

CONFERENCE PAPER • KICE 2026 • YOGYAKARTA, INDONESIA

When Can ChatGPT Foster Higher-Order Thinking? A Critical Review of Its Pedagogical Role in EFL Education

Vini Rizki Nurodiah¹, Titin Nurjanah², Muhammad Aminuddin³, Andang Saehu⁴, Cipto Wardoyo⁵,
Dedi Sulaeman⁶

¹²³⁴⁵⁶ English Education, Postgraduate Programme, UIN Sunan Gunung Djati, Bandung, Indonesia

¹Email: vinirizki93@gmail.com

Proceedings of the Kalijaga International Conference on Education 2026

Abstract. The rapid integration of generative artificial intelligence (GenAI) into education has generated both optimism regarding its pedagogical potential and concern about students' increasing reliance on AI-generated content. This paper critically examines the conditions under which ChatGPT may contribute to the development of higher-order thinking skills (HOTS) in English as a Foreign Language (EFL) education. The study employs a qualitative critical literature review to synthesise empirical studies, systematic reviews, meta-analyses, and relevant theoretical literature concerning ChatGPT, higher-order thinking, critical thinking, cognitive engagement, and cognitive offloading. The reviewed evidence is interpreted through the analysing, evaluating, and creating dimensions of the revised Bloom's taxonomy. The synthesis indicates that ChatGPT can provide meaningful cognitive and linguistic support when learners are required to question, compare, verify, evaluate, and transform AI-generated information. However, such benefits are not inherent in the technology itself. Rather, they depend substantially on instructional design, teacher guidance, and the extent to which learners retain responsibility for essential cognitive processes. The review also identifies cognitive offloading and uncritical acceptance of AI-generated content as important risks when ChatGPT is used primarily as an answer-generating tool. Overall, the evidence suggests that ChatGPT is most pedagogically valuable when positioned as a dialogic and cognitive scaffold rather than as a substitute for learner thinking. The paper therefore proposes a conditional perspective in which the relationship between ChatGPT and HOTS is mediated by instructional design and learner cognitive engagement. This perspective has particular relevance for EFL pedagogy because AI-supported language learning can be designed to develop linguistic competence while simultaneously strengthening analysis, evaluation, creation, and reflective inquiry.

Keywords: generative artificial intelligence; ChatGPT; higher-order thinking skills; EFL; critical thinking; cognitive offloading; language education

INTRODUCTION

The rapid development of artificial intelligence (AI) has transformed the ways in which students access information, receive feedback, practise language, and construct knowledge. Within this broader development, generative artificial intelligence has attracted particular attention because it can produce contextually responsive explanations, examples, written responses, and conversational exchanges. Earlier research on artificial intelligence in education focused largely on intelligent tutoring systems, adaptive learning environments, automated assessment, and personalised educational support [1]. The emergence of generative AI has expanded this landscape by introducing more flexible and interactive forms of learner-technology communication, particularly through systems that allow users to engage in sustained dialogue rather than simply retrieve predetermined information.

ChatGPT has become especially relevant to education because its conversational design enables learners to ask questions, request clarification, generate ideas, obtain feedback, and revise responses through repeated interaction. Its relevance is particularly visible in language education, where learners can use the system to practise writing, explore vocabulary and grammar, obtain explanations, and engage in simulated communication. Barrot [2] notes that ChatGPT provides accessibility, personalised interaction, immersive learning opportunities, and immediate feedback, while also emphasising that its educational value is strongest when it complements rather than replaces conventional instruction. This distinction is important because the pedagogical significance of ChatGPT does not depend simply on what the system can produce, but on how learners interact with what it produces.

The growing use of ChatGPT has also generated an important debate regarding its effect on students' cognitive development. On the one hand, generative AI may encourage inquiry by providing alternative explanations, examples, perspectives, and responses that learners can examine and discuss. A systematic review and meta-analysis of experimental studies by Deng et al. [3] found that ChatGPT interventions were associated with improved academic performance, affective-motivational states, and higher-order thinking

propensities, while also reporting reduced mental effort in learning tasks. Such evidence indicates that ChatGPT may have meaningful pedagogical potential when incorporated into educational activities.

