MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(8), em2886
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/19030

ChatGPT as cognitive scaffolding in 5E-based science instruction: Effects on middle school students' problem-solving competence

Nguyen Chi Le [1], Ha Thi Nguyen [1*], Quynh Thi Thuy Nguyen [1], Lan Phu Hoang Nguyen [1], The Van Tran [1], Tu Anh Tran [2], Quynh Viet Pham [3], Giang Thi Nguyen [1]

[1] University of Education – Vietnam National University, Hanoi, VIETNAM
[2] Nghe An University, Nghe An, VIETNAM
[3] Hanoi Metropolitan University, Hanoi, VIETNAM

Received 22 March 2026 ▪ Accepted 24 July 2026

Abstract

Problem-solving competence (PSC) is a central objective of contemporary science education; however, many students continue to experience difficulties when engaging in inquiry-based learning activities that require reasoning, reflection and solution refinement. This study investigated the effects of selective ChatGPT scaffolding embedded within the 5E instructional model—engage, explore, explain, elaborate, and evaluate—on middle school students' PSC. A quasi-experimental design was conducted with 80 grade 9 students, including an experimental group receiving ChatGPT-supported 5E instruction and a control group receiving conventional 5E instruction. Unlike continuous AI integration, ChatGPT support was selectively provided during the explore and elaborate phases to facilitate inquiry reasoning, solution generation and reflective evaluation. PSC was assessed across three dimensions: problem understanding (PU), solution planning and implementation (SPI), and evaluation and improvement (EI). Baseline equivalence between groups was confirmed using independent-samples t-tests. Analysis of covariance results revealed a significant positive effect of the intervention on overall PSC, $F(1, 77) = 6.080$, $p = .0159$, partial $\eta^2 = .073$. Significant improvement was also observed for EI, $F(1, 77) = 4.901$, $p = .0298$, partial $\eta^2 = .060$, while SPI showed a positive trend, $F(1, 77) = 2.953$, $p = .0898$. No significant effect was observed on PU. The findings suggest that selective ChatGPT scaffolding is particularly effective in supporting reflective reasoning, evidence evaluation, and solution refinement in inquiry-based science learning environments.

Keywords: ChatGPT, cognitive scaffolding, 5E instructional model, problem-solving, competence, science education

INTRODUCTION

Developing students' problem-solving competence (PSC) has become a major objective of contemporary science education, particularly within competence-based curricula that emphasize inquiry, reasoning, and the application of scientific knowledge to authentic situations (National Research Council [NRC], 2012; OECD, 2017). Rather than merely recalling scientific facts, students are increasingly expected to identify problems, evaluate evidence, justify solutions, and revise their thinking through reflective inquiry (Darling-Hammond et al., 2020; Jonassen, 2011). These competencies are closely aligned with international frameworks of scientific literacy and 21st century skills that emphasize higher-order thinking, evidence-based reasoning, and adaptive problem-solving.

Despite these expectations, many middle school students continue to experience difficulties in engaging in complex scientific problem-solving tasks. Previous research has shown that learners often struggle to transfer scientific knowledge to unfamiliar situations,

✉ Lcnguyendhhl@gmail.com ✉ nguyenha27579@vnu.edu.vn **(*Correspondence)** ✉ Quynhntt-bio@vnu.edu.vn
✉ lannph@vnu.edu.vn ✉ thetv@vnu.edu.vn ✉ tuta@nau.edu.vn ✉ Pvquynh@daihocthudo.edu.vn
✉ giangnguyen86@vnu.edu.vn

Contribution to the literature

- This study proposes a selective ChatGPT scaffolding model embedded within the 5E instructional framework for inquiry-based science learning.
- This study provides a pedagogical approach for strategically integrating ChatGPT into the explore and elaborate phases to support inquiry and reflective learning.
- The findings provide empirical evidence that selective ChatGPT scaffolding improves middle school students' PSC in 5E-based science instruction.

formulate coherent explanations, and evaluate alternative solutions in inquiry-oriented learning environments (Hmelo-Silver et al., 2007; Osborne, 2014; Windschitl et al., 2008). Such challenges are particularly evident when students are required to engage in open-ended inquiry activities that demand sustained reasoning, reflection and self-regulation (Kirschner et al., 2006).

To address these challenges, inquiry-based instructional models have been widely adopted in the field of science education. Among them, the 5E instructional model—engage, explore, explain, elaborate, and evaluate—has received substantial empirical support because it provides a coherent structure for scientific inquiry, conceptual change, and reflective learning (Bybee, 2014; Bybee et al., 2006). Through sequential inquiry experiences, students actively construct a scientific understanding rather than passively receiving information. However, inquiry-oriented learning alone does not guarantee improvements in PSC. Several studies have reported that students often require additional cognitive support to navigate complex inquiry tasks successfully (Belland, 2014; Hmelo-Silver et al., 2007).

