JEELS
(Journal of English Education and Linguistics Studies)
P-ISSN: 2407-2575 E-ISSN: 2503-2194
https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/jeels

AI-ENHANCED PROBLEM-BASED LEARNING FRAMEWORK: INTEGRATING CHATGPT AS ADAPTIVE SCAFFOLDING TO IMPROVE CRITICAL THINKING AND PERSONALIZED LEARNING

***La Sunra[1]; Suci Amaliah[2]; Fitri Radhiyani[3]**
[1,2,3]English Language Education, Faculty of Languages and Literature, Universitas Negeri Makasar, Indonesia
**la.sunra@unm.ac.id; suci.amaliah@unm.ac.id; fitri.radhiyani@unm.ac.id*
(*) Corresponding Author

Abstract: As educational institutions increasingly adopt artificial intelligence technologies, understanding how generative AI can be integrated into student-centered pedagogical frameworks has become an important area of contemporary educational research. This study examines an AI-Enhanced Problem-Based Learning (AI-PBL) Framework that integrates ChatGPT as adaptive scaffolding to improve critical thinking and support personalized learning. Using an explanatory sequential mixed-methods design, the study involved 120 eighth-grade students from four junior high schools in Ujung Pandang District, Makassar City, Indonesia. Quantitative

[1]**Citation in APA style:**
Sunra, L., Amaliah, S., & Radhiyani, F. (2026). AI-enhanced problem-based learning framework: Integrating ChatGPT as adaptive scaffolding to improve critical thinking and personalized learning. *JEELS (Journal of English Education and Linguistics Studies), 13*(2), 577-605.
DOI: 10.30762/jeels.v13i2.7039

Submission: November 2025, Revision: May 2026, Publication: June 2026

data were collected through pre- and post-tests of critical thinking, while qualitative data came from classroom observations, ChatGPT interaction logs, and interviews. The overall critical thinking score increased from 60.3 (SD = 9.7) to 73.5 (SD = 10.4), $t(119) = 9.64$, $p < .001$, $d = 1.00$, with notable gains in analysis, inference, and bias detection. Qualitative analysis showed four patterns: ChatGPT supported idea exploration, strengthened verification habits, enhanced reflection and metacognitive monitoring, and did not lead to passive dependence. These findings indicate that generative AI can function as a learning partner when it is embedded in structured PBL and guided by teachers. The study offers practical evidence for integrating ChatGPT in secondary classrooms without reducing student agency.

Keywords: *problem-based learning, ChatGPT, adaptive scaffolding, critical thinking, personalized learning*

INTRODUCTION

Critical thinking remains one of the core competencies expected in contemporary learning because students are required to evaluate information, examine evidence, compare viewpoints, and make reasoned judgments rather than receive information passively (Memarian et al., 2023). This demand has become more visible in Indonesian secondary education, where students increasingly encounter digital information, algorithmic recommendations, and AI-generated answers in everyday learning. In such conditions, classrooms need pedagogical designs that expose students not only to abundant information but also to disciplined ways of questioning, checking, and justifying what they read, hear, and use.

The emergence of generative artificial intelligence has made this issue more urgent. ChatGPT and similar large language models can produce fluent explanations, examples, summaries, and arguments within seconds. For students, this immediacy may be helpful when they face complex tasks, yet it may also create a tendency to accept

ready-made answers without sufficient scrutiny. The educational challenge, therefore, is not simply whether students should be allowed to use AI, but how teachers can design learning environments in which AI-supported inquiry strengthens rather than weakens learners' reasoning processes.

Within this context, Problem-Based Learning (PBL) is widely regarded as a suitable approach for cultivating critical thinking because it engages learners in authentic problems, collaborative inquiry, and evidence-based reasoning (Barrows, 1996; Savery, 2006). Rather than positioning students as passive recipients of answers, PBL requires them to identify problems, analyze available information, propose solutions, and reflect on the quality of their decisions. The core logic of PBL is therefore compatible with the goal of developing students who are able to reason through uncertainty and justify their conclusions.

PBL is also relevant for secondary classrooms because it helps connect school knowledge with real-life problems that students can recognize. In conventional instruction, critical thinking is sometimes treated as an abstract skill detached from classroom tasks. In PBL, however, students are asked to think critically because the problem itself demands interpretation, evidence selection, negotiation, and solution design. This makes PBL an appropriate pedagogical bridge between the need for stronger critical thinking and the realities of contemporary classroom learning.

At the same time, generative AI tools such as ChatGPT have introduced both opportunities and risks in education. Recent studies suggest that ChatGPT can support feedback, idea generation, higher-order thinking, and personalized learning when its use is guided by clear learning goals (Essel et al., 2024; Wang & Fan, 2025). AI-generated explanations can help students compare alternatives, identify weaknesses in their reasoning, and receive immediate prompts when teachers cannot respond individually to every group. These possibilities are especially important in classrooms where students differ in confidence, background knowledge, and pace of learning.

Nevertheless, the same technology may produce harmful learning patterns when it is used without structure. Several studies have raised concerns that unguided AI use may encourage overreliance, cognitive offloading, reduced reading effort, and weaker independent judgment (Abbas et al., 2024; Costa et al., 2025). Students may copy AI responses, treat fluent answers as accurate answers, or bypass the difficult process of comparing evidence. These contrasting findings indicate that the educational value of ChatGPT depends heavily on the pedagogical design in which it is embedded.

