Voices of English Language Education Society
Vol. 10, No. 2 | August 2026 | E-ISSN 2579-7484
DOI: https://doi.org/10.29408/veles.v10i2.34176

From AI Assistance to Evidence-Based Inquiry: Developing an AI-Supported Problem-Based Learning Model for EFL Academic Reading

Yuswin Harputra[1], Elissa Evawani Tambunan[1], Yulia Rizki Ramadhani*[1]
[1]Universitas Graha Nusantara, Indonesia

Received: 24 February 2026 | Revised: 24 April 2026 | Accepted: 27 April 2026
Correspondence: yuswinharputra63@gmail.com

Abstract

Artificial intelligence (AI) can support academic reading, yet its instructional value depends on whether students use it to interrogate texts or merely obtain ready-made answers. This study developed and preliminarily evaluated an AI-supported Problem-Based Learning (PBL) model designed to promote evidence-based inquiry in EFL academic reading. Following the ADDIE framework, the study produced a six-session instructional module and a structured prompting protocol that guided students to formulate questions, examine textual evidence, evaluate arguments, and reflect on their interpretations. The model was implemented with 22 undergraduate EFL students enrolled in an Academic Reading course at a public university in Indonesia. Data were collected through expert validation, pre- and post-intervention academic reading assessments, a student perception questionnaire, classroom observations, AI-interaction logs, and reflective justification sheets. Quantitative data were analysed using descriptive statistics and a paired-samples t-test, while qualitative data were examined thematically. Students' mean academic reading score increased from 61.45 to 81.32, with a statistically significant difference between the pretest and posttest scores, $t(21) = 9.42$, $p < .001$. The most substantial improvements occurred in coherence evaluation and inferential reasoning. Questionnaire, observational, and interaction-log data also indicated positive student perceptions and progressively greater engagement in inferential questioning, textual justification, argument evaluation, and reflective revision. These findings suggest that structured AI-supported PBL may help shift students' use of AI from answer seeking toward evidence-based academic inquiry. Nevertheless, the small single-institution sample, absence of a comparison group, and lack of direct cognitive-load measurement require the findings to be interpreted as preliminary.

Keywords: academic reading; artificial intelligence literacy; evidence-based inquiry; EFL higher education; problem-based learning; structured prompting.

1. Introduction

Academic reading is central to learning in higher education because students must engage with scholarly texts to acquire disciplinary knowledge, evaluate competing perspectives, and construct academically defensible interpretations. For students learning English as a foreign language (EFL), however, academic reading involves more than understanding vocabulary and identifying explicitly stated information. It requires readers to infer relationships, distinguish claims from supporting evidence, assess argumentative coherence, and monitor whether their interpretations remain consistent with the text (Fitria & Nafiah, 2025; Safarli et al., 2025; Song et al., 2025). These demands can become particularly challenging when students have limited linguistic resources and insufficient experience in analysing complex academic discourse. Consequently, many EFL students can retrieve isolated information from a text but continue to experience difficulty evaluating arguments, connecting evidence across sections, and justifying their interpretations.

Problem-Based Learning (PBL) offers a potentially relevant framework for addressing these difficulties. Rather than positioning reading as the passive reception of information, PBL engages students in investigating authentic problems, identifying what they need to understand, evaluating available evidence, and constructing reasoned solutions. Previous studies have associated PBL with active inquiry, collaborative knowledge construction, and the development of higher-order thinking skills (Hafizah et al., 2024; Nguyễn, 2021; Wijaya, 2021). Applied to academic reading, PBL can encourage students to approach a text as a source of evidence for addressing a problem rather than as a collection of information to be recalled. Nevertheless, the quality of learning within PBL depends substantially on the scaffolding provided during inquiry (Pan & Liu, 2022). When students encounter unfamiliar vocabulary, complex arguments, and multiple sources simultaneously, insufficient guidance may result in fragmented analysis or unsupported conclusions. PBL therefore requires carefully structured support that sustains inquiry without replacing students' responsibility for interpretation.

The increasing accessibility of generative artificial intelligence (AI) has created new possibilities for providing such support. AI tools can generate questions, clarify concepts, offer feedback, and respond adaptively to learners' requests. Studies have consequently examined their potential for supporting feedback, writing, tutoring, assessment, and self-regulated learning (Chang et al., 2023; Kooli & Yusuf, 2025; Ng et al., 2024). In reading-related activities, structured interaction with AI may encourage learners to ask inferential questions, reconsider preliminary interpretations, and examine relationships between textual claims and evidence (Fu & Hiniker, 2025; Thongsan & Anderson, 2025). However, unrestricted AI use also creates a pedagogical problem. When students ask AI to summarize texts, formulate interpretations, or provide direct answers, cognitive responsibility may shift from the learner to the technology. AI assistance cannot automatically be equated with meaningful academic engagement; its value depends on how it is positioned within the instructional process.

Despite increasing research on AI-assisted learning, three related gaps remain. First, much of the existing literature treats AI as a stand-alone tool for feedback, assessment, information retrieval, or answer generation rather than as an element embedded within a coherent instructional model (Bhimavarapu, 2025; Chakraborty et al., 2024; Kanthimathi et al., 2025). Second, studies frequently report learners' achievement or perceptions but provide limited evidence of how students interact with AI while interpreting texts, questioning arguments, and revising their reasoning. Third, research integrating AI with PBL has not sufficiently explained how instructional constraints can preserve evidence-based reasoning and prevent students from delegating interpretive work to AI. These gaps are especially apparent in EFL academic reading, where learners require assistance but must remain responsible for evaluating textual evidence and constructing defensible interpretations.