From a socio-cultural perspective, cognitive scaffolding provides a mechanism for supporting learners during cognitively demanding tasks (Vygotsky, 1978; Wood et al., 1976). Scaffolding can guide students' reasoning, encourage reflection, and support metacognitive regulation during inquiry activities (Holton & Clarke, 2006; Kim & Hannafin, 2011). Nevertheless, providing individualized scaffolding remains challenging in typical classroom settings, particularly when teachers must simultaneously support multiple learners who are engaged in open-ended inquiry.

Recent advances in generative artificial intelligence (GenAI), particularly large language models such as ChatGPT, have created new opportunities for providing adaptive cognitive support in educational contexts (Holmes et al., 2024; Kasneci et al., 2023). Unlike traditional educational technologies, ChatGPT can generate context-sensitive prompts, explanations, and reflective questions in real time, potentially functioning as a form of AI-mediated cognitive support. Emerging studies suggest that GenAI can support reasoning, metacognitive reflection, and inquiry learning when it is integrated into pedagogically meaningful activities (Cotton et al., 2024; Tlili et al., 2025; Zawacki-Richter et al., 2024).

Evidence from Vietnam similarly indicates a rapid increase in the adoption of GenAI in educational settings. Recent studies have reported that students and teachers are increasingly using ChatGPT for learning support, academic inquiry, and instructional design, while emphasizing the need for theoretically grounded and pedagogically guided implementation strategies (Dang, 2024; Nguyen et al., 2025; Xuan et al., 2025). However, empirical investigations examining how ChatGPT can be systematically integrated into established inquiry-based instructional models remain scarce.

Although research on AI in education has expanded rapidly, three important gaps persist. First, many studies conceptualize ChatGPT primarily as a tutoring tool or information provider, rather than as a theory-driven cognitive scaffold. Second, limited attention has been devoted to understanding how GenAI should be embedded in specific phases of inquiry-oriented instructional models. Third, empirical evidence from middle school science education, particularly in developing educational contexts, remains scarce.

To address these gaps, this study proposes a model of selective ChatGPT scaffolding embedded within the 5E instructional framework. Rather than integrating ChatGPT into all instructional activities, AI support was selectively provided during the explore and elaborate phases, where students encountered the greatest demands for inquiry reasoning, reflection, and solution refinement. This approach conceptualizes ChatGPT not as an instructional replacement but as a targeted cognitive scaffold supporting inquiry-oriented learning processes.

Based on the proposed framework, this study tested the following hypotheses:

H1. Students receiving selective ChatGPT-supported 5E instruction will demonstrate higher overall PSC than students receiving conventional 5E instruction.

H2. Students receiving selective ChatGPT-supported 5E instruction will demonstrate stronger performance in solution planning and implementation (SPI) than in problem understanding (PU).

H3. The effects of selective ChatGPT-supported 5E instruction were relatively stronger for evaluation and improvement (EI) than for other dimensions of PSC.

LITERATURE REVIEW

PSC in Science Education

PSC is widely recognized as a core outcome of science education because it enables learners to apply scientific knowledge, reasoning, and evidence-based thinking to real-world situations (NRC, 2012; OECD, 2017). Contemporary frameworks conceptualize problem-solving as a multidimensional construct involving problem identification, solution generation, implementation, evaluation, and revision (Jonassen, 2011). In science learning, effective problem-solving requires students to understand scientific concepts and engage in inquiry, argumentation, and reflective reasoning (Osborne, 2014).

Previous studies have consistently reported positive relationships between inquiry-oriented instruction and the development of PSC (Hmelo-Silver et al., 2007; Windschitl et al., 2008). However, many students continue to experience difficulties when evaluating alternative solutions, justifying decisions, and revising explanations based on evidence. These challenges suggest the need for additional forms of cognitive support in inquiry-based learning environments.

Cognitive Scaffolding and AI-Supported Learning

The concept of cognitive scaffolding originates from sociocultural theories of learning, which emphasize the role of mediated support in helping learners perform tasks beyond their current level of competence (Vygotsky, 1978; Wood et al., 1976). In inquiry-based learning, scaffolding supports reasoning, reflection, self-regulation, and metacognitive monitoring during complex learning activities (Belland, 2014; Kim & Hannafin, 2011).

Recent advances in GenAI have expanded opportunities for providing adaptive scaffolding in educational settings. Studies have shown that ChatGPT can facilitate questioning, explanation generation, reflection, and feedback processes that support student learning (Holmes et al., 2024; Kasneci et al., 2023; Tlili et al., 2025). Similarly, recent studies in Vietnam have reported the increasing use of ChatGPT for learning support and instructional activities while emphasizing the importance of pedagogically guided implementation (Dang, 2024; Nguyen et al., 2025; Xuan et al., 2025).

Despite these promising findings, existing research has primarily examined ChatGPT as a general learning assistant and tutoring tool. Comparatively little attention has been devoted to understanding how AI-generated support can be strategically embedded within specific instructional phases to address specific cognitive demands.