Previous research on AI in education has tended to focus on higher education contexts, particularly university students' perceptions, writing support, academic integrity concerns, and AI-assisted feedback practices (Baig et al., 2024; Kasneci et al., 2023; Munaye et al., 2025). Although these studies provide important insights, they do not fully explain how younger learners in secondary classrooms interact with ChatGPT during structured inquiry. Secondary students may need more explicit guidance because their information literacy, metacognitive awareness, and self-regulation are still developing.

Evidence from secondary settings is particularly important because the risks and benefits of AI may appear differently in younger classrooms. Students at this level may be curious and motivated by interactive tools, but they may also be more vulnerable to accepting AI-generated statements without checking their accuracy. For that reason, research should examine not only learning outcomes but also the classroom processes through which AI is used, monitored, discussed, and evaluated.

Another limitation of existing research is that many studies report either quantitative learning outcomes or descriptive accounts of AI use, but fewer studies combine both forms of evidence. Outcome data are necessary to show whether students' critical thinking improves, yet they cannot fully explain how students use ChatGPT during learning. Conversely, qualitative descriptions can illuminate interaction patterns, but they need to be connected to measurable

learning change. A mixed-methods approach is therefore needed to understand both the extent of improvement and the process behind that improvement.

In response to these gaps, this study examines an AI-Enhanced Problem-Based Learning (AI-PBL) Framework that integrates ChatGPT as adaptive scaffolding within structured PBL activities. The framework does not treat ChatGPT as a replacement for teacher guidance or student reasoning. Instead, it positions ChatGPT as a source of prompts, alternative perspectives, feedback, and reflective questions that must be evaluated through peer discussion, teacher facilitation, and multi-source verification.

The present study is therefore situated at the intersection of three concerns: the need to strengthen critical thinking in secondary education, the pedagogical potential of PBL, and the rapid entry of generative AI into students' learning practices. This intersection is significant for English education and broader language-based learning because students often use language to interpret sources, formulate questions, negotiate meaning, and present arguments. When ChatGPT is incorporated into these processes, students need explicit guidance to distinguish between generating text and constructing knowledge. The AI-PBL framework responds to this need by making inquiry, verification, and reflection visible within classroom practice.

This study also responds to a practical concern faced by teachers who are expected to prepare students for digital learning while still protecting academic integrity and independent thinking. In many classrooms, students already have access to AI tools outside formal instruction, but their use is rarely accompanied by explicit routines for checking accuracy, comparing evidence, or revising claims. A prohibition-oriented approach may reduce visible misuse in the classroom, yet it does not necessarily prepare students to deal with AI-generated information in authentic learning situations. For this reason, a guided pedagogical model is needed so that students can experience AI as an object of critical evaluation, not merely as a convenient source of answers.

The novelty of the present study lies in its attempt to connect ChatGPT-supported scaffolding with the established inquiry structure of PBL in lower-secondary education. Instead of evaluating ChatGPT as a stand-alone tool, the study examines how it operates when placed within a sequence of problem framing, exploration, verification, solution development, and reflection. This distinction is important because the same technology may produce different learning effects depending on whether it is used to replace thinking or to stimulate thinking. By situating ChatGPT within teacher-guided PBL, the study offers a more educationally accountable way to examine generative AI in classroom practice.

Specifically, this study asks two research questions: RQ1, to what extent does AI-PBL improve students' critical thinking in a middle school context? and RQ2, how do learners interact with ChatGPT in structured PBL activities? By answering these questions, the study offers practical evidence for integrating generative AI into secondary education while preserving critical engagement, verification, collaboration, and learner agency.

METHOD

Research Design

This study used an explanatory sequential mixed-methods design. The quantitative phase employed a one-group pre-test–post-test procedure to examine whether students' critical thinking changed after the eight-week AI-PBL intervention, and the qualitative phase was then used to explain how that change occurred through classroom observations, ChatGPT interaction logs, and interviews (Creswell & Creswell, 2023).

This design was chosen because the study addressed two complementary questions: the extent of change in students' critical thinking and the process through which students used ChatGPT during structured inquiry. It was considered more appropriate than classroom action research because the study aimed to evaluate one common intervention across four schools and to interpret shared

outcome patterns across settings, rather than to conduct repeated cycles of classroom repair within a single class.

The explanatory sequential logic was operationalized in three connected steps. First, the quantitative phase measured the change in students' critical thinking after the intervention. Second, the researchers reviewed the score patterns to identify the aspects that improved most clearly and to determine which participants should be selected for follow-up interviews. Third, qualitative evidence from observations, interaction logs, and interviews was used to interpret the classroom mechanisms behind the numerical results. This sequencing helped prevent the qualitative data from becoming a separate description; instead, it served to explain why and how the measured changes occurred during AI-PBL implementation.

Participants and Setting

The study involved 120 eighth-grade students (approximately 13–14 years old) from four public junior high schools in Ujung Pandang District, Makassar City, South Sulawesi, Indonesia. The schools were selected purposively because they represented comparable public lower-secondary settings in the district, implemented the same national curriculum, had school approval and adequate device-and-internet access for ChatGPT-supported learning, and were willing to carry out the full eight-week procedure. One Grade 8 class from each school participated. Four subject teachers served as facilitators after a short orientation on the AI-PBL procedure, verification routines, and their role as inquiry guides. Informed consent was obtained from schools, teachers, students, and parents or guardians.