The present study addresses these gaps by developing an AI-supported PBL model that positions AI as a scaffold for evidence-based inquiry. The model integrates a six-session academic reading module with a structured prompting protocol that guides students through problem orientation, textual investigation, inferential questioning, argument evaluation, collaborative discussion, and reflective revision. Before interacting with AI, students are required to formulate an initial interpretation and support it with textual evidence. AI prompts are then used to question, extend, or challenge their reasoning rather than supply final answers. This sequencing is intended to preserve learner agency while enabling AI to support inquiry at points where students require clarification or strategic direction. The novelty of the model therefore lies not simply in adding AI to PBL, but in pedagogically constraining AI use so that it redirects students from answer seeking towards evidence-based academic inquiry.

This study makes a significant contribution to scholarship and practice in three key respects. Conceptually, it reframes AI as a pedagogically regulated inquiry scaffold rather than an autonomous source of answers. Instructionally, it extends PBL into EFL academic reading by connecting problem investigation with explicit processes of inference, textual justification, argument evaluation, and reflective revision. Methodologically, it combines achievement data with student perceptions, classroom observations, AI-interaction logs, and reflective justification sheets. This combination enables the study to examine not only whether reading performance changes following implementation, but also how students engage with textual evidence and AI-supported prompts during the learning process. Nevertheless, because the model was implemented with a small, single-institution sample without a comparison group, the study provides a preliminary evaluation rather than definitive evidence of causal effectiveness.

Accordingly, this study aimed to develop and preliminarily evaluate an AI-supported PBL model for promoting evidence-based inquiry in EFL academic reading. It addressed the following research questions:

1. To what extent does the developed AI-supported PBL model meet the established criteria for content validity and instrument reliability?
2. What changes occur in students' academic reading performance following the implementation of the AI-supported PBL model?
3. How do students perceive and engage with structured AI-supported inquiry during academic reading activities?

2. Method

2.1 Research Design and Participants

This study employed a Research and Development design based on the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) framework (Branch, 2009). The framework guided the development and preliminary evaluation of an AI-supported Problem-Based Learning (PBL) model for EFL academic reading. The resulting products consisted of a six-session instructional module and structured AI prompts designed to support clarification, inferential questioning, evidence checking, argument evaluation, and reflective revision.

The implementation adopted a one-group pretest–posttest design supported by quantitative and qualitative data. Changes in academic reading performance were examined through pretest and posttest scores, while expert validation and a student perception questionnaire provided evidence regarding the quality and usability of the model. Classroom observations, AI-interaction logs, and reflective justification sheets were used to examine how students engaged with texts and AI during the intervention. As the study did not involve random assignment or a comparison group, the findings were interpreted as a preliminary evaluation of within-group change rather than conclusive evidence of causal effectiveness.

The participants were 22 undergraduate EFL students enrolled in an Academic Reading course at a public university in Indonesia. They were drawn from an intact class using convenience sampling because the class was accessible and relevant to the instructional purpose of the study (Etikan et al., 2016). All participants completed the pretest, six instructional sessions, posttest, and perception questionnaire.

2.2 Development of the AI-Supported PBL Model

The model was developed iteratively through the five phases of ADDIE (Branch, 2009). The analysis phase focused on the academic reading demands faced by students, particularly those related to understanding explicit information, drawing inferences, evaluating textual coherence, examining arguments, and justifying interpretations with evidence. It also considered the need to guide students in using AI critically rather than treating its output as an unquestioned source of answers.

During the design phase, the researchers formulated the learning objectives, instructional sequence, reading activities, assessment procedures, and principles for responsible AI use. PBL was adapted to academic reading by presenting students with text-based problems requiring investigation, interpretation, evidence selection, and evaluation of alternative explanations. This design reflected the problem-centred and inquiry-oriented principles of PBL, in which learners actively construct understanding through investigation and reasoned decision-making (Hmelo-Silver, 2004).

The development phase produced the instructional module, structured AI prompts, student worksheets, reflective justification sheets, and assessment instruments. The prompts were organised around five functions: clarification, inferential questioning, evidence checking, argument evaluation, and reflective revision. These functions provided a consistent framework but allowed students to adapt the wording of their prompts to different texts and reading problems. The module and instruments were reviewed by experts in English language education, academic reading, and instructional technology. Their ratings and written feedback were used to improve the relevance, clarity, coherence, and instructional suitability of the products before implementation.

2.3 Instructional Procedures

Before the intervention, students completed an academic reading pretest under comparable classroom conditions. They were then introduced to the PBL stages, structured AI prompts, and principles of responsible AI use. The lecturer emphasised that AI-generated information could be inaccurate, biased, or insufficiently grounded in the assigned text. Students were therefore expected to compare AI output with textual evidence and remain responsible for their final interpretations.

Each instructional session followed five stages. First, students were presented with an academic text and a contextualised problem requiring interpretation or evaluation. Second, they read the text independently, developed an initial response, and identified relevant supporting evidence. Third, they used structured AI prompts to clarify difficult information, examine possible inferences, consider alternative interpretations, evaluate arguments, or test the adequacy of their evidence. Fourth, students compared the AI-generated responses with the original text and discussed their reasoning collaboratively. Finally, they revised their responses and completed a reflective justification sheet explaining which AI suggestions they accepted, modified, or rejected and the reasons for their decisions.