On the other hand, the availability of immediate AI-generated answers may reduce the need for learners to perform effortful cognitive work. This concern can be interpreted through the concept of cognitive offloading, which refers to the use of external resources to reduce the mental effort associated with performing a task [4]. Cognitive offloading is not inherently undesirable because external tools may reduce unnecessary cognitive demands and allow learners to focus attention on more meaningful activities. The concern arises when the cognitive processes transferred to the technology are themselves the intended learning outcomes. In such circumstances, technology may no longer function as a scaffold but may instead substitute for the learner's own reasoning.

This concern is particularly important in the context of ChatGPT because the system can generate explanations, arguments, summaries, ideas, and even complete written products. Zhai et al. [5] found that over-reliance on AI dialogue systems can create risks for critical thinking, analytical reasoning, and decision-making, particularly when learners accept AI-generated recommendations without adequately evaluating their reliability. Gerlich [6] similarly reported a relationship between frequent AI use, increased cognitive offloading, and lower critical-thinking performance, although such evidence should be interpreted in relation to the specific methods and populations examined.

The tension between assistance and substitution becomes especially important when considering higher-order thinking skills. The revised Bloom's taxonomy describes analysing, evaluating, and creating as higher-level cognitive processes that require learners to examine information, make reasoned judgements, and construct new ideas or products [7]. ChatGPT may provide material that can be analysed or evaluated and can support brainstorming for creative activities, but the existence of AI assistance does not itself demonstrate that higher-order thinking has occurred. The cognitive process remains meaningful only when learners are required to engage with the information actively and retain responsibility for decisions concerning what to accept, reject, modify, or create.

This issue has particular significance in EFL education. Language learning increasingly requires students not only to acquire vocabulary and grammar but also to interpret information, evaluate arguments, express positions, solve problems, and construct meaning in complex communicative situations. In this context, ChatGPT can potentially support both linguistic and cognitive development. Learners may use the system to generate alternative expressions or ideas, but they can also be asked to compare these alternatives, identify weaknesses, verify claims, and justify their final choices. The value of AI interaction therefore depends on whether the instructional task preserves the learner's role as an active cognitive participant.

Recent EFL research provides evidence for this conditional interpretation. Cong-Lem et al. [8] examined a structured Critical AI Engagement Cycle through a quasi-experimental pretest-posttest control-group design and found significant improvement in EFL learners' critical-thinking performance following a short intervention designed around critical engagement with generative AI. The study is particularly relevant because it demonstrates that access to ChatGPT was accompanied by explicit instructional scaffolding rather than unrestricted use. Likewise, Darwin et al. [9], in a qualitative study of Indonesian EFL students, found that learners recognised AI's usefulness for academic research and theory analysis while also identifying limitations concerning personalisation, contextual nuance, and possible dependence.

The emerging literature therefore suggests that the relationship between ChatGPT and HOTS is neither uniformly positive nor uniformly negative. A recent systematic review of 67 empirical studies by Li et al. [10] found that ChatGPT supported cognitive development most consistently when embedded in inquiry-oriented and scaffolded instructional designs involving processes such as metacognitive regulation, argumentative reasoning, and idea generation. At the same time, the review identified cognitive offloading and weaker critical engagement in more unstructured forms of use. This evidence strengthens the argument that instructional design functions as an important condition connecting ChatGPT use with cognitive outcomes.

Despite the growing body of research, an important conceptual gap remains. Much of the current literature examines whether ChatGPT improves learning outcomes, perceptions, or selected cognitive measures, but comparatively less attention is given to the mechanism through which instructional design transforms AI interaction into higher-order cognitive activity. In particular, the distinction between ChatGPT as a cognitive scaffold and ChatGPT as a form of cognitive substitution requires further clarification in language-learning contexts. The present paper therefore critically examines how ChatGPT may support or constrain HOTS in EFL education by focusing on the three higher-order cognitive processes of analysing, evaluating, and creating, while also considering cognitive offloading as a central pedagogical risk. Rather than treating ChatGPT as inherently beneficial or harmful, the paper argues that its educational value is conditional upon instructional design, learner engagement, and the distribution of cognitive responsibility between the learner and the technology. Recent systematic reviews likewise frame ChatGPT in education through both

opportunities and challenges, reinforcing the need to examine the conditions under which its use supports learning [13].