The participating classes were not selected to represent all Indonesian junior high schools, but to provide comparable implementation settings in which the intervention could be conducted consistently. This decision was important because AI-supported learning requires not only student access to devices, but also school readiness, teacher cooperation, and classroom routines that allow group-based inquiry. The schools shared similar curricular

expectations and lesson schedules, which reduced procedural differences across sites. At the same time, using four schools strengthened the study because the intervention was not limited to a single classroom or one teacher's practice.

Table 1.
Participant and Setting Profile

Aspect	Profile
Students	120 Grade 8 students (approximately 13–14 years old)
Schools	4 public junior high schools in Ujung Pandang District, Makassar City
Facilitators	4 subject teachers (one per class) serving as inquiry facilitators
School selection basis	Comparable public lower-secondary settings, same national curriculum, school approval, and adequate device/internet access
Student inclusion basis	Enrolled in participating Grade 8 classes and present during the intervention schedule
Intervention context	Eight-week AI-PBL implementation in science and social studies lessons

AI-PBL Intervention Procedure

The intervention was organized through the AI-Enhanced Problem-Based Learning (AI-PBL) Framework, which integrated ChatGPT into the PBL cycle as adaptive scaffolding in science and social studies lessons. The framework was used as the operational procedure of the intervention, not as a separate theoretical discussion, because it specified how students were guided to frame problems, explore ideas, verify information, construct solutions, and reflect on their reasoning.

The framework comprises seven stages aligned with core dimensions of critical thinking, including analysis, evaluation, inference, and reflection (Cottrell, 2017). Rather than treating ChatGPT as an answer generator, the framework directs its use toward

questioning, exploration, verification, feedback, and reflective revision within the PBL cycle.

To maintain consistency across schools, the researchers prepared a common implementation guide for the teachers. The guide included the weekly problem scenario, suggested teacher prompts, student verification tasks, and reflection questions. Teachers were encouraged to adjust examples to suit classroom discussion, but they were asked to preserve the same AI-PBL stages, the same verification requirement, and the same rule that final answers had to be justified by students rather than copied from ChatGPT. This procedure was intended to keep the intervention flexible enough for natural classroom use while still sufficiently standardized for research purposes.

The role of ChatGPT was also deliberately bounded. Students were allowed to use it to ask for explanations, alternative viewpoints, examples, possible weaknesses in an argument, or questions that might be raised during presentation. However, they were not permitted to submit AI-generated text as their final group answer without discussion, revision, and evidence checking. Teachers repeatedly reminded students that an AI response was treated as a draft idea or prompt for inquiry. This boundary was central to the intervention because the study examined adaptive scaffolding, not automated answer production.

Table 2.
The stages of the AI-PBL intervention procedure

Stages	Procedure
Stage 1	Problem Framing and Contextualization. Students are presented with authentic, ill-structured problems so that inquiry begins from meaningful classroom issues and not from ready-made answers.
Stage 2	AI-Supported Exploration. ChatGPT is used to generate initial perspectives, questions, or alternative explanations that students can compare and examine critically.
Stage 3	Collaborative Sense-Making. Learners debate AI-suggested insights against textbooks, peer contributions, and online sources so that verification becomes part of collaborative discussion.

Stage 4	Adaptive Pathway Design. With teacher guidance, learners set learning targets and formulate inquiry paths, while ChatGPT provides prompts tailored to their immediate needs.
Stage 5	Multi-Source Knowledge Construction. Students triangulate ChatGPT outputs with multiple resources, turning AI into a cross-checking mechanism rather than a single source of truth.
Stage 6	Solution Development and Reflective Iteration. Groups draft solutions, use AI feedback to test alternatives or identify weaknesses, and revise their responses reflectively.
Stage 7	Presentation, Evaluation, and Knowledge Transfer. Learners present outcomes and use ChatGPT as a simulated audience to anticipate questions and strengthen the defense of their reasoning.

The strength of the AI-PBL Framework lies in its capacity to transform ChatGPT from a "shortcut tool" into a structured learning partner. Each stage deliberately channels AI use toward inquiry, verification, and reflection. The proposed AI-PBL Framework is visualized in Figure 1, which shows the relationship between the PBL cycle and ChatGPT-based adaptive scaffolding across the seven stages.