Instructional support was gradually reduced across the six sessions. In the initial sessions, the lecturer modelled how to formulate purposeful prompts and critically verify AI output. In subsequent sessions, students selected and adapted the prompts more independently according to the problems encountered in each text. This progression positioned students as active decision-makers and encouraged them to use AI as a flexible inquiry aid rather than a substitute for reading and reasoning. After the final session, students completed the posttest and perception questionnaire, and their AI-interaction logs and reflective justification sheets were collected.

2.4 Instruments and Data Collection

The academic reading assessment consisted of 16 items covering four dimensions: explicit comprehension, inferential reasoning, coherence evaluation, and critical evaluation. Each dimension was represented by four items. The same assessment framework and scoring criteria were applied to the pretest and posttest, and the resulting scores were converted to a scale of 0–100. The instrument was reviewed by experts for item relevance, clarity, construct representation, and alignment with the learning objectives. Its internal consistency was examined using Cronbach's alpha (Cronbach, 1951).

The student perception questionnaire was administered after the intervention. It examined students' perceptions of the instructional sequence, structured AI prompts, engagement with academic texts, use of textual evidence, reflective revision, and responsible AI use. The questionnaire items were rated using the Likert scale specified in the instrument. Expert review was used to establish their relevance and clarity, while Cronbach's alpha was employed to examine internal consistency. The questionnaire results represented students' experiences with the model and were not treated as direct evidence of improved reading performance.

Classroom observations were conducted during all six sessions using a structured checklist and descriptive field notes. They documented students' participation in problem investigation, inferential questioning, evidence use, argument evaluation, collaborative discussion, verification of AI output, and reflective revision. The observations were used to examine the implementation process and provide contextual evidence for interpreting the assessment and questionnaire results.

AI-interaction logs recorded the prompts submitted by students, AI-generated responses, follow-up questions, and subsequent revisions. Reflective justification sheets documented students' initial interpretations, supporting evidence, relevant AI feedback, revisions, and reasons for accepting or rejecting

AI suggestions. Together, these sources provided qualitative evidence of how students used AI while retaining responsibility for their interpretations. The observation checklist was treated as an implementation record rather than an independent psychometric scale.

2.5 Data Analysis and Ethical Considerations

Quantitative data were analysed using descriptive and inferential statistics. Expert-validation ratings and questionnaire responses were summarised using means, frequencies, and percentages, while academic reading scores were described using means and standard deviations. After the distribution of the pretest–posttest difference scores had been examined, a paired-samples **t**-test was conducted to determine whether the change between the two measurement points was statistically significant at **p** < .05. Cohen's **d** for paired observations was calculated to estimate the magnitude of the within-group change. The effect size was interpreted cautiously because the design did not include a comparison group.

Qualitative data from classroom observations, AI-interaction logs, and reflective justification sheets were analysed thematically following Braun and Clarke (2006). The analysis involved familiarisation with the data, initial coding, grouping related codes, developing themes, and reviewing the themes against the dataset. Particular attention was given to inferential questioning, textual justification, argument evaluation, verification of AI output, and reflective revision. Evidence from the three sources was compared to identify converging, complementary, or contrasting patterns. The qualitative findings were then used to contextualise and extend the quantitative results.

The study followed the principles of voluntary participation, informed consent, confidentiality, and responsible data management. Participants were informed about the objectives, procedures, types of data collected, and their right to withdraw without academic consequences. Personal identifiers were removed, and the findings were reported in aggregate or anonymised form. Students were also instructed not to enter personal, confidential, or institutionally sensitive information into the AI platform. Because AI output may contain inaccuracies or unsupported claims, they were required to verify relevant responses against the assigned texts before incorporating them into their final interpretations.

3. Findings

The findings are presented according to the three research questions. The first subsection reports the validity of the developed model, the reliability of the research instruments, and the revisions made before implementation. The second presents changes in students' academic reading performance, while the third reports students' perceptions and engagement during the intervention.

3.1 Validity and Refinement of the AI-Supported PBL Model

The first research question concerned the quality of the developed AI-supported PBL model and its supporting instruments. Three experts evaluated six components using a four-point relevance scale. The components included the alignment between the learning objectives and PBL stages, authenticity of the problem scenarios, scaffolding functions of the structured AI prompts, clarity and feasibility of the learning activities, coverage of the academic reading dimensions, and relevance of the student perception questionnaire. As presented in Figure 1, the S-CVI/Ave values ranged from 0.89 to 0.95. The structured AI prompts received the highest value (0.95), followed by the alignment between the learning objectives and PBL stages (0.93), clarity and feasibility of the learning activities (0.92), relevance of the student perception questionnaire (0.91), authenticity of the problem scenarios (0.90), and coverage of the academic reading dimensions (0.89). All components exceeded the minimum criterion of 0.80, indicating satisfactory content validity.