METHODS

Research Design

This study employed a qualitative critical literature review design to examine how ChatGPT may support or constrain higher-order thinking in EFL education. The study was not designed as a systematic review intended to identify and aggregate every available publication according to a registered protocol. Instead, it adopted a critical and interpretive approach that sought to synthesise relevant evidence, compare convergent and divergent findings, and examine the pedagogical conditions underlying the reported effects of ChatGPT. The review was guided by two central questions: how ChatGPT may support higher-order thinking in learning contexts and under what conditions its use may instead encourage cognitive offloading or reduce independent reasoning. The revised Bloom's taxonomy served as the main analytical framework, with particular attention to analysing, evaluating, and creating [7]. Cognitive offloading was used as a complementary concept for interpreting evidence related to over-reliance and the transfer of mental effort to external technological systems [4].

Data Sources and Search Strategy

The literature was identified through searches of academic databases and publisher platforms, including Scopus, Web of Science, Google Scholar, ScienceDirect, and Taylor & Francis Online. Search combinations included terms such as "ChatGPT," "generative AI," "artificial intelligence in education," "higher-order thinking," "critical thinking," "cognitive offloading," "AI-mediated learning," and "EFL." The search focused primarily on literature published from 2020 onward, with particular emphasis on research published after the emergence of ChatGPT in late 2022. Foundational literature published before this period was retained when it was necessary to establish the theoretical basis of the analysis, particularly research concerning Bloom's taxonomy and cognitive offloading. The review prioritised peer-reviewed empirical studies, systematic reviews, meta-analyses, and theoretically relevant publications that directly addressed the relationship between AI-supported learning and cognition. Particular attention was given to studies that examined higher-order thinking, critical thinking, cognitive engagement, language learning, pedagogical scaffolding, or over-reliance on AI. The review also incorporated recent evidence published in 2026 where it directly contributed to understanding the current state of research on ChatGPT and higher-order thinking.

Inclusion and Exclusion Criteria

Publications were included when they addressed generative AI or ChatGPT in relation to learning, higher-order thinking, critical thinking, cognitive engagement, language education, or cognitive offloading. Empirical studies were considered especially relevant when they provided information about research context, participant characteristics, instructional conditions, and measured or reported cognitive outcomes. Systematic reviews and meta-analyses were included when they synthesised evidence relevant to the central research questions. Publications that focused exclusively on technical aspects of AI development without educational implications were excluded. Studies that discussed ChatGPT only as a general technological phenomenon without addressing learning, cognition, or pedagogy were also excluded from the analytical synthesis. The review therefore prioritised relevance to the conceptual relationship between AI use, instructional design, and higher-order cognitive activity rather than relying solely on publication date or the visibility of a source.

Data Extraction and Analytical Procedure

The selected literature was examined with attention to research context, methodological design, participants or population where applicable, role of ChatGPT, cognitive outcomes, major findings, and pedagogical conditions or risks. These characteristics were organised into an evidence-synthesis matrix to facilitate comparison across studies. The analysis then proceeded thematically through the three higher-order cognitive processes identified in the revised Bloom's taxonomy. Studies were interpreted in relation to analysing when they involved examining information, comparing alternatives, identifying assumptions, or examining relationships among ideas. They were interpreted in relation to evaluating when learners were required to assess credibility, accuracy, relevance, evidence, or logical quality. Evidence was interpreted in relation to creating when ChatGPT was used to support brainstorming, synthesis, transformation, or the development of original ideas and products. The literature was also examined for recurring conditions that appeared to shape these outcomes, including instructional scaffolding, questioning, verification, reflective dialogue, learner autonomy, and teacher guidance. In parallel, evidence concerning cognitive offloading, over-reliance, passive acceptance, and substitution of learner effort was treated as a separate risk theme. Patterns of convergence and divergence were then interpreted to develop a broader explanation of when ChatGPT appears to function as a cognitive scaffold and when it may operate as a cognitive substitute.

Trustworthiness of the Review

The trustworthiness of the synthesis was strengthened by giving greater interpretive weight to peer-reviewed empirical studies, systematic reviews, and meta-analyses with clearly described research designs. Individual findings were interpreted in relation to their methodological characteristics and educational contexts rather than treated as universally generalisable evidence. The synthesis also retained both positive and cautionary evidence. This was important because the purpose of the review was not to demonstrate that ChatGPT is inherently beneficial but to understand the conditions under which it may support or constrain higher-order thinking. Claims were therefore framed cautiously, particularly when evidence reflected perceptions, associations, or task-specific outcomes rather than direct evidence of independently developed cognitive ability.