ChatGPT and the 5E Instructional Model

The 5E instructional model remains one of the most widely adopted inquiry-based frameworks in science education because it systematically guides learners through the engagement, exploration, explanation, elaboration, and evaluation processes (Bybee, 2014; Bybee et al., 2006). Research has demonstrated that this model can promote conceptual understanding, inquiry skills, and scientific reasoning across different educational levels.

Recent AI-in-education studies have suggested that GenAI can complement inquiry-based learning by providing adaptive prompts, explanations, and reflective dialogue (Cotton et al., 2024; Zawacki-Richter et al., 2024). However, empirical evidence regarding the integration of ChatGPT into the 5E instructional cycle remains limited. Existing studies rarely distinguish between the cognitive demands associated with different phases of inquiry learning and, therefore, provide little guidance regarding when and how AI support should be employed.

Consequently, a significant research gap remains regarding the design of phase-specific AI scaffolding strategies for inquiry-based science instruction. To address this gap, the present study proposes a selective ChatGPT scaffolding model in which AI support is strategically embedded within the explore and elaborate phases of the 5E cycle. This approach seeks to align AI-mediated support with the phases that involve the greatest demand for inquiry reasoning, reflection, and evaluative thinking.

Assumption Checking

Prior to conducting the analysis of covariance (ANCOVA), the underlying statistical assumptions were examined. The distributions of the post-test outcome variables were inspected using skewness and kurtosis indices, which were all within the acceptable ranges for parametric analyses (**Table 1**). The homogeneity of variance between the experimental and control groups was evaluated using Levene's test, and no significant violations were detected (all $p > .05$). Given the relatively balanced sample sizes ($n = 40$ per group) and the robustness of ANCOVA to minor deviations from normality, the assumptions were considered sufficiently satisfied for the subsequent analyses.

The skewness and kurtosis values for all outcome variables were within the recommended ranges for approximate normality ($|\text{skewness}| < 2$; $|\text{kurtosis}| < 7$). Levene's tests were non-significant for all dependent variables ($p > .05$), indicating homogeneity of variances between the experimental and control groups. These

Table 1. Assessment of ANCOVA assumptions for the outcome variables

Outcome variable	Skewness	Kurtosis	Levene's F	p	Interpretation
Overall PSC	-0.298	2.263	0.806	.372	Assumptions satisfied
PU	-0.427	1.731	0.400	.529	Assumptions satisfied
SPI	-0.367	1.796	0.592	.444	Assumptions satisfied
EI	-0.423	2.024	0.178	.675	Assumptions satisfied

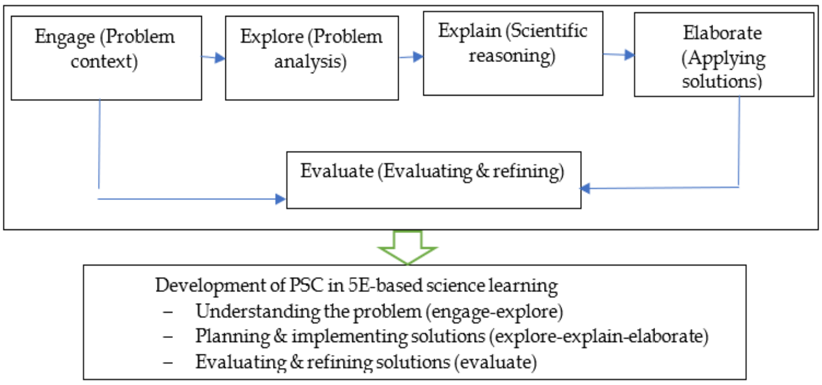

Figure 1. Development of PSC in 5E-based science learning (adapted from Bybee et al., 2006; OECD, 2017)

results support the application of ANCOVA in evaluating the intervention effects.

Following the recommendations of George and Mallery (2010), skewness values below |2| and kurtosis values below |7| were considered indicative of acceptable univariate normality for the parametric analyses.

THEORETICAL FRAMEWORK

Development of PSC Within the 5E Instructional Model

The 5E instructional model provides the pedagogical foundation for developing students' PSC in science learning. The model structures learning through five interconnected phases (5E), which collectively promote inquiry, conceptual understanding, knowledge application, and reflective reasoning (Bybee, 2014; Bybee et al., 2006).

The conceptualization of PSC adopted in this study is informed by the OECD/PISA framework, which defines problem-solving as an individual's capacity to explore and understand a problem situation, develop and implement solution strategies, and evaluate outcomes through reflective reasoning (OECD, 2017). Similarly, contemporary science education frameworks emphasize that effective problem-solving requires the integration of scientific knowledge, inquiry practices, evidence-based reasoning and reflective evaluation (Jonassen, 2011; NRC, 2012).

