



Exploring ChatGPT's Impact on Critical, Creative, and Reflective Thinking Skills: A Mixed-Methods Study in an Indonesian EFL Classroom

Winia Waziana ^{1*} , Widi Andewi ² , Damar Wibisono ³ ,
Tommy Hastomo ⁴ , Muhamad Muslihudin ⁵ 

^{1,2,5} Information System, Institut Bakti Nusantara, Lampung, Indonesia

³ Sociology, Universitas Lampung, Lampung, Indonesia

⁴ English Department, STKIP PGRI Bandar Lampung, Lampung, Indonesia

Received: 2025/07/09

Accepted: 2025/08/30

Abstract: The integration of generative AI like ChatGPT into EFL pedagogy presents both opportunities for fostering higher-order thinking skills and risks to academic integrity. A research gap exists regarding the simultaneous impact of ChatGPT on the crucial triad of critical thinking, creativity, and self-reflection within the Indonesian EFL context. This study aimed to fill this gap by quantitatively measuring the effect of ChatGPT on these skills and qualitatively exploring students' perceptions of the learning process. The study used a mixed-methods sequential explanatory design. Participants were 100 undergraduate students, randomly assigned to either an experimental group (n=50) or a control group (n=50). Data were collected using three validated instruments: the Critical Thinking Scale, the Creative Thinking Scale, and the Reflective Thinking Scale. In addition, a semi-structured interview guide was used to obtain qualitative data. Quantitative data were analyzed using ANCOVA, and qualitative data using thematic analysis. The findings revealed that the ChatGPT group achieved statistically significant gains in critical, creative, and reflective thinking scores compared to the control group. Qualitative results revealed a duality in student perceptions; they valued the AI for fostering skills through idea generation and safe practice, but expressed concerns about risks such as cognitive offloading and skill atrophy.

Keywords: ChatGPT, EFL, Higher-Order Thinking Skills, Pedagogical Integration, Student Perceptions.

* Corresponding Author.

Authors' Email Address:

¹ Winia Waziana (wwaziana@gmail.com), ² Widi Andewi (widiandewi.91@gmail.com), ³ Damar Wibisono (damar.wibisono@fisip.unila.ac.id), ⁴ Tommy Hastomo (tomhas182@gmail.com), ⁵ Muhamad Muslihudin (muslihudinstmikpsw@gmail.com)



Introduction

The swift current of global technological evolution presents the field of education with novel challenges and opportunities, particularly in the integration of Artificial Intelligence (AI) into pedagogical practices. The emergence of ChatGPT exemplifies this development. It is a generative language model that utilizes AI to produce text stylistically similar to human composition (Lund & Wang, 2023). In the pedagogical context of English as a Foreign Language (EFL), scholars view ChatGPT as a promising tool. It holds the potential to foster key higher-order thinking skills (HOTS) such as critical thinking, creativity, and reflective thinking (Borge et al., 2024; Gerlich, 2025; Lee & Low, 2024; Walter, 2024). There remains a significant imperative to critically assess the actual efficacy and consequences of using these skills in Indonesia's English language learning environments.

Critical thinking stands as a crucial competency for 21st-century education. It is also a fundamental prerequisite for shaping learners into autonomous individuals capable of rational decision-making across diverse situations (Ibrahim et al., 2024; Janse van Rensburg, 2024; Thornhill-Miller et al., 2023). On another front, creativity provides the cornerstone for developing novel concepts and navigating complex problems (Lin & Chen, 2024; Rong et al., 2022). Reflective thinking, in turn, empowers students to assess their learning journey and its results, which facilitates the sustained enhancement of their academic performance (Combrinck & Loubser, 2025; Jelodari et al., 2023). In light of the profound importance of these skills, educators and researchers have a critical responsibility to investigate the role that AI technologies, such as ChatGPT, play in either fostering or inhibiting the development of these very abilities.

Moreover, the rapid evolution of ChatGPT following its debut in late 2022 has triggered a widespread debate within academic circles and the public sphere. Proponents view it as a transformative instrument capable of accelerating learning, expanding informational access, and enhancing educational experiences for students (Song & Song, 2023; Wang et al., 2023). Conversely, critics express concerns regarding its potential to diminish original thought, foster technological dependence, and create ethical dilemmas related to academic integrity, including plagiarism and the authenticity of learning outcomes (Algaraady & Mahyoob, 2025; Mohamed, 2024). This ongoing discourse highlights the profound relevance of conducting empirical research that examines the tangible impacts of ChatGPT on cognitive skills, such as critical thinking, creativity, and reflective thinking, in the EFL context.

In practice, employing ChatGPT within EFL instructional settings shows promise for creating learning experiences that are more personal, interactive, and adaptive to individual student needs. For instance, [Waziana et al. \(2024\)](#) concluded that the tool can bolster academic writing skills, offer instantaneous feedback, and improve student motivation. Furthermore, several studies indicate that ChatGPT can foster the development of critical thinking when used in conjunction with problem-based prompts and reflective exercises ([Janse van Rensburg, 2024](#); [Tseng & Lin, 2024](#)). Despite these positive findings, a significant caveat has emerged from other research. Studies indicate that creativity may not be significantly enhanced by using ChatGPT alone, as interactions with the AI are frequently reactive and structured, which limits opportunities for genuinely exploratory thinking ([Nugroho et al., 2025](#); [Wang & Fan, 2025](#)).