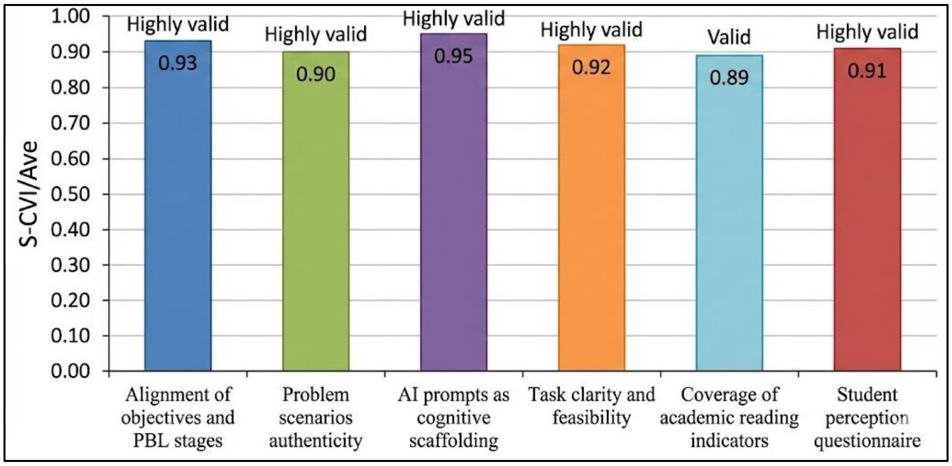

Figure 1. Expert Validation Results for the AI-Supported PBL Model

The academic reading assessment and student perception questionnaire also demonstrated satisfactory internal consistency. As shown in Table 1, the academic reading assessment produced a Cronbach's alpha coefficient of 0.84, while the student perception questionnaire obtained a coefficient of 0.91. Both instruments therefore exceeded the predetermined reliability criterion of 0.70.

Table 1
Reliability of the Research Instruments

Instrument	Cronbach's α	Interpretation
Academic reading assessment	0.84	Good
Student perception questionnaire	0.91	Excellent

In addition to providing quantitative ratings, the experts recommended several revisions to improve the instructional and assessment components. As summarised in Table 2, generic prompts were replaced with more targeted prompts requiring students to clarify information, compare interpretations, examine textual evidence, evaluate arguments, and reflect on their conclusions. The task instructions were simplified and sequenced more clearly, the difficulty of the reading texts was adjusted, and the assessment criteria were clarified across the four academic reading dimensions.

Table 2
Revisions to the AI-Supported PBL Model

Component	Revision	Intended improvement
Structured AI prompts	Generic prompts were replaced with targeted prompts	Strengthen inquiry and reasoning support
Task instructions	Instructions were simplified and sequenced	Improve procedural clarity
Reading texts	Text difficulty was adjusted	Improve accessibility for students
Assessment criteria	Criteria for the four reading dimensions were clarified	Improve scoring consistency

Together, the expert-validation results indicated that the instructional model achieved satisfactory content validity, while the reliability analysis supported the internal consistency of the two principal measurement instruments. The experts' recommendations were incorporated before classroom implementation.

3.2 Changes in Students' Academic Reading Performance

The second research question examined changes in students' academic reading performance following the six-session intervention. As presented in Table 3, the students' mean score increased from 61.45 (SD = 6.82) on the pretest to 81.32 (SD = 7.15) on the posttest, representing a mean gain of 19.87 points. The paired-samples t-test showed that the difference was statistically significant, $t(21) = 9.42$, $p < .001$. Cohen's d for paired observations was 2.01, indicating a substantial within-group change.

Table 3
Pretest and Posttest Academic Reading Results

Measurement	Mean	SD	Mean gain	t(21)	p	Cohen's d
Pretest	61.45	6.82	—	—	—	—
Posttest	81.32	7.15	19.87	9.42	< .001	2.01

Improvement was also observed across all four academic reading dimensions. Figure 2 shows that coherence evaluation recorded the largest gain of 5.8 points, followed by inferential reasoning with a gain of 5.3 points, critical evaluation with a gain of 4.5 points, and explicit comprehension with a gain of 4.2 points. The two largest gains were therefore found in dimensions requiring students to establish relationships across a text and construct meanings that were not stated directly.

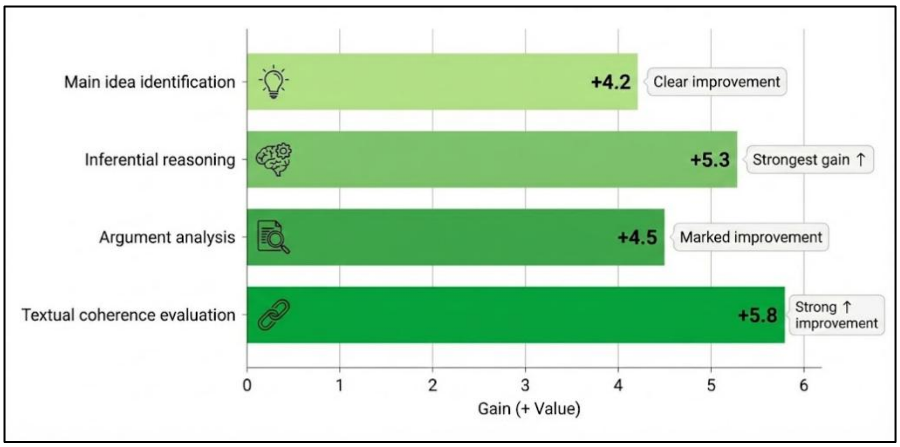

Figure 2. Gains across the Four Academic Reading Dimensions

These findings demonstrate a statistically significant and substantial improvement in students' academic reading performance after the intervention. Nevertheless, because the study employed a one-group pretest–posttest design, the results represent within-group change and should not be interpreted as conclusive evidence that the intervention alone caused the improvement.

3.3 Students' Perceptions and Engagement during AI-Supported PBL

The third research question examined students' perceptions of the AI-supported PBL model and their engagement during its implementation. Questionnaire responses were interpreted using four categories: very positive (3.26–4.00), positive (2.51–3.25), negative (1.76–2.50), and very negative (1.00–1.75). As presented in Table 4, all five evaluated aspects were rated very positively.