RESULTS AND DISCUSSION

Result

Evidence Synthesis

The reviewed literature reveals a consistent but conditional pattern regarding the relationship between ChatGPT and higher-order thinking. ChatGPT appears to provide meaningful opportunities for cognitive engagement when students are required to work critically with AI-generated information, while unstructured use creates a greater possibility that the system will replace rather than support learner reasoning. The findings also indicate that the quality of instructional design is more important than the mere presence of AI technology. This pattern is visible across experimental research, EFL studies, systematic reviews, and studies focusing specifically on cognitive offloading.

Table 1. Evidence Synthesis on ChatGPT and Higher-Order Thinking

Study	Context	Design	Role of ChatGPT	Main Cognitive Focus	Main Finding
Deng et al. [3]	Multiple educational contexts	Systematic review and meta-analysis of 69 experimental studies	AI-supported learning activities	Higher-order thinking propensity	ChatGPT interventions were associated with improved academic performance, affective-motivational outcomes, and higher-order thinking propensities, although methodological limitations remained
Barrot [2]	Language-learning contexts	Emerging technology report	Language practice, feedback, explanation, interaction	Linguistic and cognitive engagement	ChatGPT offers useful language-learning affordances but is best positioned as a complement to instruction
Zhai et al. [5]	Educational and research contexts	Systematic review of 14 studies	AI dialogue systems	Critical thinking, analytical reasoning, decision-making	Over-reliance on AI may weaken independent cognitive engagement and encourage acceptance of AI-generated recommendations without sufficient evaluation
Cong-Lem et al. [8]	EFL learners	Quasi-experimental pretest–posttest control-group study	Structured critical engagement with ChatGPT	Critical thinking	A short Critical AI Engagement Cycle improved EFL learners' critical-thinking performance
Darwin et al. [9]	Indonesian EFL master's students	Qualitative case study	AI-supported academic learning	Critical thinking	Students recognised AI's usefulness for research and theory analysis while also identifying risks related to personalisation, contextual understanding, and dependence
Al Farizi et al. [12]	Indonesian classroom context	Experimental study	ChatGPT-assisted instruction	HOTS	ChatGPT-supported instruction was associated with improved HOTS compared with the comparison condition
Li et al. [10]	University students	Systematic review of 67 empirical studies	ChatGPT in varied instructional settings	Critical and creative thinking	Cognitive benefits were more evident in inquiry-oriented and scaffolded settings, while unstructured use was associated with cognitive offloading

The synthesis indicates that ChatGPT is most consistently associated with positive cognitive outcomes when it is integrated into activities that require learners to process, question, verify, and transform the generated content. The evidence does not support the conclusion that ChatGPT itself automatically develops HOTS. Rather, the system appears to function as a resource that can either facilitate or reduce cognitive engagement depending on the instructional conditions in which it is used.

ChatGPT as a Cognitive Scaffold

Across the literature, ChatGPT emerges as potentially valuable when it provides learners with information or perspectives that become material for further reasoning. Its conversational format allows students to ask follow-up questions, request clarification, explore alternative explanations, and revisit earlier ideas. In

language education, this interaction can support both linguistic and cognitive engagement because students may practise language while simultaneously discussing concepts, developing arguments, and refining ideas. Barrot [2] highlights the interactive and personalised potential of ChatGPT for language learning, while Abdallah et al. [11] report that the technology is widely used for tutoring, drafting, clarification, and feedback in higher education.

The cognitive value of this interaction lies in what learners are required to do after receiving the AI-generated response. If students simply accept the response and move to the next task, the cognitive benefit may be limited. If they are required to compare the response with existing knowledge, verify its accuracy, identify weaknesses, or explain why they agree or disagree with it, the AI response becomes part of a meaningful cognitive process. ChatGPT can therefore be understood as a scaffold because it provides support while leaving the core responsibility for judgement and meaning-making with the learner.

The distinction is especially relevant to EFL settings because language learners may otherwise use ChatGPT primarily for surface-level purposes such as grammar correction or vocabulary lookup. These uses can be valuable, but their contribution to HOTS remains limited unless learners are also required to analyse the linguistic and conceptual implications of the generated output. When students compare alternative expressions, examine whether an argument is persuasive, or decide whether a revision changes the intended meaning, language activity becomes integrated with higher-order cognitive processing.

ChatGPT and the Analysing Dimension of HOTS

The analysing dimension of HOTS involves breaking information into components and examining relationships among those components [7]. ChatGPT can facilitate this process by generating alternative explanations, examples, arguments, or interpretations that students can compare. However, the production of multiple responses should not itself be interpreted as analysis. The analytical work occurs when learners identify similarities and differences, examine assumptions, determine relationships, and distinguish relevant from irrelevant information.