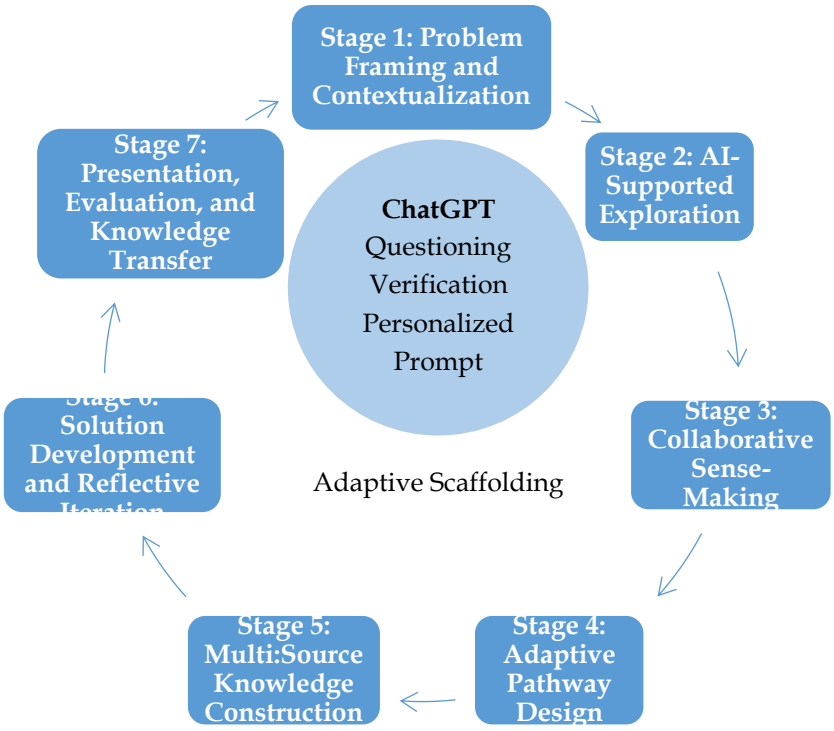

Figure 1. *AI-Enhanced Problem-Based Learning (AI-PBL) Framework*

Students worked in groups of four to five and completed a repeated sequence of problem introduction, group analysis, ChatGPT-assisted inquiry, solution formulation, and presentation/reflection. Each step served a specific function in the explanatory design because it generated both outcome data and process data showing how students questioned, verified, and refined ideas during the intervention.

Problem Introduction

The teacher introduced an authentic open-ended problem related to the lesson, distributed a short task sheet, and helped students clarify the context so that the inquiry began from a shared understanding of the problem.

Initial Group Discussion and Problem Analysis

In groups, students identified what the problem required, discussed what they already knew, and mapped what they still needed to find out. During this stage, the teacher asked probing questions without giving answers.

ChatGPT-Assisted Inquiry and Research

Students then used ChatGPT to explore concepts, generate possible explanations, and ask follow-up questions. They were explicitly required to verify AI-generated claims through textbooks, class notes, or trusted online sources so that ChatGPT functioned as a prompt for inquiry rather than a shortcut to answers.

Solution Formulation

After the inquiry phase, groups synthesized information from ChatGPT and other sources to formulate solutions. ChatGPT could also be used to test alternatives or anticipate objections, but final decisions had to be justified by evidence and group reasoning.

Presentation and Reflection

Each group presented its analysis and solution, followed by class discussion and guided reflection on both the problem and the use of ChatGPT. Reflection prompts focused on source checking, prompt revision, and the difference between receiving an answer and evaluating one.

This cycle was repeated across problem scenarios for eight weeks. Throughout the intervention, teachers maintained a consistent facilitation routine and documented implementation notes so that the same AI-PBL structure was applied across the four schools.

Implementation fidelity was monitored through teacher notes, observation records, and periodic researcher check-ins. The researchers checked whether the seven AI-PBL stages were followed, whether students worked collaboratively, whether ChatGPT outputs were verified with other sources, and whether reflective discussion occurred at the end of each problem cycle. Minor differences in classroom timing were allowed because each school had its own lesson rhythm, but the essential sequence of problem analysis, AI-supported inquiry, verification, solution formulation, and reflection remained the same across the four sites.

Ethical safeguards were integrated into the procedure. Students were instructed not to enter personal data, sensitive school information, or private details into ChatGPT. Interaction logs were collected through group-based accounts and were analyzed as learning records rather than as individual surveillance data. Names were anonymized during analysis, and interview excerpts were reported without identifying students or schools. These safeguards were necessary because the study involved minors and because generative AI use in school settings requires careful attention to privacy, consent, and responsible data handling.

Research Instruments

The study used four instruments: a critical thinking test, a classroom observation protocol, ChatGPT interaction logs, and semi-structured interview guides.

Critical Thinking Skill Test

A custom Critical Thinking Skills Test was administered one week before the intervention and immediately after the eight-week implementation. The test contained 30 items, consisting of 20 multiple-choice items and 10 open-ended items, covering analysis, evaluation, inference, bias detection, and solution design. The instrument was adapted from established critical-thinking constructs (Ennis, 2011; Facione, 1990). Three expert validators with relevant expertise in educational assessment, English language teaching, and classroom pedagogy reviewed the blueprint, content relevance, item clarity, and age appropriateness to support content validity. A pilot test with a separate Grade 8 class (n = 30) led to wording revisions. The final instrument showed acceptable internal consistency (Cronbach's α = 0.82), and two raters scored the open-ended responses with high agreement (Cohen's κ = 0.88).

The open-ended items were included because critical thinking cannot be fully captured through selected-response questions alone. Students were required to explain why a claim was weak, identify missing evidence, draw a conclusion from a short case, or propose a justified solution to a problem. The scoring rubric therefore emphasized the quality of reasoning, not only the correctness of the final answer. Responses received higher scores when students used relevant evidence, recognized assumptions, compared alternatives, and explained the logic of their decision.

Table 3.
Indicators Assessed in the Critical Thinking Test

Domain	Indicator	Item form
Analysis	Identifying relevant information and examining relationships among ideas	Multiple-choice and open-ended
Evaluation	Judging the strength of evidence, claims, or arguments	Multiple-choice and open-ended
Inference	Drawing logical conclusions from data, cases, or statements	Multiple-choice and open-ended
Bias detection	Identifying assumptions, bias, or unverified claims	Multiple-choice and open-ended
Solution design	Proposing and justifying an evidence-based response to a problem	Open-ended

Classroom Observations

A non-participant observation protocol was used to record questioning behavior, explanation of reasoning, verification behavior, collaboration patterns, and teacher facilitation during classroom sessions. The protocol was reviewed alongside the test instrument to ensure relevance to the study focus and clarity of indicators.