Table 4
Students' Perceptions of AI-Supported PBL

Aspect	Mean	Category
Ease of AI use	3.72	Very positive

Overall satisfaction	3.70	Very positive
Support for inference and critical evaluation	3.68	Very positive
Usefulness for reading comprehension	3.66	Very positive
Engagement and collaboration	3.54	Very positive

Ease of AI use received the highest mean score (M = 3.72), followed by overall satisfaction (M = 3.70). Students also perceived the model as supportive of inferential reasoning and critical evaluation (M = 3.68) and useful for reading comprehension (M = 3.66). Engagement and collaboration received the lowest mean score (M = 3.54), although it remained within the very positive category. Overall, the questionnaire results show that the model was perceived as accessible, useful, and supportive of academic reading.

The classroom observations, AI-interaction logs, and reflective justification sheets provided further evidence of changes in students' engagement. As summarised in Table 5, evidence-based reasoning developed from a low–moderate level in the early sessions to a high level in the later sessions. Inferential questioning increased from low to high, while critical evaluation progressed from moderate to high. Reflective revision also developed from low to moderate–high.

Table 5
Changes in Students' Engagement during the Intervention

Engagement indicator	Early sessions	Later sessions
Evidence-based reasoning	Low–moderate	High
Inferential questioning	Low	High
Critical evaluation	Moderate	High
Reflective revision	Low	Moderate–high

During the early sessions, students frequently produced general or literal responses and required lecturer guidance to formulate purposeful prompts. In the later sessions, their AI-interaction logs showed more frequent use of prompts for clarification, inference, evidence checking, and critical evaluation. The reflective justification sheets also contained more explicit explanations of why particular AI suggestions were accepted, modified, or rejected. These patterns indicate that students gradually exercised greater control over their AI-supported reading processes and became more accountable for their interpretations. Together, the questionnaire and process data show that students perceived the AI-supported PBL model positively and gradually demonstrated more evidence-based, inferential, critical, and reflective engagement with academic texts.

4. Discussion

This study developed and preliminarily evaluated an AI-supported Problem-Based Learning (PBL) model for EFL academic reading. The findings showed that the model was considered appropriate by experts, students' academic reading performance improved after its implementation, and students responded positively to the learning experience. The most notable improvements occurred in coherence evaluation and inferential reasoning, while classroom observations, AI-interaction records, and reflective sheets indicated increasingly critical use of AI across the six sessions. Taken together, these findings suggest that generative AI can support academic reading when it is integrated into a clear instructional sequence that requires students to read independently, examine textual evidence, and evaluate rather than simply reproduce AI-generated responses.

The expert-validation results indicate that the model established reasonable alignment among the learning objectives, PBL stages, reading activities, structured AI prompts, and assessment procedures. The experts' recommendations concerning task sequencing, text difficulty, and assessment criteria also demonstrate that effective AI integration requires careful instructional design. Kasneci et al. (2023) emphasised that the educational benefits of generative AI depend largely on how it is incorporated into teaching, while Law (2024) similarly found that technological affordances alone do not guarantee meaningful language learning. In this study, the validation and revision process helped ensure that AI supported the inquiry process without replacing the essential stages of academic reading.

This principle was reflected in the instructional sequence. Before consulting AI, students were required to read the text, formulate an initial interpretation, and identify relevant evidence. They then used structured prompts to clarify difficult information, explore possible inferences, examine arguments, and reconsider their interpretations. Finally, they compared the generated responses with the source text and justified whether those responses should be accepted, modified, or rejected. This sequence positioned AI as a source of assistance rather than a source of final answers. Such an arrangement is important because fluent and persuasive AI responses may contain inaccurate or unsupported information. As Farrokhnia et al. (2024) observed, learners need sufficient guidance to evaluate the apparent credibility of AI output, while Yan et al. (2024) highlighted verification as an important requirement for responsible educational use of large language models.

The improvement in students' academic reading performance suggests that structured AI assistance complemented the inquiry-oriented characteristics of PBL. Hmelo-Silver (2004) explained that PBL engages learners in investigating meaningful problems, identifying relevant information, considering alternative explanations, and constructing defensible conclusions. These processes are closely related to academic reading, in which students must connect ideas, interpret implicit meanings, evaluate evidence, and judge the strength of claims. In the present model, the problems gave students a meaningful purpose for reading, while AI-supported questions helped them examine the text from different perspectives. The observed improvement should nevertheless be interpreted as within-group development because the study did not include a comparison group.

The particularly strong gains in coherence evaluation and inferential reasoning correspond with the focus of the instructional activities. These two dimensions require students to move beyond identifying explicitly stated information. According to Kendeou et al. (2016), readers construct deeper comprehension by connecting information across a text and generating inferences from textual evidence and prior knowledge. The structured prompts in this study repeatedly asked students to identify relationships among ideas, explore meanings that were not directly expressed, compare alternative interpretations, and locate evidence supporting or challenging a claim. These activities may have made complex reading operations more explicit and systematic. The smaller improvement in explicit comprehension may reflect students' stronger initial ability in locating directly stated information, as well as the intervention's greater emphasis on interpretive and evaluative reading. However, further comparative testing is required before concluding that the model is more effective for certain reading dimensions.

The students' positive perceptions indicate that the model was acceptable and manageable in classroom practice. They generally considered the AI-supported activities accessible and useful for understanding texts, constructing inferences, and evaluating information. This finding is consistent with Law's (2024) conclusion that generative AI can provide immediate explanations and opportunities for independent language learning. Nevertheless, favourable perceptions do not independently demonstrate instructional effectiveness because students' responses may also have been influenced by convenience or the novelty of using AI. In this study, the perception data are more meaningful when considered together with the performance results and qualitative evidence concerning how students used AI during the intervention.