Based on these perspectives, PSC is conceptualized as a multidimensional construct comprising three interrelated components: PU, SPI, and EI. PU refers to the identification and analysis of problem situations, clarification of relevant information, and establishment of meaningful problem representations. SPI involve generating, selecting, and applying appropriate solution strategies. EI refers to evaluating evidence, reflecting on outcomes, and refining solutions based on feedback and reasoning.

Within the 5E instructional model, these competencies are progressively developed through inquiry-oriented learning. The engage and explore phases primarily support PU by encouraging students to identify problems, formulate questions, and investigate scientific phenomena. The explore, explain, and elaborate phases contribute to SPI through reasoning, knowledge construction, and application. Finally, the evaluate phase, together with reflective activities occurring throughout the inquiry cycle, supports EI by encouraging evidence-based judgment and solution refinement.

Figure 1 illustrates how the three dimensions of PSC are developed through inquiry processes embedded within the 5E instructional cycle. This framework provided the conceptual basis for evaluating students' PSC in the present study.

Selective ChatGPT Scaffolding Within the 5E Instructional Model

Cognitive scaffolding supports learners in performing tasks beyond their current level of competence by guiding their reasoning, reflection, and metacognitive regulation (Belland, 2014; Vygotsky, 1978; Wood et al., 1976). In inquiry-based science learning, scaffolding plays a particularly important role because students are often required to coordinate multiple cognitive processes simultaneously, including problem analysis, evidence evaluation, explanation construction, and reflective decision-making (Hmelo-Silver et al., 2007; Kim & Hannafin, 2011).

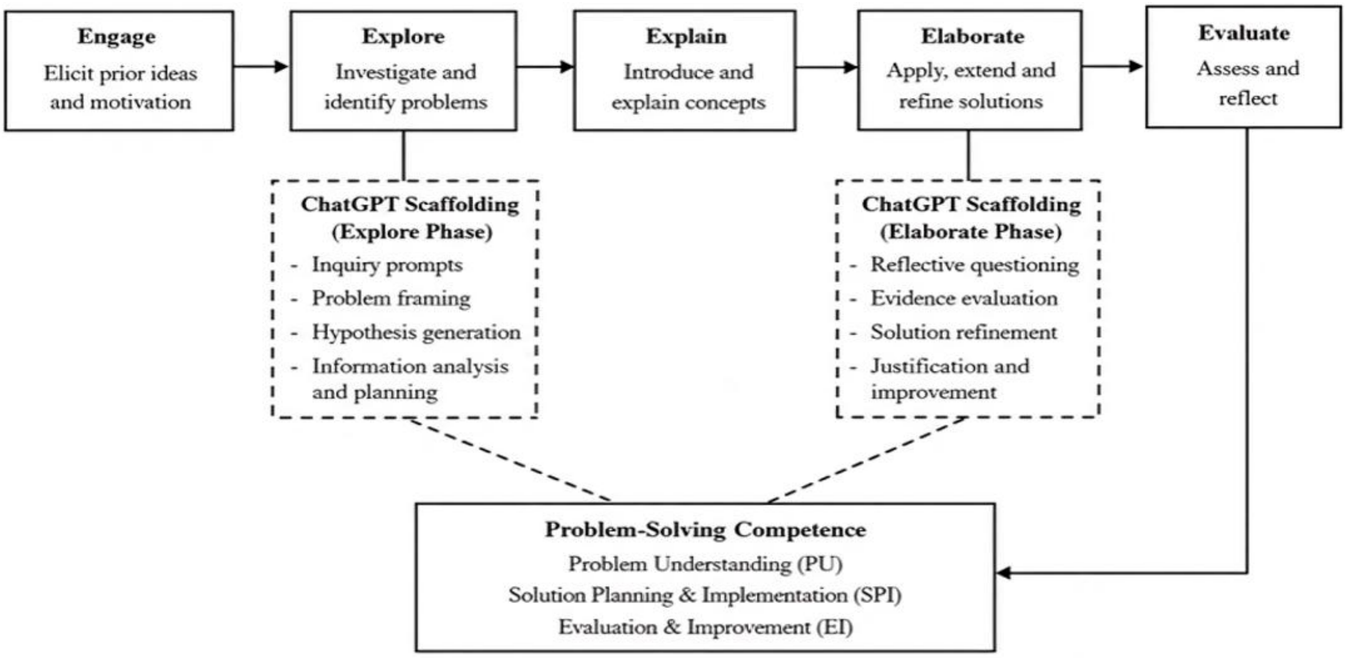

Figure 2. Selective ChatGPT scaffolding within the 5E instructional model (adapted from Bybee et al., 2006; Kasneci et al., 2023)

Recent advances in GenAI have expanded opportunities for providing adaptive cognitive support during inquiry learning. Studies suggest that ChatGPT can function as a form of AI-mediated cognitive scaffolding by generating inquiry prompts, supporting reasoning, facilitating reflection, and providing immediate feedback that promotes higher order thinking (Holmes et al., 2024; Kasneci et al., 2023; Tlili et al., 2025). Unlike traditional digital learning tools that primarily provide static information, GenAI engages learners in dynamic dialogues that adapt to emerging reasoning processes.

