



Examining the Association Between AI-Supported Cognitive Scaffolding and Conceptual Understanding in Secondary Science Learning

Anggi Widiarni¹ and Yuli Yanti²

¹Universitas Negeri Yogyakarta, Indonesia

²Universitas Islam Negeri Raden Intan Lampung, Indonesia

Received: 10 February 2026

Revised: 27 March 2026

Accepted: 23 April 2026

Online: 31 May 2026

Abstract

Artificial intelligence (AI) is increasingly used to support adaptive feedback, prompts, and learning analytics in science education, yet the mechanisms linking AI-supported scaffolding to conceptual understanding remain underexplored. This study examined whether metacognitive awareness mediated the association between AI-supported cognitive scaffolding and conceptual understanding, and whether prior knowledge moderated the scaffolding–metacognition pathway. A quasi-experimental pretest–posttest control-group design was conducted with 214 Grade 9–10 students from four intact science classes in Bandar Lampung, Indonesia. Students were assigned to either an AI-scaffolding condition ($n = 108$) or a conventional instruction condition ($n = 106$) over a 12-week intervention. Conceptual understanding was assessed using a 40-item curriculum-aligned test ($\alpha = .87$), while metacognitive awareness was measured using a 30-item adapted inventory ($\alpha = .82$). Results showed that students in the AI-scaffolding condition achieved higher post-test scores and greater learning gains. Metacognitive awareness partially mediated this association, and the indirect effect was stronger among students with lower prior knowledge.

Keywords: adaptive learning; artificial intelligence in education; cognitive scaffolding; conceptual understanding; metacognitive awareness

Corresponding Author:

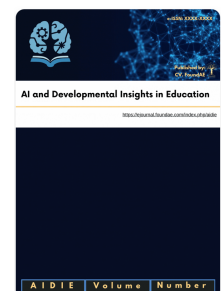
Anggi Widiarni

Email: anggiwidyb@gmail.com

 <https://orcid.org/0009-0002-8608-493X>

Author Note:

This study received academic and institutional support from the Faculty of Tarbiya and Teacher Training. The authors declare that they have no conflicts of interest. Correspondence concerning this article should be addressed to Anggi Widiarni.



Introduction

Recent work in artificial intelligence in education increasingly treats AI not as a single instructional intervention, but as a heterogeneous ecology of intelligent tutoring systems, chatbots, recommender systems, automated assessment, learning analytics, voice agents, avatars, and generative AI tools. This distinction is essential because different AI configurations support different forms of learning: some merely automate feedback, whereas others can augment explanation, diagnosis, reflection, and learner regulation. Reviews of AI in education show rapid growth in applications across school and higher education settings, but they also caution that evidence of learning impact remains uneven across technologies, domains, age groups, and implementation models (Chiu et al., 2023; Crompton & Burke, 2023; Ng et al., 2023a; Wang et al., 2024; Yim & Su, 2025). Accordingly, the central issue is no longer whether AI can be used in education, but which pedagogical mechanisms make AI-supported learning conceptually meaningful, developmentally appropriate, and instructionally defensible.

This issue is particularly important in science learning, where conceptual understanding requires students to reorganize prior ideas, connect symbolic and visual representations, identify misconceptions, and justify explanations using disciplinary evidence. AI-supported cognitive scaffolding is theoretically relevant to these demands because adaptive hints, diagnostic feedback, and reflective prompts can help learners regulate attention, compare alternative explanations, and monitor the adequacy of their reasoning. Recent STEM-focused evidence suggests that generative AI and adaptive tools can support conceptual clarification and misconception repair when embedded within structured inquiry or teacher-guided learning designs; however, the same evidence also shows limitations in advanced reasoning, numerical problem solving, and unguided student reliance on AI-generated outputs (El Fathi et al., 2025; Huang et al., 2023; Kasneci et al., 2023; Stadler et al., 2024; Yan et al., 2024).

The current literature therefore presents a more nuanced picture than early celebratory accounts of AI-enhanced learning. Meta-analyses and systematic reviews report positive effects of AI chatbots, personalized recommendations, and ChatGPT-based interventions on selected learning outcomes, yet the magnitude and durability of these effects vary substantially by task type, intervention design, outcome measure, and level of teacher orchestration (Deng et al., 2025; Labadze et al., 2023; Wu & Yu, 2024). Critical syntheses further warn that performance improvements during AI-supported tasks should not be equated automatically with durable learning, particularly when studies rely on short interventions, self-reported usefulness, or immediate post-tests without transfer and retention measures (Bauer et al., 2025; Yan et al., 2024). For secondary science education, this means that AI scaffolding must be examined not only as an outcome-producing tool but also as a learning process that may activate, redirect, or sometimes bypass students' cognitive effort.