The process data showed that students' engagement became more purposeful over time. During the initial sessions, many students produced relatively literal interpretations and relied on lecturer modelling to formulate suitable prompts. In subsequent sessions, their interaction records and reflective sheets showed more frequent attempts to explore inferences, question AI-generated explanations, identify textual evidence, and revise initial interpretations. This progression suggests that students gradually moved from receiving AI assistance to managing it more critically. Although the term interpretive agency should not be treated as a separately measured construct in this study, it usefully describes the students' growing responsibility for deciding how AI responses contributed to their interpretations. Their agency was maintained by requiring an initial reading, textual verification, and justification of the final decision.

One result that requires further attention is that engagement and collaboration, despite remaining in the very positive category, received the lowest questionnaire mean. This suggests that access to AI does not automatically strengthen interaction among students. PBL typically benefits from collaborative discussion through which learners compare interpretations, challenge assumptions, and negotiate solutions. Yew and Goh (2016) regarded this social process as an important element of PBL, but individual consultation with AI may reduce opportunities for peer negotiation when collaboration is not deliberately incorporated. Future implementation should therefore provide more structured opportunities for students to compare AI responses, discuss unsupported claims, defend competing interpretations, and jointly determine which conclusion is most consistent with the text.

Overall, the study proposes a simple conceptual pathway in which problem-based inquiry provides the purpose for reading, structured AI assistance supports particular interpretive activities, and textual verification and reflection maintain students' responsibility for the resulting interpretation.

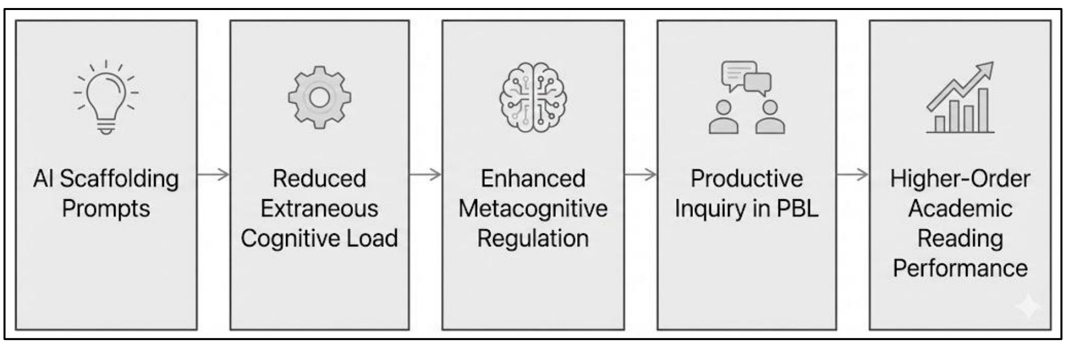

Figure 4. Conceptual Pathway of AI-Supported PBL for Academic Reading Development

Figure 4 should be understood as a conceptual representation of the instructional process rather than a statistically tested model. It proposes that academic reading development does not arise from AI use alone but from the sequence through which students investigate a problem, consult AI for specific purposes, verify the generated information, and reflect on their interpretations. This process-oriented contribution is relevant because, as Lo et al. (2024) reported, research on generative AI in ESL/EFL education has concentrated predominantly on writing and learner perceptions, while reading performance and classroom interaction processes remain less extensively investigated. Pedagogically, lecturers should therefore introduce AI after students have attempted the text independently, align prompts with specific reading purposes, and require all generated information to be checked against textual evidence. Instruction should also address inaccurate output, bias, privacy, transparency, and appropriate disclosure, which Yan et al. (2024) identified as central concerns in educational AI use.

5. Limitations and Directions for Future Research

The findings should be interpreted within several limitations. The study involved only 22 students from one intact class at a single Indonesian university and did not include random assignment or a comparison group. Consequently, the observed improvement cannot be attributed exclusively to AI-supported PBL. Regular instruction, repeated exposure to academic texts, lecturer assistance, familiarity with the assessment, and novelty effects may also have influenced the results. In addition, the six-session implementation does not establish whether the gains would be retained or transferred to unfamiliar texts, other disciplines, or independent reading contexts. The qualitative findings were also based on classroom observations, AI-interaction records, and reflective documents; therefore, interpretations concerning students' increasingly critical use of AI should not be treated as direct measurements of a separate psychological construct.

Future research should evaluate the model with larger and more diverse samples and compare it with conventional PBL, unrestricted AI use, and non-AI reading instruction. Delayed posttests would help determine whether improvements are sustained and transferable. Researchers could also examine individual reading dimensions more closely, analyse the quality and timing of students' prompts, assess the accuracy of AI-generated responses, and investigate how AI consultation affects peer collaboration. Such studies would provide stronger evidence concerning whether the observed development results specifically from the model and would clarify the instructional conditions under which generative AI supports academic reading without replacing students' responsibility for interpretation.

6. Conclusion

This study developed and preliminarily evaluated an AI-supported Problem-Based Learning model for EFL academic reading. Expert validation indicated that the model appropriately aligned learning objectives, PBL stages, structured AI prompts, reading activities, and assessment procedures. Following the six-session implementation, students demonstrated significant improvement in academic reading, with the strongest gains occurring in coherence evaluation and inferential reasoning. Their positive perceptions and progressively more critical use of AI also suggest that the model was feasible and acceptable in classroom practice. These findings indicate that generative AI can support academic reading when it is embedded in a structured learning process rather than used merely to generate summaries or answers.