ChatGPT Interaction Logs

Student–ChatGPT interaction logs were archived through group-specific accounts so that prompts and responses could be reviewed as text transcripts. The logs were examined for prompt types, requests for justification, verification attempts, and moments when students accepted, questioned, or revised AI-generated ideas.

Semi-Structured Interviews

Semi-structured interviews were conducted after the intervention with 12 students and 4 facilitating teachers. Students were selected purposively to represent high-, medium-, and low-gain patterns across the four classes, while all facilitating teachers were

interviewed. The interview guides explored how participants used ChatGPT, how they judged its trustworthiness, and whether it supported or hindered independent thinking.

Data Collection

Data were collected sequentially in line with the explanatory design. Students completed the pre-test one week before the intervention and the post-test immediately after week 8. Classroom observations and ChatGPT interaction logs were gathered throughout the intervention, and interviews were conducted after the quantitative results had been reviewed so that the qualitative phase could explain the pattern of change.

During the intervention, observation data were collected in selected sessions across the four schools to capture how students moved from initial questioning to final presentation. The researchers paid particular attention to moments when students accepted, challenged, or revised ChatGPT responses. These moments were important because they showed whether students treated AI output as an endpoint or as a stimulus for further reasoning. The interview phase was then used to clarify students' intentions behind their prompts and teachers' interpretations of classroom behavior.

Table 4.
Interview Participant Profile

Participant group	N	Selection basis
Students	12	Purposive sample representing high-, medium-, and low-gain patterns across the four classes
Teachers	4	All facilitating teachers involved in the intervention

Data Analysis

The data analysis addressed the two research questions: RQ1 on the extent to which AI-PBL improved critical thinking, and RQ2 on how students interacted with ChatGPT during structured PBL activities.

Quantitative Analysis

Pre- and post-test scores were analyzed using paired-samples t-tests for the overall score and subskills. Cohen's d was calculated to estimate effect size, and one-way ANOVA was used to examine possible differences in gain scores across the four schools. All analyses were conducted in SPSS v26.

Qualitative Analysis

Qualitative data from observations, chat logs, and interviews were analyzed using thematic analysis (Braun & Clarke, 2006). The analysis followed six steps: familiarization, initial coding, searching for themes, reviewing themes, defining and naming themes, and reporting the final themes with representative evidence from the dataset.

To enhance trustworthiness, the qualitative findings were triangulated across the three sources of data. Member checking was also conducted with participating teachers and several student interviewees, while an audit trail and independent review of coded samples were used to reduce interpretive bias.

FINDINGS

This section presents the findings of the study based on both quantitative and qualitative analyses conducted to examine the effectiveness of the AI-Enhanced Problem-Based Learning (AI-PBL) framework integrating ChatGPT as adaptive scaffolding. The findings are organized according to the two research questions. First, quantitative findings from the pre-test and post-test analyses are presented to determine the extent to which the intervention improved students' critical thinking skills across several subskills. Second, qualitative findings derived from classroom observations, interaction logs, and interviews are discussed to explore how students interacted with ChatGPT during structured PBL activities. Together, these findings provide comprehensive evidence regarding the pedagogical role of generative AI in supporting critical inquiry, reflective thinking, and student agency within collaborative learning environments.

To What Extent Did AI-PBL Improve Students' Critical Thinking

RQ1 addressed whether students' critical thinking improved after the intervention. Pre-test and post-test comparisons showed a substantial increase in the overall score, from M = 60.3 (SD = 9.7) to M = 73.5 (SD = 10.4), $t(119) = 9.64$, $p < .001$, $d = 1.00$. This indicates a large and statistically significant improvement.

The size of the gain suggests that the intervention was not limited to minor test familiarity or routine classroom progress. Although the study did not use a control group, the improvement across the overall score and multiple subskills indicates that students became more capable of analyzing information, drawing conclusions, and identifying unsupported claims after participating in the AI-PBL sequence. The consistency of improvement across schools also indicates that the pattern was not concentrated in one unusually high-performing class.

Improvement was also found in the three reported subskills. Analysis increased from M = 62.0 to 77.1, inference from M = 59.5 to 72.0, and bias detection from M = 57.8 to 68.9, with all differences statistically significant. Table 4 presents the complete results.

Table 5.
Pre- and Post-Test Results on Critical Thinking Skills (N = 120)

Critical Thinking Measure	Pre-test M (SD)	Post-test M (SD)	t(119)	p	Cohen's d
Overall Score	60.3 (9.7)	73.5 (10.4)	9.64	< .001	1.00
Analysis	62.0 (12.0)	77.1 (11.2)	10.88	< .001	0.95
Inference	59.5 (11.1)	72.0 (10.3)	9.45	< .001	0.88
Bias Detection	57.8 (13.3)	68.9 (12.5)	7.90	< .001	0.75

No significant differences were found among the four schools ($p > .05$), suggesting that the pattern of improvement was comparable across the participating classes.