This distinction helps explain why ChatGPT can be cognitively useful without being inherently cognitively developmental. The technology can increase the quantity and variety of information available to students, but the learner must still perform the process of examining that information. When a teacher asks students to compare two ChatGPT-generated explanations and determine which better accounts for a phenomenon, the system supplies material for analysis. When students simply choose the first answer generated by ChatGPT, the cognitive demand is substantially reduced. The findings of Li et al. [10] reinforce this interpretation because the review found that inquiry-oriented and scaffolded uses of ChatGPT were more strongly associated with cognitive development. The implication is that analytical tasks should not stop at AI output generation. Students need explicit opportunities to examine the structure and quality of the output and to explain their reasoning.

ChatGPT and the Evaluating Dimension of HOTS

Evaluation requires learners to make judgements based on evidence, standards, or criteria [7]. This dimension is particularly significant in AI-supported learning because ChatGPT can produce responses that are linguistically polished and apparently authoritative without guaranteeing their accuracy or contextual appropriateness. Learners therefore need to evaluate both the information and the reasoning contained in AI-generated content. Evidence from Indonesian EFL learners supports this concern. Darwin et al. [9] found that participants considered AI useful for academic research and theory examination while also identifying limitations related to contextual nuance, personalisation, and the possibility of restricted perspectives. These findings suggest that AI-supported critical thinking requires learners to maintain a degree of epistemic distance from the generated response. Instead of asking whether the answer sounds convincing, students should be required to ask whether it is sufficiently supported, relevant to the context, and consistent with reliable evidence. The study by Cong-Lem et al. [8] provides stronger intervention-based support for this approach. Their Critical AI Engagement Cycle was designed to encourage EFL learners to interact critically with ChatGPT rather than accept its outputs passively. The intervention produced significant improvement in critical-thinking performance, suggesting that the act of evaluating AI-generated information can become a meaningful learning activity when it is deliberately incorporated into instructional design.

ChatGPT and the Creating Dimension of HOTS

The creating dimension of HOTS involves generating, organising, and transforming ideas into a new product or solution [7]. ChatGPT can contribute to this process by expanding the range of ideas available to learners. Brainstorming with ChatGPT, for instance, may help students overcome the initial difficulty of generating possibilities. However, the educational value of this process depends on whether students retain ownership over subsequent decisions. In EFL writing, ChatGPT may suggest several possible arguments, examples, or organisational structures. The learner can then select the most relevant ideas, combine them with

personal knowledge and external evidence, modify them to fit the intended audience, and construct an original final product. In such a process, the AI system contributes to ideational expansion while the learner remains responsible for creation. By contrast, when the learner submits an AI-generated product with little substantive modification, the system performs much of the creative labour that the task is intended to cultivate. The experimental evidence from Indonesia also demonstrates the potential of ChatGPT-supported instruction to enhance HOTS. Al Farizi et al. [12] reported significant improvement in HOTS in an experimental classroom context involving direct-current physics and Merdeka Belajar. Although the context differs from EFL education, the study supports the broader proposition that ChatGPT can contribute to higher-order cognitive outcomes when incorporated into structured teaching rather than treated simply as an unrestricted answer source.

Cognitive Offloading as a Major Risk

The positive potential of ChatGPT is accompanied by a substantial risk of cognitive offloading. Zhai et al. [5] found that over-reliance on AI dialogue systems can influence critical thinking, analytical reasoning, and decision-making because users may favour rapid AI-generated solutions over slower processes of independent evaluation. Gerlich [6] likewise reported an association between more frequent AI use, higher cognitive offloading, and lower critical-thinking performance. These findings do not imply that AI use is inherently detrimental, but they indicate that the distribution of cognitive effort between learners and technology matters. The distinction can be illustrated through writing. If students use ChatGPT to generate possible topics but are subsequently required to develop, verify, and defend their own arguments, AI assistance may reduce an unnecessary barrier while preserving the intended cognitive work. If students ask ChatGPT to generate the entire argument, organise the evidence, write the final text, and revise it without independent engagement, the technology performs much of the cognitive work that the writing task is intended to develop. The risk of cognitive offloading therefore reinforces rather than contradicts the argument for pedagogical integration. The appropriate response is not necessarily to prohibit AI but to design learning environments in which the cognitive responsibilities assigned to students remain explicit. Students can use AI as a resource while still being required to make the substantive analytical and evaluative decisions.