The main contribution of the model lies in combining problem-based inquiry, purposeful AI assistance, textual verification, and reflection. Students were required to develop an initial interpretation, use AI for specific reading purposes, evaluate its responses against textual evidence, and justify their final decisions. Nevertheless, the findings represent preliminary within-group evidence and should not be interpreted as proof of causal effectiveness because the study involved a small sample, a short intervention, and no comparison group. Further studies with larger samples, control conditions, and delayed assessments are needed to examine the model's effectiveness, transferability, and long-term contribution to EFL academic reading.

7. Acknowledgements

The authors would like to express their sincere appreciation to the experts who evaluated the instructional model and provided constructive recommendations for its refinement. The authors also thank the lecturer and students who participated in the implementation of the model and contributed valuable data and reflections throughout the study.

8. Declaration of Generative AI Use

During the preparation of this manuscript, the authors used Scopus AI to assist in identifying relevant scholarly literature and Grammarly to support language refinement and improve the clarity of selected sentences. These tools were not used to generate or analyse the research data, determine the findings, or formulate the study's conclusions. The authors critically reviewed, verified, and revised all AI-assisted content and take full responsibility for the accuracy, integrity, and originality of the final manuscript.

References

Alifah, N., & Hidayat, A. R. (2025). Effectiveness of artificial intelligence-based learning analytics tools in supporting personalized learning in higher education. Jurnal Pendidikan Progresif, 15(1), 74–84. https://doi.org/10.23960/jpp.v15i1.pp74-84

Arani, S. M. N. (2025). AI-mediated problem-based language learning: Easing the load for teachers, empowering students. Computer Assisted Language Learning, 1–27. https://doi.org/10.1080/09588221.2025.2599156

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. Review of Educational Research, 87(2), 309–344. https://doi.org/10.3102/0034654316670999

Berenji, S. (2021). Enhancing metacognitive scaffolding and comprehension ability through problem-based learning in an EFL context. Education Research International, 2021, Article 6766793. https://doi.org/10.1155/2021/6766793

Bhimavarapu, U. (2025). Personalized learning in the digital era: Unlocking potential through AI and data analytics. In J. Basha, T. Alade, M. Al Khazimi, R. Vasudevan, & J. Khan (Eds.), Optimizing research techniques and learning strategies with digital technologies (pp. 187–210). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-7863-2.ch007

Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. https://doi.org/10.1007/978-0-387-09506-6

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

Chakraborty, S., Misra, B., & Dey, N. (2024). AI-empowered knowledge management in education. Springer. https://doi.org/10.1007/978-981-97-2574-8

Chang, D. H., Lin, M. P., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbots that support self-regulated learning in education: Goal setting, feedback, and personalization. Sustainability, 15(17), Article 12921. https://doi.org/10.3390/su151712921

Chaudhary, J., & Goyal, J. S. (2025). The future of AI in education: A conceptual exploration. International Journal of Indian Psychology, 13(1). https://doi.org/10.25215/1301.289

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. https://doi.org/10.1007/BF02310555

Dignath, C., & Veenman, M. V. J. (2021). The role of direct strategy instruction and indirect activation of self-regulated learning: Evidence from classroom observation studies. Educational Psychology Review, 33(2), 489–533. https://doi.org/10.1007/s10648-020-09534-0

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. Journal of Educational Psychology, 109(6), 761–781. https://doi.org/10.1037/edu0000180

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American Journal of Theoretical and Applied Statistics, 5(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. Journal of Educational Computing Research, 62(8), 1741–1774. https://doi.org/10.1177/07356331241268349

Fitria, W., & Nafiah, U. (2025). Enhancing students' inferential reading skills through a higher-order-thinking-skills-based module: An experimental study. International Journal of Education and Teaching Zone, 4(2).

https://doi.org/10.57092/ijetz.v4i2.434

Fu, Y., & Hiniker, A. (2025). Supporting students' reading and cognition with AI [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2504.13900

Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed., pp. 83–106). Routledge.

Hafizah, M., Solin, S., Purba, C. T., Sihotang, M. M., Rahmad, R., & Wirda, M. A. (2024). Meta-analysis: The impact of problem-based learning (PBL) models on students' critical thinking skills. Journal of Digital Learning and Education, 4(3), 1–13. https://doi.org/10.52562/jdle.v4i3.1393

Handage, S., & Chander, M. (2021). Development of an instrument for measuring student learning outcomes: A content validation process. Indian Journal of Extension Education, 57(3), 1–7. https://doi.org/10.48165/ijee.2021.57302

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? Educational Psychology Review, 16(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Jaiswal, P. (2025). Cognitive, metacognitive, and digital reading strategies in English education: A 20-year review. International Journal of Linguistics, Literature and Translation, 8(9), 10–18. https://doi.org/10.32996/ijllt.2025.8.9.2

Johnson, R. B., & Christensen, L. (2020). Educational research: Quantitative, qualitative, and mixed approaches (7th ed.). SAGE.

Kanthimathi, S., Nanda, P., Venkatraman, S., Renganayagan, V., & Eswaran, S. (2025). Transforming education through AI-powered personalized assessment models. In Adopting artificial intelligence tools in higher education (pp. 136–154). CRC Press.

