the learning process (Yoon et al., 2026; Gabriel et al., 2025). The benefits of AI in mathematics learning also warrant examination in relation to students' performance after support is withdrawn, as improvements in the quality of work produced with AI assistance may not be accompanied by gains in knowledge or independent transfer of learning (Bastani et al., 2025; Fan et al., 2025). These findings draw attention to how students use AI support and the thinking processes involved in its use.

The ease of obtaining solutions raises questions about the relationship between successful task completion and the development of students' abilities. Research on peer feedback and science inquiry suggests that AI assistance may be accompanied by reliance on support and more limited engagement in reasoning, indicating a need to distinguish ease of task completion from the quality of the learning process (Darvishi et al., 2024; Stadler et al., 2024). This issue draws attention to how assistance is designed and how students use it. Research in physics education demonstrates the benefits of AI tutors designed around pedagogical principles. Evidence from writing tasks indicates that improvements in work quality do not necessarily translate into greater knowledge acquisition or transfer (Kestin et al., 2025; Fan et al., 2025). These findings across disciplines inform the evaluation of AI use in mathematics education by drawing attention to students' thinking processes and independent capabilities alongside the quality of work produced with assistance.

In mathematics education, this issue relates to the goal of developing understanding that students can articulate and apply independently. This article uses the term *deep learning* in its pedagogical sense, encompassing conceptual understanding, connections among ideas, justification, and knowledge transfer. Procedural proficiency is an important aspect of mathematical activity. Explanations of why procedures work and under what conditions they apply provide further insight into students' understanding. These forms of knowledge can develop in mutually supportive ways (Skemp, 2006; Rittle-Johnson et al., 2015). Perspectives on cognitive engagement and generative learning emphasize students' active processing of information, generation of explanations, and integration of new ideas with prior knowledge (Chi & Wylie, 2014; Fiorella & Mayer, 2016). In the context of AI, these perspectives are used to examine students' intellectual contributions during conversations, such as constructing explanations, comparing strategies, or testing claims.

Such engagement requires awareness of one's understanding and regulation of thinking processes. Metacognition encompasses knowledge of one's own cognitive processes and the ability to monitor and regulate these activities in accordance with learning goals (Flavell, 1979; Veenman et al., 2006). In mathematical problem solving, students need to recognize what they understand, identify difficulties, and determine how to proceed. This regulation can be examined through strategic planning, progress monitoring, and evaluation of results and their alignment with the goals of the problem-solving task (Artzt & Armour-Thomas, 1992; Schoenfeld, 1992). When AI contributes suggestions, this framework directs attention to students' decisions about when assistance is needed, whether the explanations received are trustworthy, and how to assess their relevance to the problem.

Research on AI-assisted learning provides reasons to pay particular attention to metacognition. Studies involving writing tasks and peer feedback suggest that improvements in work quality during AI use may be accompanied by reliance on assistance and may not necessarily reflect gains in knowledge, transfer, or the ability to work independently (Fan et al., 2025; Darvishi et al., 2024). These findings across disciplines provide a basis for examining the relationships among product quality, regulation processes, and learning outcomes in mathematics education. Research on mathematical problem solving and programming education also indicates that high-performing students engage more actively in interpreting, evaluating, or critically examining AI-generated information to develop their solutions (Zhou et al., 2025; Li et al., 2025). These differences in engagement patterns highlight the relevance of evaluative engagement in AI-assisted learning but do not establish that metacognitive activity causes improvements in performance.

Scaffolding provides a foundation for designing assistance that sustains student engagement through support tailored to their needs and developing capabilities (Wood et al., 1976; Van de Pol et al., 2010). Contingency, the gradual withdrawal of support (*fading*), and the transfer of responsibility are key principles for examining such assistance. Applying these principles to AI requires attention to how well prompts and hints align with students' responses and demonstrated abilities. Prompt design can guide the

provision of hints and shape learning interactions, as research on AI tutors highlights the importance of assistance informed by pedagogical considerations (Bastani et al., 2025; Kestin et al., 2025). In this article, scaffolding principles serve as a conceptual basis for examining how AI can elicit initial strategies, provide limited hints, and encourage students to examine their reasoning. The effectiveness of these practices requires further assessment through the quality of interactions and evidence of student understanding.