A key mechanism that may distinguish productive AI scaffolding from superficial AI assistance is metacognitive awareness. When AI feedback is designed as a prompt for monitoring, explanation, and revision, students may become more capable of identifying what they know, what remains uncertain, and which strategy should be used next. This mechanism is increasingly emphasized in research on self-regulated learning and generative AI. Recent studies indicate that metacognitive support improves learners' self-regulated learning experiences in generative AI environments, while a recent meta-analysis of AI support for self-regulated learning reports that effects are positive but heterogeneous and strongly dependent on the design of scaffolding across forethought, performance, and reflection phases (Xu et al., 2025; Xu et al., 2026). These findings suggest that the educational value of AI is likely

mediated by learners' regulatory engagement rather than produced solely by the availability of automated responses.

Prior knowledge is also likely to shape how students benefit from AI-supported scaffolding. Learners with limited prior knowledge may need structured hints, timely corrective feedback, and explicit metacognitive prompts to locate misconceptions and maintain progress during conceptually demanding science tasks. Learners with stronger prior knowledge, by contrast, may interpret AI feedback more selectively, evaluate explanations more critically, or require less external scaffolding. Research on AI literacy, personalized learning, and students' conceptions of AI suggests that learner characteristics influence how AI tools are interpreted, trusted, and used (Markauskaite et al., 2022; Mertala et al., 2022; Nguyen et al., 2023; Shoufan, 2023). Thus, the same AI intervention may function differently across students depending on their baseline conceptual resources and capacity to regulate learning.

Against this background, the unresolved problem is not simply whether AI-supported cognitive scaffolding improves science achievement, but how such scaffolding is associated with conceptual understanding, through which learner-level mechanism, and under what prior-knowledge condition. Much existing research has examined AI effects as direct treatment contrasts, while fewer studies have tested conditional explanatory models that connect AI scaffolding, metacognitive awareness, and prior knowledge within a single empirical framework. Addressing this problem is especially important in secondary education, where students are still developing disciplinary reasoning, metacognitive control, and critical AI literacy. The present study responds to this need by testing a moderated mediation model in which instructional condition is associated with conceptual understanding, metacognitive awareness serves as a mediator, and prior knowledge moderates the pathway from instructional condition to metacognitive awareness.

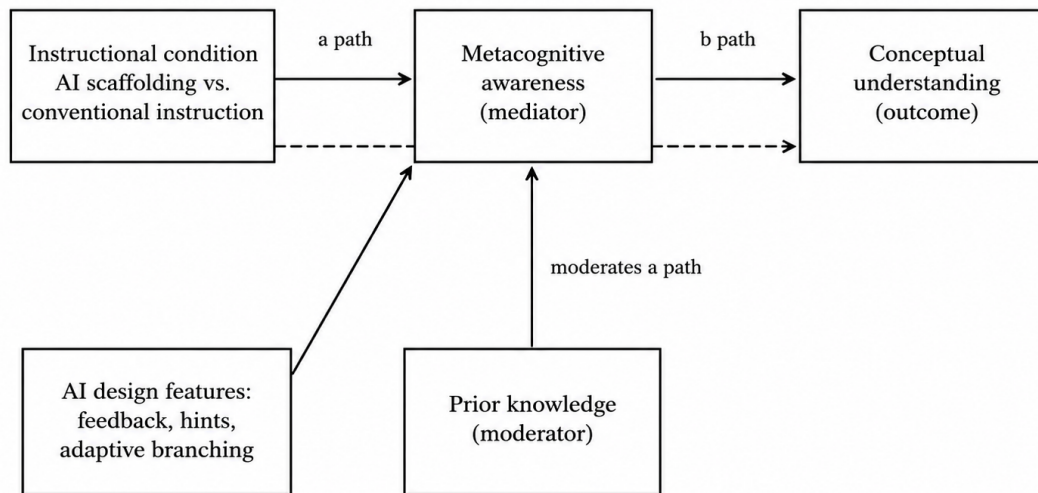
The study was guided by three hypotheses. H1 predicted that students in the AI-scaffolding condition would show larger conceptual understanding gains than students in conventional instruction. H2 predicted that metacognitive awareness would partially mediate the association between AI scaffolding and conceptual understanding. H3 predicted that prior knowledge would moderate the mediated pathway, such that the indirect association through metacognitive awareness would be stronger among students with lower prior knowledge. By focusing on mechanism and boundary condition rather than treating AI as a universally beneficial technology, the study aims to provide a more cautious and instructionally useful account of AI-supported science learning for secondary science contexts.

Method

Research Design

This study used a quasi-experimental pretest-posttest control-group design. Four intact science classes from two public secondary schools were assigned to either an AI-scaffolding condition or a conventional instruction condition. Because assignment occurred at the class level rather than the individual-student level, causal language is treated cautiously throughout the manuscript. The intervention lasted 12 weeks and included 36 instructional sessions of 80 minutes each.