A closer reading of the subskill pattern shows that analysis produced the largest mean increase. This is pedagogically meaningful

because the AI-PBL tasks repeatedly required students to break down problems, identify relevant information, and separate main issues from supporting details before asking ChatGPT for help. Inference also improved substantially, suggesting that students became more confident in drawing conclusions from compared sources. Bias detection improved more moderately, which may indicate that recognizing assumptions and judging source reliability require longer and more explicit practice than identifying information or drawing conclusions.

How Did Learners Interact with ChatGPT in Structured PBL Activities

RQ2 examined how students used ChatGPT during the AI-PBL cycle. Thematic analysis of observations, interaction logs, and interviews produced four recurring patterns.

ChatGPT Encouraged Exploration and Idea Generation

Students frequently described ChatGPT as a "thinking partner" or "brainstorming companion." They used it to generate multiple perspectives and hypotheses rather than to seek direct answers. In the interaction logs, groups often opened inquiry by asking exploratory questions and then discussed the responses before drawing conclusions.

For example, a group tackling a local environmental issue used ChatGPT to list possible solutions to river pollution, debated their feasibility, and then synthesized a response that combined the AI's suggestions with local knowledge. Such episodes indicate that ChatGPT functioned as a cognitive catalyst rather than a substitute thinker.

This pattern was especially visible when students used prompts that asked for comparison, clarification, or examples instead of direct final answers. For instance, several groups asked ChatGPT to explain a problem from different stakeholder perspectives before deciding which explanation was most relevant

to the local case. In classroom discussion, students did not simply repeat the AI output; they selected, rejected, and reorganized ideas based on the task sheet and peer arguments. This shows that the exploratory function of ChatGPT became useful when students were required to process the response socially and analytically.

Verification and Information Literacy Were Strengthened

A second pattern was the strengthening of verification and information literacy. Students increasingly cross-checked ChatGPT's claims with textbooks, online sources, and peer discussion, and many interviewees explicitly stated that they learned not to accept every AI response at face value.

Teachers reinforced this habit through prompts such as "What is the source?" and "Can we confirm that statement elsewhere?" Over time, students began to do this spontaneously, indicating that verification had become part of the inquiry routine.

The interaction logs showed a gradual shift in prompt behavior. In early sessions, students tended to ask broad questions and accept the first response as useful. In later sessions, they increasingly asked follow-up questions such as whether a claim had exceptions, what evidence supported a statement, or how a response could be checked. This shift suggests that verification was not only imposed by teachers but also internalized as part of students' inquiry habits. The classroom routine trained students to view AI output as provisional information that required confirmation.

Reflection and Metacognitive Awareness Were Enhanced

A third pattern concerned reflection and metacognitive awareness. Students became more able to monitor how they framed prompts, recognized when a response was unclear or misleading, and revised their reasoning accordingly.

Teachers also observed that students increasingly verbalized their reasoning during group discussion by revisiting earlier answers, comparing alternative viewpoints, and adjusting their strategies. In this sense, ChatGPT functioned as a mirror for students' thinking rather than as an endpoint of learning.

Metacognitive growth was also reflected in students' awareness of prompt quality. Some students reported that unclear prompts produced general or confusing answers, while more specific prompts helped them receive responses that were easier to evaluate. This awareness encouraged students to revise not only the answer but also the question. In this sense, the AI-PBL cycle made questioning itself a visible learning activity. Students learned that better reasoning often begins with a clearer problem formulation and a more precise request for explanation.

ChatGPT Did Not Lead to Dependency or Passivity

The final pattern was the absence of sustained dependency or passivity. Although initial concern focused on possible overreliance, students generally remained active, inquisitive, and collaborative throughout the intervention.

Students themselves distinguished between receiving ideas and making judgments. As one student stated, "ChatGPT gives ideas, but we are the ones who decide which idea is right." Strong post-test performance without AI assistance also suggests that the gains were internalized rather than externally generated.

Taken together, the qualitative evidence shows that ChatGPT supported inquiry, verification, reflection, and agency when it was embedded in a structured PBL cycle and moderated by teacher guidance.

DISCUSSION

ChatGPT as a Cognitive Catalyst for Exploration and Idea Generation

The first major finding is that AI-PBL was associated with measurable improvement in students' critical thinking, especially in analysis and inference. This pattern is consistent with prior evidence that ChatGPT can support higher-order thinking when it is used purposefully (Wang & Fan, 2025), but the present study extends that evidence to Indonesian lower-secondary classrooms, where research has remained limited.

The finding is important because lower-secondary students are often assumed to need protection from AI rather than structured opportunities to learn how to evaluate it. The present evidence suggests a more balanced position. When AI use is framed by authentic problems, group discussion, and teacher questioning, students can use ChatGPT to widen their thinking without surrendering responsibility for judgment. Thus, AI-PBL may help teachers move beyond a binary choice between unrestricted AI use and complete prohibition.

Development of Verification and Information Literacy Skills

A second key finding is that students developed stronger habits of verification and information literacy. Rather than accepting ChatGPT responses automatically, students increasingly cross-checked claims and asked follow-up questions. This supports findings from Nasr et al. (2025) and Zhou et al. (2024), but it also adds evidence from teacher-guided PBL in secondary education, where verification routines were embedded into everyday inquiry.

This result also indicates that information literacy should be taught as a classroom habit rather than as a separate warning about technology. Students became more careful not because they were simply told that AI could be wrong, but because every problem cycle required them to compare ChatGPT responses with other sources and defend their decisions during discussion. Verification therefore became a repeated social and cognitive practice. This is a key contribution of

the framework because it translates ethical AI use into observable classroom routines.