The sequence of AI use also warrants consideration in the design of support. Research on problem solving prior to instruction and AI use in mathematical proving provides a basis for considering students' initial efforts and how subsequent assistance responds to the work they have already produced (Loibl et al., 2017; Yoon et al., 2026). Although these studies address different contexts, both draw attention to the relationship among initial activity, support needs, and opportunities to revisit solutions. In the context of AI, students' initial attempts can serve as a basis for comparing strategies and identifying aspects that require improvement, drawing on research on problem solving prior to instruction and the comparison of solution methods in mathematics (Sinha & Kapur, 2021; Rittle-Johnson & Star, 2007). The benefits of these attempts require assessment through evidence of understanding, distinguishing students' perceptions of the usefulness of assistance from the abilities they demonstrate through independent explanations or problem solving. Feelings of learning and performance during instruction do not always reflect lasting learning (Deslauriers et al., 2019; Soderstrom & Bjork, 2015). On this basis, this article examines the timing of AI assistance and its role in relation to the work students have already completed.

These studies provide complementary evidence but differ in task context, support design, and measures of learning outcomes. Some examine learning outcomes through pedagogically designed AI tutors and changes in performance after support is withdrawn (Kestin et al., 2025; Darvishi et al., 2024). Others relate work quality to learning regulation, student engagement, and knowledge transfer (Fan et al., 2025; Li et al., 2025). These differences limit the conclusion that particular prompts directly produce deep learning through metacognition. This article focuses on the relationships among these components by distinguishing the design of the support provided, the metacognitive activities students undertake, and the understanding they can demonstrate. The question *Who is doing the thinking?* is used to examine how responsibility for selecting strategies, evaluating claims, and revising solutions is distributed during interactions with AI.

This article examines the role of metacognition in the use of AI to support deep learning in mathematics education. The discussion connects theories of metacognition and scaffolding with empirical research to examine opportunities, risks, and principles for designing support. The article also proposes a conceptual framework that positions metacognitive activity as a potential mechanism linking AI assistance to deep learning. Its contributions include a conceptual synthesis, pedagogical illustrations, and directions for empirical research that assesses students' thinking processes and their ability to work independently.

2. METHOD

This study is designed as a narrative literature review with conceptual synthesis to examine the relationships among artificial intelligence use, metacognitive activity, and deep learning in mathematics education. This approach was selected to integrate theoretical perspectives and empirical findings across different contexts and methods and to inform the development of a conceptual framework. Literature reviews and conceptual articles require clarity in source selection, the role of theory, and the integration of ideas to ensure that the resulting arguments are traceable.

Literature searches will be conducted using ERIC, the ScienceDirect platform, and Google Scholar. Priority will be given to AI publications from 2020 to 2026, up to the final search date. Earlier publications will be included to establish the theoretical foundations of metacognition, scaffolding, and deep learning. Search terms will include *generative artificial intelligence*, *AI tutoring*, *mathematics education*, *metacognition*, *self-regulated learning*, *scaffolding*, *prompt design*, *conceptual understanding*, and *knowledge transfer*. These terms will be combined according to the search capabilities of each service. Additional sources will be identified by examining the reference lists of relevant articles. Search dates, keyword combinations, and source selection decisions will be documented to ensure the traceability of the review process.

Articles will be selected based on their relevance to AI support design, metacognitive activity, or the processes and outcomes of deep learning. Primary sources will comprise peer-reviewed journal articles available in full text that provide sufficient theoretical or methodological detail. Priority will be given to research in mathematics education. Studies from other disciplines may be included as supporting evidence when they address relevant mechanisms, with explicit attention to contextual differences and limits to their applicability. Articles will be excluded if they address deep learning solely as a machine learning technique, assess AI performance in solving problems without examining student learning, or fall outside the scope of

the review. Selection will involve title and abstract screening, followed by full-text review and duplicate removal.

Information from the selected articles will be compiled in a matrix covering publication details, context and participants, research design, forms of AI support, prompt characteristics, metacognitive activities, learning outcome indicators, main findings, and limitations. Critical appraisal will consider the alignment among research questions, methods, outcome measures, and conclusions. Particular attention will be paid to distinctions among performance during AI use, independent performance after support is withdrawn, participants' perceptions, and evidence of understanding and transfer. Studies reporting associations will be distinguished from those whose designs support causal inference.