The analytic model that guided the quasi-experimental design and subsequent conditional process analysis is presented in Figure 1. Placing the model here clarifies the role of each variable before the sampling, measurement, and analytic procedures are described.

Figure 1*Moderated mediation model examined in the study*

Note. Solid arrows indicate hypothesized paths; the dashed arrow indicates the direct effect.

This model therefore guides the subsequent subsections, which describe the participants, instruments, intervention procedures, and analyses used to test the specified pathways.

Participants and Context

Participants were 214 secondary school students in Grades 9-10 in Bandar Lampung, Indonesia. Two classes participated in the AI-scaffolding condition ($n = 108$) and two classes in conventional instruction ($n = 106$). The final sample included 109 female and 105 male students (M age = 14.83 years, $SD = 0.91$, range 14-16). Participation required parental consent and student assent. Students requiring individualized educational programming were not included in the intervention sample because the platform and assessment instruments had not been adapted for those needs; this exclusion limits generalizability and should be addressed in future inclusive designs. The distribution of key demographic characteristics across the two instructional conditions is summarized in Table 1.

Table 1*Demographic characteristics of participants by condition*

Characteristic	AI-scaffolding ($n = 108$)	Conventional ($n = 106$)
Age, M (SD)	14.81 (0.89)	14.85 (0.93)
Female, n (%)	55 (50.9%)	54 (50.9%)
SES: Low, n (%)	41 (38.0%)	41 (38.7%)
Prior science achievement, M (SD)	71.4 (8.3)	70.9 (8.7)

Note. SES = socioeconomic status. Baseline differences were small and not statistically significant in the available analyses.

The two groups were broadly comparable on observable baseline characteristics, which supports but does not prove the adequacy of class-level assignment. Remaining unobserved classroom differences are therefore addressed later as a limitation.

Sampling and Recruitment

Cluster sampling was used because classes were the practical instructional unit. Schools were selected purposively to be broadly comparable in urbanicity, public funding status, and prior science achievement. Recruitment involved letters to school principals, teacher briefings, and parental consent forms. The participation rate was 96.0% (214 of 223 eligible students).

Sample Size and Missing Data

An a priori power calculation for the main between-condition comparison indicated that approximately 204 students would be sufficient to detect a medium effect at $\alpha = .05$ with power = .80. The achieved sample of 214 students was sufficient for the main comparison, but the moderated mediation analysis should be interpreted as model-based and exploratory because cluster-level design effects and classroom nesting were not fully modeled. Missing outcome data were below 2% and were addressed using multiple imputation with 20 imputations after diagnostics suggested no systematic missingness pattern.

Measures

Conceptual understanding was assessed using the Science Conceptual Understanding Test, a 40-item curriculum-aligned multiple-choice instrument targeting Grade 9-10 physics and chemistry concepts. Content validity was reviewed by five science education specialists (content validity ratio = .87), and pilot testing with 62 non-participant students yielded internal consistency of $\alpha = .87$. Because this instrument was researcher-developed, score interpretations should remain context-bound until external validation is completed.

Metacognitive awareness was measured using a 30-item adapted inventory covering metacognitive knowledge and metacognitive regulation. Internal consistency in the present sample was $\alpha = .82$ for the total scale. Prior knowledge was measured at baseline using a 20-item criterion-referenced test reviewed by two science education specialists (content validity index = .91; $\alpha = .78$).

Table 2 reports the descriptive and internal-consistency information for the study instruments. This summary is included because the interpretation of the findings depends on whether the measures showed acceptable reliability in the present sample.

Table 2

Psychometric properties of study instruments

Instrument / subscale	Items	M	SD	Range	alpha
Conceptual understanding pre-test	40	52.3	8.76	20-90	.87
Conceptual understanding post-test	40	71.4	11.24	30-98	.87
Metacognitive awareness total	30	62.8	9.41	18-90	.82
Metacognitive knowledge	15	31.4	5.12	9-45	.79
Metacognitive regulation	15	31.4	5.18	9-45	.80
Prior knowledge test	20	57.1	9.33	20-90	.78

Note. M and SD reflect baseline values unless otherwise indicated.

The coefficients indicate acceptable internal consistency for all measures used in the main analyses. However, because the conceptual understanding and prior-knowledge tests were locally developed, the values should be treated as sample-specific evidence rather than definitive validation.

Intervention Procedures

Data collection occurred across three phases during the 2023-2024 academic year. In Week 1, all participants completed the prior knowledge test, conceptual understanding pre-test,

and metacognitive awareness inventory. During Weeks 2-11, the AI-scaffolding condition used an adaptive learning platform that provided individualized hint sequences, immediate corrective feedback, reflection prompts, and adaptive question branching based on learner responses. The conventional condition received equivalent science content through teacher-led exposition, textbook exercises, and whole-class discussion. Teachers received standardized preparation for their respective conditions. Classroom observations indicated acceptable implementation consistency (Cohen's kappa = .84). In Week 12, students completed the conceptual understanding post-test and metacognitive awareness post-measure.