Reflection and Metacognition as Drivers of Deep Learning

The qualitative data also show that ChatGPT supported reflection and metacognitive awareness. Students became more conscious of how they framed prompts, evaluated responses, and revised their own reasoning. In this respect, the findings align with Garrison's (2022) emphasis on reflective discourse and with Hmelo-Silver's (2004) account of PBL as a process of knowledge construction and self-regulation.

Balancing Human Facilitation and AI Assistance

Another important finding is that the intervention did not lead to sustained dependency or passivity. Although this concern has been raised in studies of unguided AI use (Abbas et al., 2024; Costa et al., 2025), the present findings suggest that teacher-guided verification and group discussion can redirect AI use toward active judgement rather than cognitive offloading.

Teacher facilitation was therefore central. ChatGPT provided immediate prompts and alternative viewpoints, but teachers maintained the conditions for questioning, source checking, and final decision making. This helps explain why the present study differs from concerns reported in less-structured AI contexts and why the framework can address the gap in secondary school evidence identified in the introduction.

Pedagogical Implications of the AI-PBL Framework

The overall findings suggest that AI-PBL is not merely a digital extension of PBL but a structured model for integrating generative AI into inquiry-based learning. By embedding ChatGPT within problem framing, exploration, verification, solution development, and reflection, the framework helps preserve the collaborative and analytical character of PBL while adding adaptive support.

This study therefore contributes not only a conceptual framework, but also mixed-methods evidence showing how AI-mediated scaffolding can improve critical thinking while sustaining verification, reflection, and learner agency in secondary education. Practically, these findings suggest that generative AI should be positioned as a thinking amplifier rather than an answer generator, that teacher guidance remains indispensable, that reflective checkpoints should be built into classroom tasks, and that school-level AI policies should emphasize ethical and critical use.

Theoretical Contribution

This study also contributes to the theoretical dialogue on the cognitive function of generative AI in learning by showing how ChatGPT can operate as a cognitive companion within a teacher-guided PBL structure (Kasneci et al., 2023). In this study, that role was observable not only in the final outcomes but also in the processes of questioning, checking, discussing, and revising.

In summary, the findings suggest that AI can support rather than erode critical thinking when it is integrated with clear pedagogical intent, structured inquiry, and sustained teacher oversight.

Nevertheless, the findings should be interpreted within the limits of the study design. The one-group pre-test and post-test procedure allowed the researchers to identify meaningful improvement, but it did not fully isolate the intervention from all possible external influences. The study was also conducted in four schools within one district, which means that the findings should not be generalized automatically to schools with different resources, teacher readiness, or digital access. Future research should involve comparison groups, longer implementation periods, and more diverse school contexts to test the robustness of the AI-PBL framework.

Another limitation concerns the scope of the measured outcomes. This study focused on critical thinking and classroom interaction, but it did not examine long-term retention, transfer to other subjects, or possible changes in students' writing quality, motivation,

or digital ethics over time. Because AI-supported learning is likely to influence several dimensions of learning simultaneously, future studies may combine critical thinking tests with measures of self-regulation, academic integrity, and language production. Such work would provide a more comprehensive picture of how generative AI reshapes learning processes in secondary education.

CONCLUSION

This study addressed two questions. First, AI-PBL was associated with significant improvement in students' critical thinking. Second, students used ChatGPT mainly to explore ideas, verify information, reflect on their reasoning, and refine solutions during structured PBL activities. Taken together, the findings suggest that ChatGPT can support critical thinking in secondary education when it is embedded in a guided PBL sequence and combined with teacher facilitation and multi-source verification. The study is limited to four junior high schools in one district, so broader and longitudinal studies are still needed.

Despite these limitations, the study provides a practical model for schools that are beginning to address generative AI in teaching and learning. The evidence suggests that ChatGPT can be educationally productive when students are not left to use it independently as an answer machine, but are guided to question, compare, verify, and revise. For teachers, the implication is that AI integration should begin with pedagogy: clear problems, structured collaboration, explicit verification routines, and reflective checkpoints. For researchers, the study demonstrates the need to examine both outcomes and interaction processes so that claims about AI-supported learning are grounded in what students actually do during classroom inquiry.

ACKNOWLEDGMENT

The authors thank the participating schools, teachers, students, and parents or guardians for their cooperation during the study.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this manuscript, the authors used AI-assisted and digital writing tools, including ChatGPT, Grammarly, Mendeley, and Microsoft Word, for limited technical support. ChatGPT was used to assist with language refinement, coherence checking, sentence restructuring, and the improvement of academic readability. Grammarly was used to identify grammatical, mechanical, and stylistic issues, while Mendeley was used to support reference management and citation organization. Microsoft Word was used for formatting, editing, and manuscript preparation.

The use of these tools did not replace the authors' scholarly judgment, conceptual development, research design, data analysis, interpretation of findings, or final decision-making. All AI-assisted suggestions were critically reviewed, revised, and validated by the authors to ensure accuracy, originality, ethical compliance, and consistency with the study's objectives. The authors take full responsibility for the content, integrity, and originality of the submitted manuscript.

REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*, 19. https://doi.org/10.1186/s41239-024-00444-7

Baig, M. I., et al. (2024). ChatGPT in higher education: A systematic literature review. *Education and Information Technologies, 29*, 5875–5899. https://doi.org/10.1007/s10639-024-12345

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning, 68*, 3–12.