The analysis will follow a thematic approach, using initial categories comprising AI support design, planning, monitoring, evaluation, conceptual understanding, justification, and transfer. These categories may be refined during reading to accommodate findings beyond the initial framework. The synthesis will compare convergent and divergent findings, including evidence that challenges or qualifies the proposed arguments. The analysis will inform the formulation of conceptual relationships and the conditions that may influence them. The resulting framework will be presented as a proposal for empirical testing.

3. RESULTS AND DISCUSSION

The empirical studies reviewed reveal variations in the outcomes of AI use across task contexts, forms of support, and learning indicators. In mathematics and physics education, AI tutors can support performance or learning outcomes, but improvements during assisted work need to be distinguished from independent performance after support is withdrawn (Bastani et al., 2025; Kestin et al., 2025). Research in other contexts also indicates that work quality, engagement in learning regulation, and knowledge transfer may follow different patterns (Fan et al., 2025; Li et al., 2025). These findings provide a basis for examining the relationships among the support provided, students' thinking processes, and the understanding they can demonstrate independently.

Table 1 summarizes the empirical studies informing the discussion. Direct evidence from mathematics education is distinguished from supporting research across disciplines to clarify the scope of interpretation. The table presents a preliminary synthesis of the articles identified to date rather than the final results of the full literature selection process.

Table 1. Summary of Empirical Studies on AI Support, Metacognition, and Learning Outcomes

Study	Context and Method	Main Findings	Relevance to the Review
Bastani et al. (2025)	Nearly 1,000 high school students; a randomized experiment in mathematics learning involving GPT Base, GPT Tutor, and control groups.	AI improved practice performance. After support was withdrawn, the GPT Base group scored approximately 17% lower than the control group. Pedagogical safeguards in GPT Tutor largely mitigated this negative effect, but independent examination results did not differ significantly from those of the control group.	Support design needs to be assessed alongside independent performance. The study did not test metacognition as a mediator.
Zhou et al. (2025)	Twenty-nine university students; a mixed-methods study of ChatGPT-assisted mathematical problem solving.	The researchers reported improved performance. High-performing students engaged more frequently in interpretation and evaluation, whereas low-performing students showed more procedural interaction patterns.	Interaction patterns are relevant to understanding how students use support. Associations between activity and performance do not establish causality.
Yoon et al. (2026)	Thirty university students; a comparison of sequences of AI use in mathematical proving.	Students who drafted proofs first used AI to review and revise their work. Those who viewed AI-generated proofs first tended to use them as a reference.	Highlights the importance of the timing and role of support. Participants' perceived conceptual benefits are distinct from understanding measured through tests.

Fan et al. (2025)	A total of 117 university students; a randomized experiment involving writing tasks with several forms of support.	The ChatGPT group showed greater improvements in essay scores but no significant advantage in knowledge gains or transfer. Learning regulation processes differed across groups.	Provides evidence from another discipline on distinctions among product quality, learning regulation, and learning outcomes.
Darvishi et al. (2024)	A total of 1,625 university students across 10 courses; a randomized experiment involving peer feedback.	Participants tended to rely on AI assistance. A self-monitoring checklist helped after support was withdrawn but did not match performance under AI assistance. Combining the two was no more effective than AI alone.	Relevant to reliance on assistance and the transfer of responsibility. The system used was not a generative conversational tutor.
Stadler et al. (2024)	Ninety-one university students; an experiment comparing ChatGPT and Google use in socioscientific inquiry.	The ChatGPT group reported lower cognitive load but produced weaker reasoning or justifications than the web search group.	Reduced task demands need to be distinguished from depth of engagement. The study did not specifically test metacognitive scaffolding.
Kestin et al. (2025)	A total of 194 university students; a randomized experiment with a crossover design in physics education.	An AI tutor designed around pedagogical principles produced greater learning gains on post-instruction tests in less time.	Provides supporting STEM evidence on the potential of tutor design. The study did not establish effects on long-term retention, transfer, or metacognitive mediation.
Li et al. (2025)	A quasi-experimental study of independent programming learning, including an analysis of learning process patterns.	AI use supported performance and self-efficacy but showed limitations in long-term transfer. High-performing students engaged more actively in critically examining and processing AI outputs.	Provides evidence from another discipline concerning critical engagement, reliance on assistance, and transfer.