Data Analysis

Preliminary analyses examined distributional assumptions, missingness, and baseline equivalence. H1 was evaluated using gain-score and post-test comparisons with effect sizes. Because pre-test scores were similar but not experimentally controlled at the individual level, post-test comparisons were interpreted alongside gain-score evidence rather than as stand-alone causal estimates. H2 was tested using mediation analysis with 5,000 bootstrap resamples. H3 was tested using moderated mediation, with prior knowledge treated as a continuous baseline moderator. All inferential results were interpreted in light of the class-level assignment, the limited number of clusters, and the locally developed outcome measure.

Ethical Considerations

Ethical approval was obtained from the institutional review board of Universitas Islam Negeri Raden Intan Lampung. Written informed consent was obtained from parents or legal guardians, and written assent was obtained from students. Data were anonymized before analysis. Because participants were minors, teachers supervised platform use and no automated recommendation was treated as a substitute for teacher judgment.

Results

Participant Flow and Preliminary Analyses

Of 223 eligible students, 214 provided consent and assent. Five students were absent for more than three sessions, but intention-to-treat analyses yielded a substantively similar pattern to per-protocol analyses. Missing outcome data were minimal and imputed. Descriptive statistics and correlations are reported in Table 3.

Table 3

Descriptive statistics and intercorrelations among key variables

Variable	M	SD	Range	1	2	3	4
1. Condition (AI = 1)	0.50	0.50	-	-			
2. Prior knowledge	57.1	9.33	20-90	.03	-		
3. Metacognitive awareness	62.8	9.41	18-90	.41***	.38***	-	
4. Conceptual understanding post-test	71.4	11.24	30-98	.62***	.44***	.57***	-

Note. Condition was coded 0 = conventional and 1 = AI-scaffolding. *** $p < .001$.

The correlation pattern is consistent with the proposed model: instructional condition, prior knowledge, and metacognitive awareness were positively related to post-test conceptual understanding. These zero-order associations justified proceeding to the mediation and moderated mediation analyses while retaining cautious interpretation.

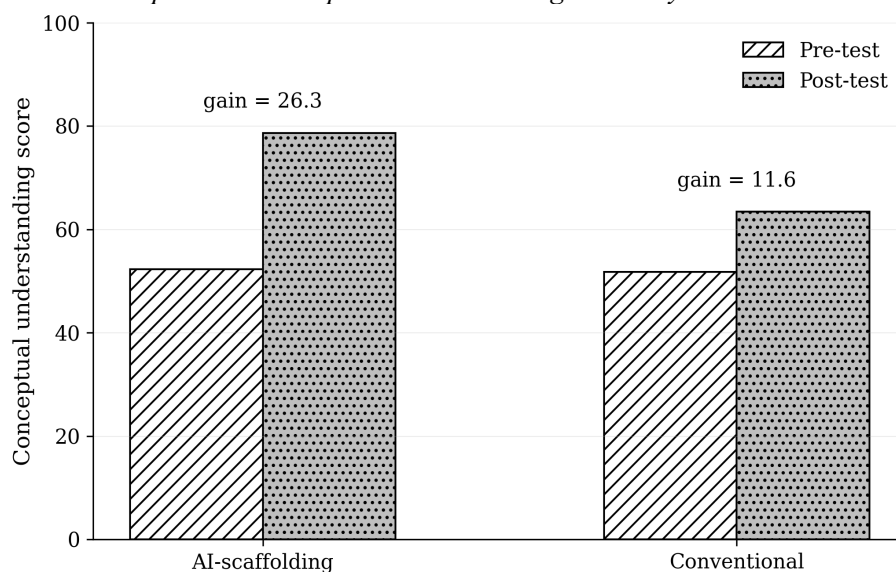
Conceptual Understanding Gains

Students in the AI-scaffolding condition obtained higher post-test conceptual understanding scores ($M = 78.6$, $SD = 9.83$, 95% CI [76.73, 80.47]) than students in conventional instruction ($M = 63.4$, $SD = 10.12$, 95% CI [61.46, 65.34]), $t(212) = 10.83$, $p < .001$. The gain-score comparison showed an average gain of 26.3 points in the AI-scaffolding condition and 11.6 points in the conventional condition. The direction of the evidence supported H1, although the magnitude should be interpreted cautiously, as classroom-level assignment and novelty effects may have contributed to the observed difference.

Figure 2 illustrates the descriptive pre-test and post-test pattern for both instructional conditions. The visualization shows that the post-intervention separation was larger than the baseline difference.

Figure 2

Pre-test and post-test conceptual understanding scores by instructional condition



Note. Bars represent reported group means. Gain values were calculated from reported pre-test and post-test means; the figure is descriptive and does not replace inferential analyses.

The visual pattern reinforces the statistical comparison, but it should be read together with the design limitations of class-level assignment and the absence of delayed retention data.