Blahopoulou, J. & Ortiz-Bonnin, S. Student perceptions of ChatGPT: benefits, costs, and attitudinal differences between users and non-users toward AI integration in higher education. *Educ Inf*

Technol **30**, 19741–19764 (2025). https://doi.org/10.1007/s10639-025-13575-9

Blundell, C. N., Mukherjee, M., & Nykvist, S. (2022). A scoping review of the application of the SAMR model in research. *Computers & Education Open, 3*, 100093. https://doi.org/10.1016/j.caeo.2022.100093

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*(4), 616–630. https://doi.org/10.1007/s11528-022-00734-3

Cottrell, S. (2017). *Critical thinking skills: Developing effective analysis and argument* (3rd ed.). Palgrave Macmillan.

Costa, C., Husain-Habib, N., & Reiter, A. (2025). Integrating AI into education: Successful strategies, ideas, and tools from psychology instructors. *Teaching of Psychology, 52*(3), 330–338. https://doi.org/10.1177/00986283241297635

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.)*. Sage Publications.

Divito, C. B., Katchikian, B. M., Gruenwald, J. E., & Burgoon, J. M. (2024). The tools of the future are the challenges of today: The use of ChatGPT in problem-based learning medical education. *Medical Teacher, 46*(3), 320–322. https://doi.org/10.1080/0142159X.2024.2308841

Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities.* University of Illinois.

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and*

Education: Artificial Intelligence, 6, 100198. https://doi.org/10.1016/j.caeai.2024.100198

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. California Academic Press.

Garrison, D. R. (2022). Theoretical foundations and epistemological insights of the community of inquiry. *Interactive Learning Environments, 30*(3), 359–371. https://doi.org/10.4018/978-1-4666-2110-7.ch001

Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education, 100*(12), 4876–4883. https://doi.org/10.1021/acs.jchemed.3c00505

Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence, 8*, Article 100361. https://doi.org/10.1016/j.caeai.2025.100361

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*, 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://www.sciencedirect.com/science/article/abs/pii/S1041608023000195

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., & Colombo, L. A. (2024). Critical thinking in the age of generative AI. *Academy of Management Learning & Education, 23*(3), 373–378. https://doi.org/10.5465/amle.2024.0338

Mabina, A. (2025). Improving collaboration in higher education through ChatGPT using the ADDIE framework. *COMPETITIVE: Journal of Education, 4*(1), 81–96. https://competitive.pdfaii.org/index.php/i/article/view/162

Memarian, B., et al. (2023). ChatGPT in education: Methods, potentials, and limitations. *Computers in Human Behavior: Artificial Humans,*

1(2), 100022. https://www.sciencedirect.com/science/article/pii/S2949882123000221

Munaye, Y. Y., Admass, W., Belayneh, Y., et al. (2025). ChatGPT in education: A systematic review on opportunities, challenges, and future directions. *Algorithms, 18*(6), 352. https://doi.org/10.3390/a18060352

Nasr, N. R., Tu, C.-H., Werner, J., Bauer, T., Yen, C.-J., & Sujo-Montes, L. (2025). Exploring the impact of generative AI ChatGPT on critical thinking in higher education: Passive AI-directed use or human–AI supported collaboration? *Education Sciences, 15*(9), Article 1198. https://doi.org/10.3390/educsci15091198

Olivier, C., & Weilbach, L. (2026). GenAI chatbot tutors as agents of inclusive education: A review guided by ethics of care and critical pedagogy. In *Innovative technologies and learning: 8th International Conference, ICITL 2025, Oslo, Norway, August 5–7, 2025, Proceedings, Part I* (pp. 84–93). Springer. https://doi.org/10.1007/978-3-031-98185-2_10

Ruiz Viruel, S., et al. (2025). The role of artificial intelligence in project-based learning. *Education Sciences, 15*(2), 150. https://doi.org/10.3390/educsci15020150

Savery, J. R. (2006). Overview of problem-based learning: Definitions and distinctions. *Interdisciplinary Journal of Problem-Based Learning, 1*(1), 9–20. https://doi.org/10.7771/1541-5015.1002

Suriano, R., Plebe, A., Acciai, A., & Fabio, R. A. (2025). Student interaction with ChatGPT can promote complex critical thinking skills. *Learning and Instruction, 95*, Article 102011. https://doi.org/10.1016/j.learninstruc.2024.102011

Villan, F., & Dos Santos, R. P. (2023). ChatGPT as co-advisor in scientific initiation: Action research with project-based learning in elementary education. *arXiv Preprint, arXiv:2311.14701.*

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

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*, Article 238. https://doi.org/10.1057/s41599-025-04787-y

Younis, B. (2025). Exploring in-service teachers' lived experiences in a professional development program: An AI-TPACK framework approach. *Proceedings of the International Conference on New Trends in Teaching and Education, 2*(1), 1–24. https://doi.org/10.33422/ntteconf.v2i1.1232

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39. https://doi.org/10.1186/s41239-019-0171-0

Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. *Education Sciences, 14*(12), Article 1302. https://doi.org/10.3390/educsci14121302

Zhu, G., Fan, X., & Hou, C. (2023). Embrace opportunities and face challenges: Using ChatGPT in undergraduate students' collaborative interdisciplinary learning. *International Journal of Educational Research Open, 5*, 100298. https://doi.org/10.1016/j.ijedro.2023.100298