AI Support and the Development of Mathematical Understanding

The findings in Table 1 indicate that the benefits of AI support need to be interpreted in relation to the indicators used. Research on writing tasks and programming education shows that improvements in work quality or performance may coexist with limitations in knowledge development and transfer (Fan et al., 2025; Li et al., 2025). This distinction is important in mathematics education because correct answers can emerge from varying levels of engagement. Students may understand conceptual relationships, follow the guidance provided, or use a solution without being able to explain its underlying reasoning. Assessing deep learning requires evidence that distinguishes among these possibilities.

Conceptual understanding and procedural proficiency are mutually supportive. The ability to execute procedures can contribute to the development of understanding, while an understanding of mathematical relationships can support more appropriate procedural choices and applications (Skemp, 2006; Rittle-Johnson et al., 2015). AI support can be examined in terms of its contribution to this relationship. Step-by-step explanations need to be accompanied by opportunities for students to articulate their reasoning, examine the conditions under which a strategy applies, and connect the representations used. Within the framework of this article, deep learning encompasses conceptual understanding, justification, and the application of knowledge in different situations.

Evidence of understanding also needs to be distinguished from perceptions of the benefits of support. Research on learning experiences and the relationship between performance and learning shows that feelings of learning and fluency during task completion do not always reflect lasting competence (Deslauriers et al., 2019; Soderstrom & Bjork, 2015). Satisfaction with AI responses provides information about the user experience but cannot replace the assessment of students' abilities. Independent tasks,

ISSN. 2621-9832
JURNAL MathEdu (Mathematic Education Journal)
http://journal.ipts.ac.id/index.php/MathEdu
Vol. 8 . No. 3 November 2025

explanations of reasoning, and application to new problems can provide evidence that complements the work produced while support is available.

Metacognitive Engagement in Interactions with AI

Metacognition provides a framework for examining how students use support and regulate their activities. Knowledge of their own abilities, task characteristics, and strategies informs actions taken to monitor and control learning processes (Flavell, 1979; Veenman et al., 2006). In interactions with AI, students need to determine what information they require, recognize gaps in their understanding, and assess whether the support received addresses their needs. This engagement can be examined through decisions made during task completion, alongside students' reports of understanding the material.

During planning, attention is directed toward goal setting and strategy selection. Monitoring involves checking progress and identifying errors or gaps in understanding, while evaluation involves assessing results and determining whether revisions are needed. Research on mathematical problem solving identifies the regulation of these activities as an important element in understanding the solution process (Artzt & Armour-Thomas, 1992; Schoenfeld, 1992). **Figure 1 illustrates how metacognitive knowledge informs these regulatory activities and identifies learner actions through which their engagement can be examined during interactions with AI.** The return arrow indicates that students may revisit their plans as difficulties or new information emerge, rather than follow a fixed sequence. Checking claims, explaining decisions, changing strategies, and revising reasoning provide observable evidence to be interpreted alongside students' explanations and self-reports.

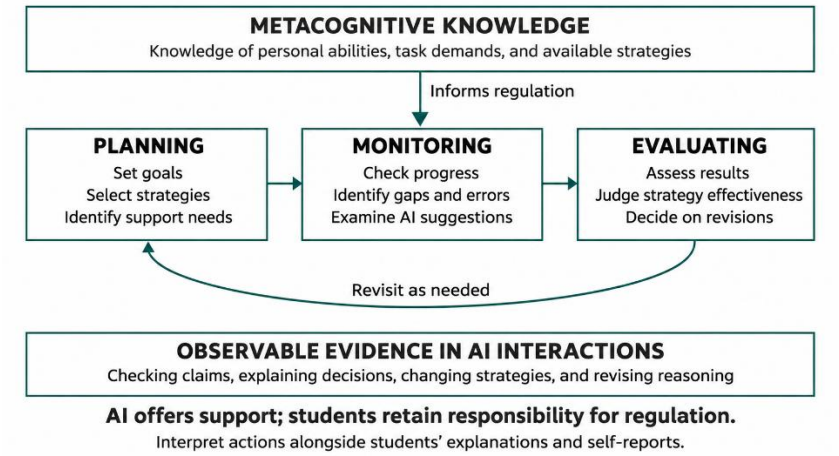


Figure 1. Metacognitive knowledge, regulation, and observable learner actions in interactions with AI