Discussion

Instructional Design as the Central Mechanism

The most important implication of the findings is that instructional design functions as the central mechanism linking ChatGPT use with higher-order thinking. The evidence does not support a straightforward relationship in which increased access to ChatGPT automatically produces improved cognitive abilities. Instead, positive outcomes appear more likely when students encounter instructional tasks that require them to perform meaningful cognitive operations while interacting with AI. This interpretation is consistent with the findings of Deng et al. [3], whose meta-analysis identified positive associations between ChatGPT-supported interventions and higher-order thinking propensities while also highlighting methodological limitations in the existing evidence base. It is also consistent with Li et al. [10], whose review of 67 empirical studies found stronger cognitive benefits when ChatGPT was embedded within inquiry-oriented and scaffolded learning designs. Together, these findings suggest that the pedagogical task is more consequential than the mere presence of the technology. The implication is that teachers should begin AI-supported lesson planning from the cognitive process they want students to develop rather than from the technological feature they want students to use. If the instructional objective is analysis, students should be required to examine AI-generated information. If the objective is evaluation, they should be required to verify evidence and justify judgements. If the objective is creation, they should be required to transform and defend the ideas they develop. ChatGPT then becomes part of a larger cognitive process rather than the destination of the task.

The Distribution of Cognitive Responsibility

The findings also suggest that the concept of learner responsibility is central to understanding why ChatGPT produces different outcomes across educational contexts. ChatGPT may provide cognitive assistance, but learners must retain responsibility for the decisions that define the quality and meaning of the final response. This principle helps distinguish productive AI use from excessive dependence. For EFL learning, this issue becomes particularly important because ChatGPT can influence both linguistic and conceptual dimensions of student work. Students may rely on the system to generate grammatically accurate sentences while simultaneously allowing it to determine what ideas should be expressed. In this situation, the linguistic quality of the final product may increase without a corresponding increase in students' ability to formulate and defend ideas independently. Educationally meaningful AI use therefore requires learners to remain involved in the conceptual work of the task even when some linguistic support is delegated to the system. This perspective also provides a useful explanation for apparently contradictory findings in the literature. Studies reporting positive outcomes may involve structured tasks in which learners remain cognitively active, whereas studies identifying negative effects may involve more substantial delegation of

cognitive responsibility. The difference may therefore lie less in the technology than in how cognitive labour is distributed between humans and AI.

Cognitive Scaffolding Versus Cognitive Substitution

The distinction between cognitive scaffolding and cognitive substitution offers a useful way to integrate the positive and negative findings of the literature. Scaffolding occurs when AI provides support that enables learners to perform a task while preserving the higher-order cognitive processes that the task is intended to develop. Substitution occurs when AI performs those processes instead of the learner. This distinction is compatible with the concept of cognitive offloading described by Risko and Gilbert [4]. External support becomes educationally productive when it reduces unnecessary difficulty without eliminating meaningful cognitive work. Conversely, offloading becomes problematic when learners transfer precisely those cognitive activities that they need to practise. The distinction is particularly important because the same ChatGPT function can be pedagogically productive or problematic depending on the task. Idea generation can support creativity when learners critically select and transform AI suggestions, but it can undermine creativity when the AI-generated ideas are accepted without substantial learner contribution. Automated language revision can support writing when students analyse why a revision improves clarity, but it may reduce learning when the revised text is simply adopted without reflection. Consequently, the educational challenge is not to determine whether ChatGPT should be used but to determine which parts of a cognitive task should remain human responsibilities. This provides a more nuanced basis for AI integration than either unrestricted adoption or complete prohibition.