Mediation by Metacognitive Awareness

The mediation model indicated that instructional condition was positively associated with metacognitive awareness ($\beta = 0.43$, $SE = 0.07$, $t = 6.14$, $p < .001$, 95% CI [0.29, 0.57]). Metacognitive awareness was positively associated with conceptual understanding while controlling for condition ($\beta = 0.38$, $SE = 0.06$, $t = 6.33$, $p < .001$, 95% CI [0.26, 0.50]). The indirect effect through metacognitive awareness was statistically significant (indirect effect = 0.19, boot SE = 0.04, 95% CI [0.11, 0.28]), supporting H2. The direct pathway from condition to conceptual understanding remained significant ($\beta = 0.51$, $SE = 0.08$, $t = 6.38$, $p < .001$, 95% CI [0.35, 0.67]), indicating partial rather than full mediation.

Moderated Mediation

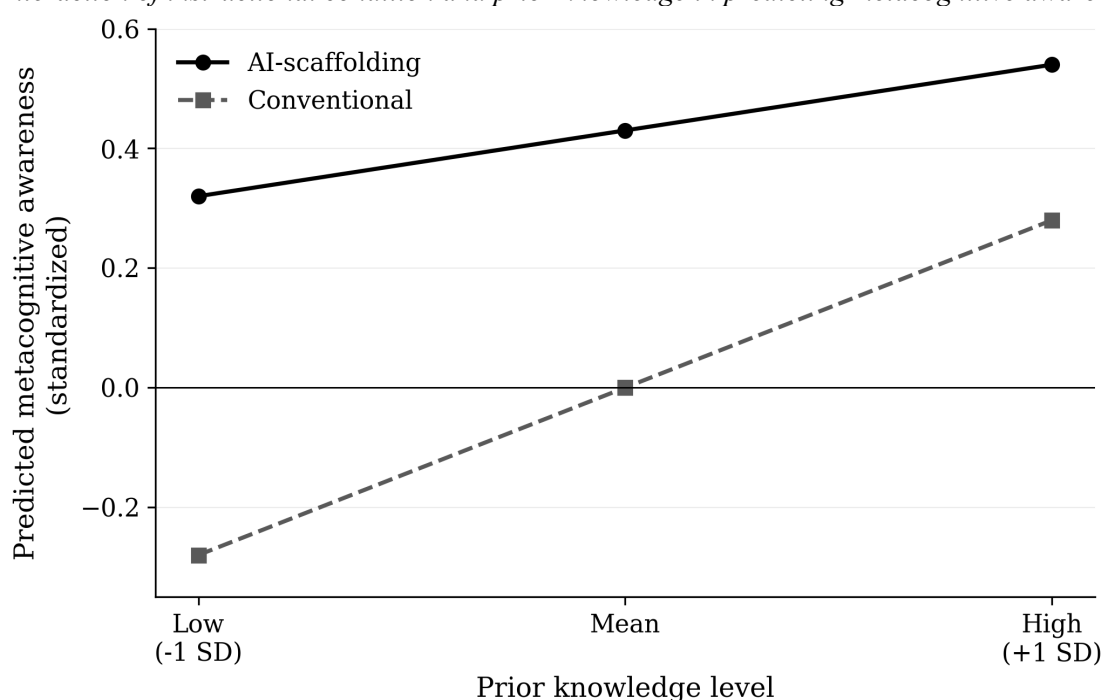
The interaction between instructional condition and prior knowledge was associated with metacognitive awareness ($\beta = 0.31$, $SE = 0.10$, $t = 3.10$, $p = .003$, 95% CI [0.11, 0.51]).

The index of moderated mediation was statistically significant (index = 0.12, boot SE = 0.04, 95% CI [0.05, 0.21]). Conditional indirect effects were larger among learners with lower prior knowledge (indirect effect = 0.26, 95% CI [0.14, 0.40]) than among learners with higher prior knowledge (indirect effect = 0.12, 95% CI [0.04, 0.22]). H3 was therefore supported, with the qualification that the model should be replicated using multilevel methods.

Figure 3 displays the conditional pattern underlying the moderated mediation result. The plotted values clarify that the AI-scaffolding condition was associated with higher metacognitive awareness across prior-knowledge levels, with the strongest relative advantage among students with lower prior knowledge.

Figure 3

Interaction of instructional condition and prior knowledge in predicting metacognitive awareness



Note. The figure displays the model-implied pattern for the moderated a path. Values are standardized estimates plotted at low (-1 SD), mean, and high (+1 SD) prior knowledge.

This pattern supports the interpretation that prior knowledge functioned as a boundary condition for the metacognitive pathway. It does not imply that low-prior-knowledge students automatically benefit from AI in all settings; rather, benefits appear tied to structured prompts and teacher-mediated use.