Findings concerning reliance on assistance and reduced depth of reasoning highlight the need to sustain students' engagement in evaluating information. Research on peer feedback and science inquiry shows that support can improve or facilitate task completion without ensuring corresponding competence when assistance is reduced or reasoning is assessed (Darvishi et al., 2024; Stadler et al., 2024). These findings across disciplines cannot be directly generalized to mathematics learning but provide a basis for examining the transfer of responsibility. Metacognitive questions need to be followed by concrete actions, such as checking claims, changing strategies, or revising reasoning.

Scaffolding and Prompt Design as Learning Support

Scaffolding emphasizes assistance that enables students to complete aspects of a task they cannot yet perform independently. Contingent support, gradual withdrawal of assistance, and the transfer of responsibility provide a basis for assessing the quality of such support (Wood et al., 1976; Van de Pol et al., 2010). Applying these principles to AI requires information about students' responses and developing abilities. The same hint may provide adequate support for one student but too little or too much for another.

Research on AI tutors highlights the importance of pedagogical design in structuring interactions and assistance. Tutors incorporating hints and safeguards designed for learning produce different outcome patterns from general AI access. Tutors informed by learning principles can also support learning gains in specific contexts (Bastani et al., 2025; Kestin et al., 2025). This evidence supports examining the purpose

ISSN. 2621-9832
JURNAL MathEdu (Mathematic Education Journal)
http://journal.ipts.ac.id/index.php/MathEdu
Vol. 8 . No. 3 November 2025

of assistance and the activities students are asked to undertake. These studies have not established that any particular prompt type, in isolation, causes improvements in metacognition or deep learning.

The timing of assistance also warrants consideration. Research on the sequence of AI use in mathematical proving and on problem solving prior to instruction indicates that initial work can provide a basis for subsequent learning (Yoon et al., 2026; Loibl et al., 2017). Students have ideas to compare and revise, while those providing support gain insight into their thinking. Within the proposed framework, prompts can direct AI to elicit initial strategies, provide limited hints, and encourage students to examine their reasoning. Subsequent withdrawal of support needs to be guided by students' demonstrated abilities.

Conceptual Synthesis and Implications for Deep Learning

The research synthesis identifies three components that need to be distinguished: AI support design, students' metacognitive activities, and evidence of deep learning. Support design encompasses the purpose of questions, the form of hints, the timing of assistance, and its gradual withdrawal. Metacognitive activities involve decisions about planning, monitoring, evaluating, and revising solutions. Evidence of deep learning concerns conceptual understanding, justification, retention, and transfer, assessed through appropriate measures. The proposed relationships among these components, with student responsibility extending throughout the interaction, are presented in Figure 2. These relationships constitute a conceptual proposal rather than a fully established causal pathway.

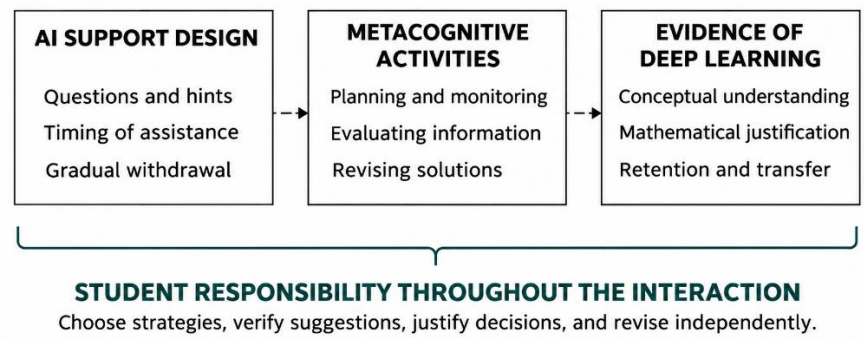


Figure 2. Proposed conceptual relationships among AI support design, metacognitive activities, and evidence of deep learning.