However, few studies have systematically examined how AI support can be aligned with the cognitive demands of inquiry-based science instruction.

To address this gap, the present study proposes a selective ChatGPT scaffolding model in which AI support is strategically embedded only within the explore and elaborate phases of the 5E instructional cycle. Rather than functioning as a continuous instructional assistant, ChatGPT can be conceptualized as a phase-specific cognitive scaffold designed to support students during the most cognitively demanding stages of inquiry learning.

The proposed framework assumes that different phases of the 5E cycle involve different cognitive demands and therefore require different types of support (see **Figure 2**). During the explore phase, students are expected to identify problems, formulate questions, generate hypotheses, analyze available information, and organize initial ideas. These processes involve inquiry, reasoning, cognitive activation, and problem representation. In this phase, ChatGPT functions as an exploratory scaffold by generating inquiry prompts, encouraging alternative perspectives, supporting scientific questioning, and helping students to clarify relationships among concepts. Such support is expected to strengthen PU and contribute to the SPI.

During the elaborate phase, students apply knowledge to new situations, compare alternative solutions, justify decisions, evaluate evidence, and refine their explanations. These activities depend heavily on reflective thinking, metacognitive monitoring and evaluative reasoning (Schraw et al., 2006; Zimmerman, 2002). Accordingly, ChatGPT functions as a reflective scaffold by engaging students in iterative dialogue, prompting justification, challenging assumptions, and encouraging the evidence-based revision of solutions. Consequently, AI support is expected to exert stronger effects on EI, which represents the highest level of PSC in this study.

In contrast, the engage, explain, and evaluate phases remain primarily teacher-mediated phases. This design preserves the instructional role of teachers in motivating inquiry, introducing scientific concepts, facilitating classroom interactions, and conducting formal assessments. Therefore, ChatGPT is conceptualized not as a replacement for teaching but as a complementary cognitive mediator that supports students' reasoning and reflection during specific inquiry phases of the learning process.

The central proposition of this framework is that the educational value of GenAI derives not from continuous exposure to AI but from the alignment between AI scaffolding functions and the cognitive demands of specific inquiry activities. By linking exploratory support to PU and reflective support to EI, the model provides a theoretically grounded explanation of how selective ChatGPT integration may enhance students' PSC in 5E-based science instruction. This mechanism forms the theoretical foundation for the hypotheses tested in this study.

METHODS

Research Design and Participants

This study employed a quasi-experimental pre-/post-test control group design to examine the effects of selective ChatGPT scaffolding on 5E-based science instruction. The participants were 80 grade 9 students from a lower secondary school in Vietnam. Two intact classes were assigned to either the experimental group (n = 40) or the control group (n = 40). Both groups studied the same science content, followed the same curriculum, and were taught by the same teacher during the intervention period.

Both groups received instructions based on the 5E instructional model. The key difference was the integration of ChatGPT.

In the control group, all learning activities were conducted using conventional teacher-guided 5E instruction. In the experimental group, ChatGPT was selectively integrated during the explore and elaborate phases. During the explore phase, students used ChatGPT to support inquiry, questioning, problem analysis, information organization, and hypothesis generation. During the elaborate phase, ChatGPT was used to facilitate solution refinement, evidence evaluation, justification of decisions and reflective improvement of proposed solutions.

The engage, explain, and evaluate phases remained teacher-mediated in both the groups. This design ensured that ChatGPT functioned as a targeted cognitive scaffold, rather than as a substitute for classroom instruction.

Measures

PSC was assessed using a rubric developed from the OECD (2017) problem-solving framework and the contemporary science education literature. The rubric comprises three dimensions:

- PU: Identifying and analyzing problem situations, clarifying relevant information, and constructing meaningful problem representations.
- SPI: Generating, selecting, and applying solution strategies.
- EI: Evaluating evidence, reflecting on outcomes, and refining solutions.

The dimension scores were combined to produce an overall PSC score. Students completed equivalent assessment tasks before and after the intervention. Independent scoring procedures were used to ensure the consistency of the evaluation.

Table 2. Baseline equivalence of the experimental and control groups

Variable	t	p
PU_Pre	1.240	0.219
SPI_Pre	1.227	0.224
EI_Pre	0.000	1.000
PSC_Pre	1.330	0.187

Data Analysis

Data were analyzed using the R statistical software. First, independent-samples t-tests were conducted to examine baseline equivalence between the experimental and control groups for all pre-test measures. Subsequently, a series of ANCOVA models were performed with post-test scores as dependent variables, instructional condition as the independent variable, and corresponding pre-test scores as covariates.