Discussion

This study examined whether AI-supported cognitive scaffolding was associated with conceptual understanding in secondary science learning, whether metacognitive awareness helped explain that association, and whether prior knowledge conditioned the metacognitive pathway. The results supported the hypothesized pattern: students in the AI-scaffolding condition showed larger conceptual understanding gains, metacognitive awareness partially mediated the association, and the indirect pathway was stronger among learners with lower prior knowledge. These findings are best understood as evidence of a conditional instructional

mechanism rather than evidence that AI is universally effective across learners, tasks, and contexts.

The main effect should be interpreted as an augmentation effect rather than as evidence that AI can replace teacher-guided science instruction. Recent theoretical and empirical syntheses argue that AI improves learning most defensibly when it adds well-designed cognitive support, increases opportunities for explanation and feedback, and preserves students' engagement with disciplinary reasoning (Bauer et al., 2025; Yan et al., 2024). In the present study, the AI condition included adaptive hints, immediate corrective feedback, and structured question branching; these features plausibly reduced unproductive impasses while maintaining opportunities for conceptual work. This interpretation aligns with recent evidence that AI-enabled personalized recommendations and AI-based classroom discourse tools can support engagement and outcomes when they are embedded within coherent instructional routines (Huang et al., 2023; Wang et al., 2024; Yim & Su, 2025).

The findings also connect with emerging STEM-focused research showing that generative AI can support conceptual clarification and misconception repair, but only within carefully designed learning sequences. For example, recent quasi-experimental work in STEM education suggests that ChatGPT-supported instruction may improve conceptual understanding while still leaving weaknesses in advanced quantitative reasoning and problem solving (El Fathi et al., 2025). This distinction is important for the present study: AI scaffolding appears useful for guiding conceptual progress, but the results should not be read as evidence that AI alone can develop complete scientific expertise. Conceptual understanding in science still requires teacher mediation, peer explanation, laboratory or representational activity, and opportunities to transfer ideas beyond the supported task.

The mediation result strengthens the argument that AI scaffolding may operate partly through students' regulatory engagement. Students did not merely receive feedback; they were prompted to assess their understanding, revise responses, and monitor progress. This interpretation is consistent with recent work showing that metacognitive support is a critical design feature in generative AI learning environments and that AI-supported self-regulated learning effects are positive but heterogeneous across instructional phases (Xu et al., 2025; Xu et al., 2026). Because the mediation was partial rather than full, other pathways remain plausible, including reduced extraneous cognitive load, increased time on task, improved motivation, or more frequent corrective feedback. Future studies should therefore avoid treating metacognition as the only explanatory mechanism and should test multiple mediators simultaneously.

The moderated mediation pattern provides a more differentiated interpretation of equity and learner variability. Students with lower prior knowledge appeared to benefit more from the metacognitive pathway, suggesting that adaptive scaffolding may help novices organize reasoning and identify misconceptions. However, this result should not be overstated as proof that AI automatically equalizes learning opportunities. AI may widen or reduce learning differences depending on access, digital infrastructure, teacher guidance, task design, language proficiency, and students' AI literacy (Markauskaite et al., 2022; Nguyen et al., 2023; Shoufan, 2023). The safer interpretation is that AI scaffolding can be especially promising for students with weaker initial conceptual foundations when it is integrated with explicit monitoring prompts and teacher-supported interpretation of feedback.

The magnitude of the observed learning difference also requires caution. The intervention lasted 12 weeks, used class-level assignment, and involved a locally developed conceptual understanding test. These features make the study more ecologically relevant than a brief laboratory trial, but they also introduce interpretive constraints. Classroom effects, teacher enthusiasm, novelty, and differences in implementation quality may have contributed

to the outcomes. Moreover, immediate post-test gains cannot by themselves demonstrate durable conceptual change. This is consistent with recent warnings that AI-education research should distinguish supported performance from retained learning and should avoid overgeneralization from immediate outcomes (Bauer et al., 2025; Deng et al., 2025; Yan et al., 2024).

Taken together, the findings position AI-supported scaffolding as a potentially valuable component of science instruction when it is embedded in a pedagogical sequence that requires explanation, monitoring, and revision. The results are less supportive of unstructured AI use, in which students may receive fluent responses without engaging in the conceptual work needed for durable understanding.

Implications

For science teachers, the findings support using AI-supported scaffolding as a guided learning aid rather than as an autonomous tutor. Teachers should require students to justify AI-assisted answers, compare AI feedback with scientific evidence, identify remaining uncertainties, and explain why a solution is conceptually valid. Such practices are especially important for secondary students who may not yet have sufficient disciplinary judgment or AI literacy to evaluate automated responses independently (Kasneci et al., 2023; Mertala et al., 2022; Nguyen et al., 2023).