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., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. Policy Insights from the Behavioral and Brain Sciences, 3(1), 62–69. https://doi.org/10.1177/2372732215624707

Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. Educational Psychology Review, 30(2), 397–429. https://doi.org/10.1007/s10648-017-9419-1

Kooli, C., & Yusuf, N. (2024). Transforming educational assessment: Insights into the use of ChatGPT and large language models in grading. International Journal of Human–Computer Interaction, 1–12. https://doi.org/10.1080/10447318.2024.2338330

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI-empowered self-regulated learning in higher education. npj Science of Learning, 10(1), Article 21. https://doi.org/10.1038/s41539-025-00319-0

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. Computers and Education Open, 6, Article 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11, Article 50. https://doi.org/10.1186/s40561-024-00342-5

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson.

Matcha, W., Gašević, D., Jovanović, J., Uzir, N. A. A., Oliver, C. W., & Murray, A. (2020). Analytics of learning strategies: The association with personality traits. In Proceedings of the 10th International Conference on Learning Analytics & Knowledge (pp. 151–160). Association for Computing Machinery.

Molenaar, I. (2022). The concept of hybrid human–AI regulation: Exemplifying how to support young learners' self-regulated learning. Computers and Education: Artificial Intelligence, 3, Article 100070. https://doi.org/10.1016/j.caeai.2022.100070

Nguyễn, T. M. T. (2021). Using problem-based tasks to promote higher-order thinking skills for TESOL MA students in Vietnam. Interdisciplinary Journal of Problem-Based Learning, 15(1). https://doi.org/10.14434/ijpbl.v15i1.28772

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. Psychological Science, 30(9), 1344–1351. https://doi.org/10.1177/0956797619862276

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, Article 100020. https://doi.org/10.1016/j.caeai.2021.100020

Pan, Z., & Liu, M. (2022). The role of an adaptive scaffolding system in supporting middle school problem-based learning activities. Journal of Educational Technology Systems, 51(2), 117–145. https://doi.org/10.1177/00472395221133855

Rokeman, N. R. M. (2024). Likert measurement scale in education and social sciences: Explored and explained. EDUCATUM Journal of Social Sciences, 10, 77–88. https://doi.org/10.37134/ejoss.vol10.1.7.2024

Safarli, N. T., Jafarova, I. M., Nazirzada, L. N., Jumayeva, L., Deepa, B., Antar, D., & Mehdizade, N. S. (2025). Foreign language proficiency among EFL students in higher education and usage of artificial intelligence. Forum for Linguistic Studies, 7(8), 646–665. https://doi.org/10.30564/fls.v7i8.11067

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally focused analytics of self-regulated learning: A systematic review of literature. Computers and Education: Artificial Intelligence, 3, Article 100060. https://doi.org/10.1016/j.caeai.2022.100060

Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. In A. Walker, H. Leary, C. E. Hmelo-Silver, & P. A. Ertmer (Eds.), Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows (pp. 5–15). Purdue University Press.

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.

Shin, M. (2021). Development of an English teaching model applying artificial intelligence through maker education. Journal of the Korea Convergence Society, 12(3), 61–67. https://doi.org/10.15207/JKCS.2021.12.3.061

Song, X., Razali, A. B., & Jeyaraj, J. J. (2025). How project-based learning improves college EFL learners' critical thinking skills and reading comprehension ability: A case study. Language Teaching Research. Advance online publication. https://doi.org/10.1177/13621688251352275

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Taky-Eddine, O., & Madaoui, R. (2024). Cognitive overload in the hypertext reading environment. International Journal of English Language Studies, 6(2), 94–100. https://doi.org/10.32996/ijels.2024.6.2.13

Thongsan, N. C., & Anderson, N. J. (2025). From passive answers to active inquiry: How AI supports critical reading in EFL classrooms. LEARN Journal: Language Education and Acquisition Research Network, 18(2), 795–820. https://doi.org/10.70730/kmkl8505

Torres, V. (2025). Unlocking potential: Teaching bilingual students how to think using metacognitive strategies [Unpublished manuscript or graduate project]. California State University, Northridge.

Tuononen, T., Hyytinen, H., Räisänen, M., Hailikari, T., & Parpala, A. (2022). Metacognitive awareness in relation to university students' learning profiles. Metacognition and Learning, 18(1), 37–54. https://doi.org/10.1007/s11409-022-09314-x

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. Educational Psychology Review, 17(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. In Proceedings of the 10th International Conference on Learning Analytics & Knowledge (pp. 524–533). Association for Computing Machinery. https://doi.org/10.1145/3375462.3375483

Wijaya, K. F. (2021). Collaborative and problem-based learning in promoting Indonesian EFL learners' learning autonomy. Polyglot: Jurnal Ilmiah, 17(2), 191–211. https://doi.org/10.19166/pji.v17i2.3590

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed., pp. 36–48). Routledge.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), 90–112. https://doi.org/10.1111/bjet.13370

Yew, E. H. J., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. Health Professions Education, 2(2), 75–79. https://doi.org/10.1016/j.hpe.2016.01.004

Yin, J., & Gulnaz, A. (2025). Enhancing teaching effectiveness in ancient architecture colour painting through blended design. [The supplied journal title, SHILAP Revista de Lepidopterología, does not match DOI https://doi.org/10.48619/ais.v6i1.1093; verify the actual journal title before inclusion.]

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, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk & B. J. Zimmerman (Eds.), Self-regulation of learning and performance: Issues and educational applications (pp. 3–21). Routledge. https://doi.org/10.4324/9780203763353-1