The question *Who is doing the thinking?* is used to examine the distribution of responsibility within these interactions. Reflective questions generated by AI are insufficient to demonstrate that students engage in metacognitive regulation. Attention needs to focus on their subsequent responses and decisions: whether they check information, explain their reasons for accepting or rejecting suggestions, and revise solutions based on mathematical considerations. This analysis allows interaction quality to be assessed through students' activities rather than solely through the characteristics of AI outputs.

Implications for teaching include opportunities to formulate initial strategies, receive support tailored to need, examine reasoning, and work independently. Assessment can combine completed solutions with explanations, records of revision, and tasks requiring the application of concepts under different conditions. The limitations of the evidence require consideration because the studies reviewed vary in context, duration, AI systems, and outcome measures. Future research needs to test whether metacognitive activity serves as a mechanism linking support design to deep learning and identify the conditions under which this relationship emerges.

4. CONCLUSION

This review indicates that the benefits of AI in mathematics education need to be examined through the quality of learning processes and the abilities students can demonstrate independently. Improved performance while support is available is insufficient to establish gains in conceptual understanding, justification, retention, and knowledge transfer. Metacognition is central to understanding how students identify their support needs, select strategies, evaluate AI outputs, and revise solutions based on mathematical reasoning. Scaffolding and prompt design can support planning, monitoring, and evaluation through reflective questions, tailored hints, and the gradual withdrawal of assistance. Their effectiveness needs to be examined through students' responses and decisions, alongside assessments of understanding after support is reduced. The proposed conceptual framework positions metacognitive activity as a potential

mechanism linking AI support design to deep learning. Variation in the contexts and methods of the studies reviewed limits causal inference. Empirical testing is needed to clarify these relationships and the conditions under which they apply while preserving students' responsibility for regulating their mathematical thinking.

5. REFERENSI

Artzt, A. F., & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction, 9*(2), 137–175. https://doi.org/10.1207/s1532690xci0902_3

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

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

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*, 717–741. https://doi.org/10.1007/s10648-015-9348-9

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gabriel, F., Kennedy, J., Marrone, R., & Leonard, S. (2025). Pragmatic AI in education and its role in mathematics learning and teaching. *npj Science of Learning, 10*, Article 26. https://doi.org/10.1038/s41539-025-00315-4

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, Article 17458. https://doi.org/10.1038/s41598-025-97652-6

Li, S., Liu, J., & Dong, Q. (2025). Generative artificial intelligence-supported programming education: Effects on learning performance, self-efficacy and processes. *Australasian Journal of Educational Technology, 41*(3), 1–25. https://doi.org/10.14742/ajet.9932

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715. https://doi.org/10.1007/s10648-016-9379-x

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*, 587–597. https://doi.org/10.1007/s10648-015-9302-x

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Macmillan.

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Skemp, R. R. (2006). Relational understanding and instrumental understanding. *Mathematics Teaching in the Middle School, 12*(2), 88–95. https://doi.org/10.5951/MTMS.12.2.0088

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, Article 108386. https://doi.org/10.1016/j.chb.2024.108386

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*, 271–296. https://doi.org/10.1007/s10648-010-9127-6

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14. https://doi.org/10.1007/s11409-006-6893-0

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

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM–Mathematics Education, 56*, 1531–1551. https://doi.org/10.1007/s11858-024-01629-0

Yoon, H., Lee, Y., Lee, K., & Kwon, O. N. (2026). Before or after? Beyond timing of students' interaction with generative artificial intelligence in proving mathematical statements. *ZDM–Mathematics Education, 58*, 945–962. https://doi.org/10.1007/s11858-026-01800-9

Zhou, R., He, X., Fan, Q., Li, Y., Li, Y., Xiao, X., & Fang, J. (2025). Exploring ChatGPT-facilitated scaffolding in undergraduates' mathematical problem solving. *Journal of Computer Assisted Learning, 41*(4), Article e70077. https://doi.org/10.1111/jcal.70077