Separate ANCOVA models were conducted for overall PSC and for each competence dimension (PU, SPI, and EI). Effect sizes were reported using partial eta-squared (partial η^2). To facilitate the interpretation of the intervention effects, adjusted post-test means and 95% confidence intervals were estimated using estimated marginal means.

The analyses were designed to test three hypotheses: (**H1**) selective ChatGPT-supported 5E instruction would improve overall PSC; (**H2**) stronger effects would be observed for SPI than for PU; and (**H3**) the strongest effects would occur for EI.

RESULTS

Before examining the intervention effects, baseline equivalence between the experimental and control groups was assessed using independent-samples t-tests on the pre-test measures of PU, SPI, EI, and overall PSC.

As shown in Table 2, no statistically significant differences were observed between the two groups on any pre-test measure (PU: $t = 1.240$, $p = 0.219$; SPI: $t = 1.227$, $p = 0.224$; EI: $t = 0.000$, $p = 1.000$; and PSC: $t = 1.330$, $p = 0.187$). These findings indicate that the experimental and control groups were comparable before the intervention, supporting the validity of subsequent comparisons. This additional analysis directly addresses the concerns regarding potential baseline inequivalence raised during the review process.

H1-H3. To examine the effects of the intervention, a series of ANCOVA were conducted with post-test scores as dependent variables, instructional condition as the independent variable, and corresponding pre-test scores as covariates.

The results in Table 3 revealed a statistically significant group effect for overall PSC, $F(1, 77) = 6.080$, $p = 0.0159$, partial $\eta^2 = 0.073$. Students in the ChatGPT-supported 5E condition achieved significantly higher

Table 3. ANCOVA results for testing **H1-H3**

Outcome	F (1, 77)	p	Partial η^2	Decision
PSC	6.080	0.0159	0.073	Supported
PU	0.621	0.4330	0.008	Not supported
SPI	2.953	0.0898	0.037	Marginal trend
EI	4.901	0.0298	0.060	Supported

Table 4. Estimated marginal means

Outcome	Group	Adjusted mean	Standard error
PSC	ChatGPT-5E	7.05	0.22
	Control	6.31	0.22

adjusted post-test scores than those in the conventional 5E condition. Therefore, **H1** is supported.

Different patterns emerged at the component level. For PU, the group effect was not statistically significant, F (1, 77) = 0.621, p = 0.4330, partial η^2 = 0.008. In contrast, SPI showed a positive trend favoring the experimental group, F (1, 77) = 2.953, p = 0.0898, partial η^2 = 0.037, although this effect did not reach the conventional .05 significance threshold.

Consequently, **H2** received only partial support, indicating that selective ChatGPT scaffolding contributed more strongly to solution generation processes than to initial PU.

For EI, the ANCOVA indicated a statistically significant group effect, F (1, 77) = 4.901, p = 0.0298, partial η^2 = 0.060. Students receiving selective ChatGPT support demonstrated stronger performance in reflective evaluation and solution refinement than those in the control condition.

These findings support **H3**, suggesting that AI-mediated scaffolding exerted the strongest influence on the evaluative and reflective dimensions of problem-solving.

Importantly, the results indicate that the intervention did not affect all dimensions equally. Rather than producing broad improvements across all aspects of problem-solving, the effects were concentrated on higher-order evaluative processes, which are theoretically consistent with the reflective scaffolding functions assigned to ChatGPT during the elaborate phase of the instructional model.

To further examine **H3**, effect sizes were compared across the three dimensions of PSC.

As shown in **Table 4**, EI produced the largest effect size (partial η^2 = .060), followed by SPI (partial η^2 = 0.037), whereas PU showed only a negligible effect (partial η^2 = .008). This ranking indicates that selective ChatGPT scaffolding was most strongly associated with students' capacity to evaluate evidence, justify decisions and improve solutions through reflective reasoning.

These findings align with the theoretical framework presented in **Figure 2**, which positions ChatGPT primarily as a reflective scaffold during the elaborate phase rather than as a tool for direct content delivery to students. The results suggest that AI-supported dialogue may be particularly effective for supporting metacognitive reflection and evaluative thinking while exerting more limited influence on initial problem identification and representation processes.

Table 5. Adjusted means (ANCOVA estimated marginal means)

Outcome	Group	Adjusted mean	Standard error	95% CI
PSC	ChatGPT-5E	7.05	0.22	6.61-7.49
	Control	6.31	0.22	5.88-6.75
PU	ChatGPT-5E	2.28	0.12	2.03-2.53
	Control	2.16	0.12	1.91-2.40
SPI	ChatGPT-5E	2.35	0.12	2.11-2.58
	Control	2.07	0.12	1.83-2.30
EI	ChatGPT-5E	2.43	0.11	2.21-2.64
	Control	2.09	0.11	1.87-2.30

Note. CI: Confidence interval

Adjusted Post-Test Means

To facilitate the interpretation of the intervention effects, adjusted post-test means estimated from the ANCOVA models were calculated.