Implications for EFL Pedagogy

The findings have direct implications for EFL pedagogy. ChatGPT can be integrated into language learning activities in ways that connect language development with higher-order cognitive objectives. Rather than limiting AI use to grammar correction, vocabulary explanation, or sentence generation, teachers can design tasks in which students examine AI-generated texts, compare alternative formulations, evaluate arguments, verify information, and revise their own responses on the basis of explicit criteria. Writing instruction provides a particularly useful context for this approach. Students can begin with their own ideas, use ChatGPT to generate alternative perspectives, and then evaluate those alternatives against the purpose and audience of the assignment. They can subsequently explain which suggestions they accepted, which they rejected, and why. Such a procedure preserves learner agency while still allowing AI to provide useful cognitive and linguistic support. AI-supported reading and discussion can be organised in a similar way. Students can ask ChatGPT to provide alternative interpretations of a text and then compare those interpretations with textual evidence. They can identify unsupported assumptions, locate potential inaccuracies, and formulate their own conclusions. These activities combine language practice with the cognitive processes of analysis and evaluation. The role of assessment is equally important. When students are evaluated only on the quality of a final product, it becomes difficult to determine how much of the cognitive process was performed independently. Assessment can therefore include reflective commentary, oral justification, process-based writing, source verification, and explanation of revisions. Such approaches provide evidence of how learners arrived at their answers rather than simply examining the final answer itself. Teacher readiness is another important factor. Educators need sufficient AI literacy to recognise the strengths and limitations of generative AI and to design tasks that preserve student agency. The emphasis should be placed less on monitoring whether students use AI and more on ensuring that their use of AI remains consistent with the cognitive objectives of instruction.

A Conditional Model of AI-Supported HOTS

The combined evidence supports a conditional model in which ChatGPT's contribution to HOTS is mediated by instructional design and realised through learner cognitive engagement. The first component of this model is the technological affordance provided by ChatGPT, including dialogue, explanation, feedback, brainstorming, and content generation. These affordances create possibilities but do not determine educational outcomes. The second component is instructional design, which determines how learners are expected to use those affordances. Structured questioning, comparison, evidence verification, justification, revision, and reflection encourage students to engage actively with AI-generated information. The third component is cognitive engagement, through which learners perform analysing, evaluating, and creating. When these conditions are present, ChatGPT can function as a scaffold for higher-order thinking. A contrasting pathway emerges when AI is used without sufficient structure. When learners seek complete answers, accept AI-generated claims without scrutiny, or transfer the central reasoning process to the system, cognitive offloading becomes more likely. The resulting relationship can therefore be represented as a conditional model rather than a direct causal relationship.

ChatGPT affordances → instructional design → learner cognitive engagement → HOTS
Unstructured AI use → cognitive offloading → reduced independent reasoning

The significance of this model lies in its ability to reconcile the positive and negative findings in the literature. It does not position ChatGPT as either inherently beneficial or inherently harmful. Instead, it identifies pedagogy as the condition that determines whether technological support becomes a resource for thinking or a substitute for thinking.

CONCLUSION

This critical literature review examined the relationship between ChatGPT and higher-order thinking in EFL education. The evidence indicates that ChatGPT can provide meaningful support for the cognitive processes of analysing, evaluating, and creating when learners interact with AI-generated information through structured and purposeful learning activities. The technology can provide explanations, alternative perspectives, examples, feedback, and preliminary ideas that create opportunities for students to examine information and develop their own responses. At the same time, the review demonstrates that the positive contribution of ChatGPT should not be treated as automatic. The strongest evidence points to the importance of instructional scaffolding, explicit cognitive demands, learner reflection, and verification of AI-generated content. When students are required to question, compare, verify, justify, and transform AI-generated information, ChatGPT can function as a cognitive scaffold. When students use it primarily to obtain complete answers or delegate essential reasoning, the same technology may facilitate cognitive offloading and reduce independent cognitive engagement.

The central contribution of this review is therefore a conditional interpretation of ChatGPT and HOTS. Rather than asking whether ChatGPT is inherently good or bad for higher-order thinking, the more meaningful question concerns the instructional conditions under which its affordances are translated into cognitive development. In EFL education, this perspective suggests that AI-supported language learning should preserve learners' responsibility for analysis, evaluation, and creation while allowing technology to provide support where it reduces unnecessary barriers. The theoretical implication is that AI-supported learning should not be conceptualised as a direct relationship between technological use and learning outcomes. Instructional design and learner engagement constitute important mediating conditions. The pedagogical implication is that teachers should design AI-supported tasks around explicit cognitive purposes and develop assessment procedures that reveal students' reasoning rather than simply measuring the quality of AI-assisted products.

Several limitations should nevertheless be acknowledged. This study adopted a critical literature review rather than a systematic review with a registered protocol and exhaustive screening procedure, meaning that the selected literature does not necessarily represent all available evidence. The studies included in the synthesis also vary in educational context, participant characteristics, research design, duration, implementation, and outcome measures, which limits direct comparison. Furthermore, research on generative AI remains relatively recent, and long-term evidence concerning whether cognitive gains associated with structured AI use are sustained after scaffolding is removed remains limited. Some studies also rely on self-reported perceptions or task-specific outcomes rather than direct measures of independent cognitive performance.