For instructional designers and platform developers, the results support the inclusion of metacognitive checkpoints, explainable feedback, adaptive hint escalation, misconception alerts, and prompts that ask students to articulate reasoning rather than copy answers. AI-generated videos, avatars, and conversational agents may expand access to learning support, but they should be designed around cognitive engagement, transparency, and error checking rather than convenience alone (Aslan et al., 2024; Fink et al., 2024; Netland et al., 2025; Stadler et al., 2024).

For school leaders and policymakers, the findings suggest that AI implementation should be accompanied by teacher professional development, infrastructure readiness, data privacy safeguards, and clear norms for responsible student use. Investment in AI tools without parallel investment in pedagogy and governance risks producing shallow use, overreliance, or inequitable access. The results are therefore best understood as support for carefully mediated AI integration in science classrooms, not as a blanket justification for rapid or unsupervised adoption.

Limitations

The study was conducted in two public secondary schools in one Indonesian municipality. The findings may therefore transfer only to contexts with comparable curriculum, infrastructure, teacher preparation, language background, and student characteristics.

Class-level assignment reduced practical disruption but did not eliminate cluster-level confounding. Because only four intact classes were involved, the analysis could not fully separate student-level effects from classroom- or teacher-level variation. Future studies with more schools and classes should apply multilevel or cluster-randomized designs.

The conceptual understanding test and prior knowledge test were locally developed. Although content validity and reliability were acceptable in the present sample, independent validation is required before broader claims about measurement quality can be made. In addition, the metacognitive awareness measure relied on self-report, which may not fully capture dynamic monitoring and regulation during AI interaction.

The study also did not include delayed retention or transfer testing. Consequently, the findings speak primarily to immediate post-intervention conceptual outcomes rather than to long-term retention, flexible transfer, or independent problem solving without AI support.

Recommendations for Future Research

Future research should use multi-site cluster-randomized or multilevel designs involving more schools, teachers, and classroom clusters. Such designs would allow researchers to estimate student-level, classroom-level, and teacher-level sources of variation and would provide stronger evidence about the scalability of AI-supported cognitive scaffolding.

Future studies should combine self-report measures with platform log data, think-aloud protocols, eye-tracking, screen recordings, and classroom observations. These data sources would make it possible to test whether students actually engage in planning, monitoring, evaluation, and revision during AI-supported learning rather than merely reporting higher metacognitive awareness.

Future experiments should isolate the active ingredients of AI scaffolding by comparing adaptive hints, corrective feedback, metacognitive prompts, explainable feedback, and teacher-mediated reflection as separate design components. Component-level experiments would clarify whether learning gains are driven by adaptivity, feedback timing, prompt quality, teacher orchestration, or their interaction.

Future research should include delayed retention tests, transfer tasks, and assessments completed without AI support. These outcomes would help distinguish temporary performance support from durable conceptual understanding. Studies should also examine whether students can apply concepts in unfamiliar science contexts after the AI scaffold is removed.

Finally, qualitative and design-based studies are needed to examine how teachers and students negotiate trust, accuracy, dependence, and responsibility during AI-supported science learning. Such work would be particularly valuable in under-resourced or linguistically diverse settings, where infrastructure, language, access, and teacher readiness may strongly condition the effectiveness of AI-supported scaffolding. Future research should also examine inclusive adaptations for learners with disabilities, because the present intervention did not evaluate whether the platform and measures were accessible for students requiring individualized educational programming.

Conclusion

The study provides evidence that AI-supported cognitive scaffolding was associated with larger gains in conceptual understanding in secondary science learning and that metacognitive awareness partially explained this association. Prior knowledge moderated the mediation pathway, with stronger indirect associations among learners who began with lower prior knowledge. The contribution of the study lies in specifying a conditional mechanism rather than asserting a universal AI effect. AI-supported scaffolding appears most defensible when integrated with teacher oversight, metacognitive prompting, careful assessment, and transparent limitations. Replication across broader samples, longer time frames, inclusive learner groups, and independently validated measures is necessary before stronger claims can be made.

Author Contributions

Conceptualization, Anggi Widiarni and Yuli Yanti; Methodology, Anggi Widiarni; Formal analysis, Anggi Widiarni; Investigation, Anggi Widiarni and Yuli Yanti; Resources, Yuli Yanti; Data curation, Anggi Widiarni; Writing - original draft, Anggi Widiarni; Writing - review and editing, Yuli Yanti; Visualization, Anggi Widiarni; Supervision, Yuli Yanti; Project administration, Anggi Widiarni. Both authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

This study was conducted in accordance with ethical guidelines for educational research involving minors and was approved by the institutional review board of Universitas Islam Negeri Raden Intan Lampung.

Informed Consent Statement

Written informed consent was obtained from parents or legal guardians, and written assent was obtained from all participating students.

Trial Registration

Not applicable, because this study was an educational quasi-experimental study and did not involve clinical intervention.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request, subject to anonymization and institutional ethical restrictions.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Generative AI Statement

Generative AI tools were used only for language refinement and editorial checking. All scholarly content, interpretations, and final revisions were reviewed and approved by the authors.