The adjusted means consistently favored the ChatGPT-supported 5E group across all the outcome variables (**Table 5**). For overall PSC, the experimental group achieved an adjusted mean of 7.05 compared with 6.31 for the control group. Similar patterns were observed for PU (2.28 vs. 2.16), SPI (2.35 vs. 2.07), and EI (2.43 vs. 2.09) in the two groups. The largest adjusted difference was observed for EI, further supporting the conclusion that selective ChatGPT scaffolding contributed most strongly to the reflective and evaluative aspects of scientific problem-solving.

DISCUSSION

This study investigated whether selective ChatGPT scaffolding embedded within a 5E instructional framework could support the development of middle school students' PSC in science learning. Rather than integrating AI continuously throughout the instruction, ChatGPT was strategically employed during the explore and elaborate phases, where students encountered the greatest demands for inquiry reasoning, solution generation, and reflective evaluation. The findings provide evidence that the effectiveness of AI-supported learning may depend not only on access to GenAI but also on how AI functions align with the cognitive demands of specific instructional activities.

The positive effect observed for overall PSC is consistent with emerging research suggesting that GenAI can support higher-order thinking when it is embedded in pedagogically structured learning environments. Recent studies have argued that ChatGPT is most effective when functioning as a cognitive scaffold that guides inquiry, reflection, and reasoning, rather than as a simple provider of information (Holmes et al., 2024; Kasneci et al., 2023). Similarly, Tlili et al. (2025) emphasized that the educational value of GenAI depends largely on instructional design and the extent to which AI interactions promote active cognitive engagement in learning activities. The present findings extend this emerging body of evidence by demonstrating that selective AI scaffolding can be integrated into an established inquiry-based framework, such as the 5E instructional model.

An important contribution of this study concerns the differential effects observed across the three dimensions of PSC. The strongest effects were associated with EI, whereas considerably weaker effects were found for PU. This pattern closely aligns with the theoretical assumptions underlying the proposed framework. During the elaborate phase, students used ChatGPT to compare alternative solutions, justify their decisions, evaluate evidence, and revise their responses. These activities require reflective judgment and metacognitive monitoring, processes that GenAI appears particularly well suited to support through iterative dialogue and feedback generation. Similar conclusions have been reported in recent studies, indicating that language models may be especially valuable for facilitating reflection, argumentation, and evidence-based reasoning (Cotton et al., 2024; Zawacki-Richter et al., 2024).

In contrast, the relatively limited effects observed for PU suggest that AI scaffolding may be less influential during the initial stages of inquiry, where learners attempt to identify and represent problem situations. Problem framing often depends on prior knowledge, contextual understanding, and direct engagement with learning tasks, factors that may not be substantially altered by AI-mediated support alone. This finding is consistent with previous research showing that inquiry learning requires substantial teacher guidance during the early phases of problem identification and conceptual exploration (Belland, 2014; Hmelo-Silver et al., 2007). Consequently, GenAI should not be viewed as a replacement for teacher facilitation, but rather as a complementary tool that supports specific cognitive processes within inquiry-oriented instruction.

The observed trend for SPI provides additional insight into how AI scaffolding may influence scientific problem-solving. Although the effect was smaller than that observed for EI, the direction of the findings suggests that AI-supported questioning and idea generation may contribute to students' ability to formulate and apply solution strategies. This interpretation aligns with recent studies reporting that ChatGPT can stimulate brainstorming, support alternative solution generation, and encourage deeper exploration of scientific ideas (Holmes et al., 2024; Tlili et al., 2025). Nevertheless, the modest magnitude of this effect indicates that solution planning remains strongly dependent on learners' existing conceptual knowledge and their inquiry skills.

Collectively, these findings support the growing consensus that the educational potential of GenAI lies not in automating learning processes but in augmenting students' reasoning and reflection within carefully designed pedagogical environments. The results suggest that selective integration may be more beneficial than unrestricted AI use because it preserves teachers' instructional role while providing targeted support during cognitively demanding phases of inquiry. This perspective is consistent with recent calls for human-centered approaches to AI integration that emphasize pedagogical alignment, learner agency, and reflective engagement, rather than technological substitution (Holmes et al., 2024; Zawacki-Richter et al., 2024).

This study has several limitations. First, the study involved a relatively small sample drawn from two grade 9 classes within a single educational context, which may limit the generalizability of its findings. Second, although significant effects were observed for overall competence and evaluative reasoning, some dimensions demonstrated only modest improvements, suggesting that larger studies are needed to confirm the robustness of the proposed model. Third, this study focused primarily on quantitative outcomes and did not examine the quality of student-AI interactions. Future research should incorporate qualitative analyses of dialogue patterns, reasoning processes, and metacognitive development to better understand how GenAI supports inquiry learning. Future studies should investigate how different prompting strategies, levels of teacher mediation, and subject domains influence the effectiveness of AI-supported scaffolding.