Future research should therefore examine the mechanisms through which different instructional designs shape AI-supported cognition. Experimental and quasi-experimental studies could compare structured and unstructured ChatGPT use while controlling for instructional content, teacher involvement, and duration. Research should also employ assessments that require learners to demonstrate independent reasoning without unrestricted AI assistance. In EFL contexts, further studies could examine whether critical AI engagement influences writing quality, language proficiency, metacognitive awareness, and HOTS simultaneously. Longitudinal research would also be valuable in determining whether repeated interaction with ChatGPT produces durable cognitive development or gradually increases dependence on external technological support. Finally, research in Indonesian educational settings could provide valuable evidence regarding how curriculum, teacher readiness, digital infrastructure, and AI literacy shape the relationship between ChatGPT and higher-order thinking.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the English Education Studies Postgraduate Program, UIN Sunan Gunung Djati Bandung, in the preparation of this manuscript.

REFERENCES

- [1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education—Where are the educators?,” *International Journal of Educational Technology in Higher Education*, vol. 16, p. 39, 2019. doi: 10.1186/s41239-019-0171-0.
- [2] J. S. Barrot, “ChatGPT as a language learning tool: An emerging technology report,” *Technology, Knowledge and Learning*, vol. 29, no. 2, pp. 1151–1156, 2024. doi: 10.1007/s10758-023-09711-4.
- [3] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, “Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies,” *Computers & Education*, vol. 227, p. 105224, 2025. doi: 10.1016/j.compedu.2024.105224.
- [4] E. F. Risko and S. J. Gilbert, “Cognitive offloading,” *Trends in Cognitive Sciences*, vol. 20, no. 9, pp. 676–688, 2016.
- [5] C. Zhai, S. Wibowo, and L. D. Li, “The effects of over-reliance on AI dialogue systems on students’ cognitive abilities: A systematic review,” *Smart Learning Environments*, vol. 11, p. 28, 2024. doi: 10.1186/s40561-024-00316-7.
- [6] M. Gerlich, “AI tools in society: Impacts on cognitive offloading and the future of critical thinking,” *Societies*, vol. 15, no. 1, p. 6, 2025. doi: 10.3390/soc15010006.
- [7] D. R. Krathwohl, “A revision of Bloom’s taxonomy: An overview,” *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, 2002. doi: 10.1207/S15430421TIP4104_2.
- [8] N. Cong-Lem, T. T. Nguyen, and K. N. H. Nguyen, “Critical thinking in the age of generative AI: Effects of a short-term experiential learning intervention on EFL learners,” *International Journal of TESOL Studies*, pp. 1–21, 2025. doi: 10.58304/ijts.250522.
- [9] D. Darwin, D. Rusdin, N. Mukminatien, N. Suryati, E. D. Laksmi, and Marzuki, “Critical thinking in the AI era: An exploration of EFL students’ perceptions, benefits, and limitations,” *Cogent Education*, vol. 11, no. 1, p. 2290342, 2024. doi: 10.1080/2331186X.2023.2290342.
- [10] C. Li, H. Cui, and L. S. Hagedorn, “The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes,” *Computers and Education: Artificial Intelligence*, vol. 10, p. 100571, 2026. doi: 10.1016/j.caeai.2026.100571.
- [11] N. Abdallah, R. Katmah, K. Khalaf, and H. F. Jelinek, “Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration,” *Social Sciences & Humanities Open*, vol. 12, p. 101866, 2025. doi: 10.1016/j.ssaho.2025.101866.
- [12] Z. Al Farizi, S. Sahlan, E. Prasetyo, N. Fitriani, E. M. Laleng, H. H. Tiar, and Z. Anwar, “Analysis of the impact of ChatGPT usage in direct current physics material on enhancing students’ higher-order thinking skills (HOTS) during the Merdeka Belajar,” *Berkala Ilmiah Pendidikan Fisika*, vol. 12, no. 1, pp. 92–105, 2024.
- [13] Y. Y. Munaye, W. Admass, Y. Belayneh, A. Molla, and M. Asmare, “ChatGPT in education: A systematic review on opportunities, challenges, and future directions,” *Algorithms*, vol. 18, no. 6, p. 352, 2025. doi: 10.3390/a18060352.