References

- Aslan, S., Durham, L. M., Alyuz, N., Okur, E., Sharma, S., & Nachman, L. (2024). Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild. *Computers and Education: Artificial Intelligence*, 6, 100220. <https://doi.org/10.1016/j.caeai.2024.100220>
- Bai, B., Yeung, S. S., & Lo, C. K. (2026). Enhancing the effect of AI-assisted learning: The use of scaffolding strategies to develop students' prompt engineering skills. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2026.2649547>
- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37, 45. <https://doi.org/10.1007/s10648-025-10020-8>
- Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education*, 20, 62. <https://doi.org/10.1186/s41239-023-00432-3>
- Bhutoria, A. (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence*, 3, 100068. <https://doi.org/10.1016/j.caeai.2022.100068>
- Butler, D., & Starkey, L. (2024). OK Google, help me learn: Children's use of AI voice assistants for learning. *Technology, Pedagogy and Education*, 33(2), 135-148. <https://doi.org/10.1080/1475939X.2024.2311779>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahim, E. M. (2025). Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. *Disciplinary and Interdisciplinary Science Education Research*, 7, Article 6. <https://doi.org/10.1186/s43031-025-00125-z>
- Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19, 57. <https://doi.org/10.1186/s41239-022-00362-6>
- Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: Benefits and challenges. *Frontiers in Education*, 9, 1416307. <https://doi.org/10.3389/feduc.2024.1416307>
- Galindo-Dominguez, H., Delgado, N., Campo, L., & Losada, D. (2024). Relationship between teachers' digital competence and attitudes towards artificial intelligence in education. *International Journal of Educational Research*, 126, 102381. <https://doi.org/10.1016/j.ijer.2024.102381>
- Gao, X., Noroozi, O., Gulikers, J. T. M., Biemans, H. J. A., & Admiraal, W. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence*, 6, 100206. <https://doi.org/10.1016/j.caeai.2024.100206>
- Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, 13(7), 692. <https://doi.org/10.3390/educsci13070692>

- Huang, X., Lu, G., & Yang, X. (2023). Effects of AI-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 104684. <https://doi.org/10.1016/j.compedu.2022.104684>
- Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review*, 36, 25. <https://doi.org/10.1007/s10648-024-09868-z>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Shum, S. B., Gasevic, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, 3, 100056. <https://doi.org/10.1016/j.caeai.2022.100056>
- Mertala, P., Fagerlund, J., & Calderon, O. (2022). Finnish 5th and 6th grade students' pre-instructional conceptions of artificial intelligence and their implications for AI literacy education. *Computers and Education: Artificial Intelligence*, 3, 100095. <https://doi.org/10.1016/j.caeai.2022.100095>
- Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235-251. <https://doi.org/10.1080/21532974.2023.2247480>
- Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, 224, 105164. <https://doi.org/10.1016/j.compedu.2024.105164>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2023a). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, 28, 8445-8501. <https://doi.org/10.1007/s10639-022-11491-w>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2023b). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71, 137-161. <https://doi.org/10.1007/s11423-023-10203-6>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28, 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Ninaus, M., & Sailer, M. (2022). Closing the loop: The human role in artificial intelligence for education. *Frontiers in Psychology*, 13, 956798. <https://doi.org/10.3389/fpsyg.2022.956798>
- Samala, A. D., Rawas, S., Wang, T., Reed, J. M., Kim, J., Howard, N. J., & Ertz, M. (2025). Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects. *Education and Information Technologies*, 30(3), 3239-3278. <https://doi.org/10.1007/s10639-024-12936-0>
- Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access*, 11, 38805-38818. <https://doi.org/10.1109/ACCESS.2023.3268224>
- 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, 108386. <https://doi.org/10.1016/j.chb.2024.108386>
- Wang, H., Tao, Y., & Chen, C. (2024). Artificial intelligence in classroom discourse: A systematic review of the past decade. *International Journal of Educational Research*, 123, 102275. <https://doi.org/10.1016/j.ijer.2023.102275>

- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>
- Wu, R., & Yu, Z. (2024). Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10-33. <https://doi.org/10.1111/bjet.13334>
- Xu, Z., Qiao, H., Cheng, J., Liu, X., & Zhao, Y. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56(5), 1842-1863. <https://doi.org/10.1111/bjet.13599>
- Xu, J., Luo, Y., Wang, C., Wang, M., & Wu, Y. (2026). AI support in self-regulated learning: A decade of technological evolution and meta-analysis. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.70058>
- Yan, L., Greiff, S., Teuber, Z., & Gasevic, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8, 1839-1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yim, I. H. Y., & Su, J. (2025). Artificial intelligence learning tools in K-12 education: A scoping review. *Journal of Computers in Education*, 12, 93-131. <https://doi.org/10.1007/s40692-023-00304-9>