CONCLUSION

This study examined the effects of selective ChatGPT scaffolding embedded within the 5E instructional model on middle school students' PSC in science. The findings indicate that students receiving ChatGPT-supported 5E instruction achieved higher overall PSC than those receiving conventional 5E instruction did. The strongest effects were observed for EI, followed by SPI, whereas PU showed limited improvement.

This study contributes to the growing literature on AI-supported learning by proposing a theoretically grounded model of selective ChatGPT scaffolding. Rather than functioning as a general instructional assistant, ChatGPT was conceptualized as a phase-specific cognitive scaffold aligned with the cognitive demands of the explore and elaborate phases of the 5E instructional model. The findings suggest that the educational value of GenAI depends on pedagogically purposeful integration rather than continuous use.

From a practical perspective, selective AI integration offers a feasible approach to support inquiry-based science learning while preserving the central instructional role of teachers. In particular, GenAI appears to be especially valuable for promoting reflective reasoning, evidence evaluation, and solution refinement.

Author contributions: NCL: conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, writing – original draft; **HTN:** data curation, investigation, validation, writing – review & editing; **QTTN & GTN:** data curation, validation, visualization, writing – review & editing; **LPHN:** data curation, validation, writing – review & editing; **TVT & TAT:** investigation, resources, writing – review & editing; **QVP:** data curation, investigation, resources, writing – review & editing. All authors agreed with the results and conclusions.

Funding: Part of this research was funded by the VNU – University of Education under Research Project No. QS.24.05 (LPHN).

Ethical statement: This study was conducted in accordance with ethical principles governing educational research involving human participants. Written informed consent was obtained from all participating students and their parents or legal guardians before data collection. Participation was voluntary, all data were anonymized before analysis, and confidentiality was maintained throughout the study. The research complied with the institutional regulations of the participating school and the ethical principles of the Declaration of Helsinki applicable to educational research.

AI statement: Selected sections were translated from Vietnamese to English using GPT-5 (OpenAI's ChatGPT) and lightly polished for language. All outputs were reviewed, revised, and fact-checked by the study authors. No AI system influenced the study design, analysis, or conclusion.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 505-518). Springer. https://doi.org/10.1007/978-1-4614-3185-5_39

Bybee, R. W. (2014). The BSCS 5E instructional model: Personal reflections and contemporary implications. *Science and Children, 51*(8), 10-13. https://doi.org/10.2505/4/sc14_051_08_10

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. BSCS.

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

Dang, V. E. (2024). Current status of ChatGPT use in students' learning and research activities. *Vietnam Journal of Education*.

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97-140. https://doi.org/10.1080/10888691.2018.1537791

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 18.0 update* (11th ed.). Allyn & Bacon.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107. https://doi.org/10.1080/00461520701263368

Holmes, W., Miao, F., Viéville, T., & Gaffney, C. (2024). Generative artificial intelligence and the future of education. *UNESCO*. https://unesdoc.unesco.org/ark:/48223/pf0000385877

Holton, D., & Clarke, D. (2006). Scaffolding and metacognition. *International Journal of Mathematical Education in Science and Technology, 37*(2), 127-143. https://doi.org/10.1080/00207390500285818

Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? Opportunities and challenges of large language models in education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research, theory, and practice. *Computers & Education, 56*(2), 403-417. https://doi.org/10.1016/j.compedu.2010.08.024

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86. https://doi.org/10.1207/s15326985ep4102_1

Nguyen, N. Q. T., Tran, N. M., Vu, H. T., Tran, L. T., & Nguyen, H. V. T., Bui, T. T., & Nguyen, T. T. (2025). Evaluating the feasibility of using generative AI for educational research within the context of Vietnam's 2018 general education curriculum. *Vietnam Journal of Education, 9*(2), 253-264. https://doi.org/10.52296/vje.2025.481

NRC. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

OECD. (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy, and collaborative problem solving*. OECD Publications. https://doi.org/10.1787/9789264281820-en

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25*(2), 177-196. https://doi.org/10.1007/s10972-014-9384-1

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139. https://doi.org/10.1007/s11165-005-3917-8

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2025). Generative AI and educational transformation: Emerging trends and future directions. *Smart Learning Environment*.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm for school science investigations. *Science Education, 92*(5), 941-967. https://doi.org/10.1002/sce.20259

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

Xuan, S. H., Nguyen, A. T., Nguyen, T., Nguyen, L., Nguyen, H., Pham, N., Phung, T., Ngo, B., Nguyen, V., Nguyen, M., Tran, T., Le, T., Nguyen, K., & FNU, P. (2025). Evaluating the impact of generative AI on mathematics education: A comparative study in Vietnamese high schools. *Human Behavior and Emerging Technologies*. https://doi.org/10.1155/hbe2/8886206

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2024). Systematic review of artificial intelligence applications in higher education and implications for future research. *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64-70. https://doi.org/10.1207/S15430421TIP4102_2