A similar concern arises with reflective thinking, a critical part of self-directed learning. This skill is frequently overlooked when educational technologies prioritize efficiency and immediate results over deliberate thought. For example, research by [Ibrahim et al. \(2024\)](#) demonstrates that while AI-based instruction can enhance reflective thinking and academic resilience, its effectiveness depends on an instructional design that fosters students' internal dialogue and is structured around constructive feedback. The articulate nature of AI-generated responses may therefore lead learners to accept answers at face value and bypass the crucial step of internal validation ([Chiu, 2024](#)). In this context, it is necessary to analyze whether ChatGPT is genuinely capable of prompting deep reflection or if it merely creates an illusion of superficial understanding.

With the increasing prevalence of AI in education, especially within the Indonesian context, understanding the student perspective on ChatGPT is essential. Research indicates that many students embrace the tool. They appreciate it as a constantly available study partner that offers immediate feedback ([Hastomo et al., 2025](#); [Marzuki et al., 2023](#); [Oktarin et al., 2024](#); [Waziana et al., 2024](#); [Wulyani et al., 2024](#)). However, this positive reception is tempered by significant concerns about the potential erosion of autonomous thought, reduced human-to-human interaction, and the blurred line between helpful support and detrimental dependency ([Luther et al., 2024](#)). For this reason, a thorough examination of student attitudes and perceptions toward ChatGPT as an English learning tool represents a significant and necessary area of inquiry.

Existing literature suggests that while ChatGPT can support higher-order skills like critical thinking, this effectiveness often requires significant pedagogical scaffolding from instructors ([Borge et al., 2024](#); [Combrinck & Loubser, 2025](#); [Ibrahim et al., 2024](#); [Janse van](#)

Rensburg, 2024; Jelodari et al., 2023; Lee & Low, 2024; Rong et al., 2022; Wang & Fan, 2025). Furthermore, its utility in fostering creativity and deep reflective thinking remains a subject of debate. Studies indicate that structured AI interactions may not be sufficient for promoting original thought or profound introspection (Baskara, 2023; Wang, 2024). This creates an ambivalent picture for educators, who recognize the tool's potential but also harbor valid concerns about its pedagogical risks (Wang et al., 2023). A significant research gap therefore emerges from the literature. While many studies address one of these skills in isolation, there is a notable scarcity of empirical research that holistically and simultaneously investigates the combined impact of ChatGPT on students' critical thinking, creativity, and reflective thinking within the Indonesian EFL context. Addressing this gap is crucial, as understanding the multifaceted influence of AI on higher-order thinking can inform evidence-based instructional design in EFL settings. This study is significant because it provides integrated insights into how generative AI, when used with pedagogical intention, can enhance students' cognitive development in ways that are both measurable and meaningful for EFL education in Indonesia. To address this gap, this study proposes a mixed-methods approach guided by the following four research questions:

1. Is there a significant difference in the critical thinking scores of university students who use ChatGPT in English language learning compared to those taught using traditional methods?
2. Is there a significant difference in the creative thinking scores between students using ChatGPT and those learning through conventional methods in the context of English language acquisition?
3. How do the reflective thinking skill scores of students who utilize ChatGPT compare to those of students in a traditional learning setting?
4. From the students' perspective, what specific interactions with ChatGPT are perceived as fostering or hindering the development of their critical thinking, creativity, and reflective thinking?

Literature Review

Generative AI in English Language Teaching

The advent of Generative Artificial Intelligence (GenAI), particularly large language models like ChatGPT, represents a significant paradigm shift in English Language Teaching (ELT). These technologies offer unprecedented opportunities for personalized learning experiences

that cater to the individual needs of each student. For instance, [Rasul et al.'s \(2023\)](#) research suggests that AI tutors can adjust content complexity and instructional pacing, thereby creating a more efficient and motivating learning path for each student. GenAI also functions as a tireless conversational partner. This provides a low-anxiety environment for learners to practice their speaking and interactional skills outside the traditional classroom ([Tu, 2020](#)). Furthermore, the capacity of these tools to provide instantaneous grammatical and lexical feedback is a frequently cited advantage in recent literature.

In terms of specific skill development, the application of GenAI in fostering writing proficiency is extensively documented. AI tools can assist learners in brainstorming ideas, structuring arguments, and refining drafts through iterative feedback ([Marzuki et al., 2023](#); [Waziana et al., 2024](#)). Beyond writing, these platforms have proven effective for vocabulary acquisition. They can generate contextualized sentences and create personalized vocabulary quizzes, which enhance retention and understanding of new lexical items ([Slamet, 2024](#)). Recent studies have also explored its use as a pronunciation tutor. Research by [Yang et al. \(2022\)](#) found that AI-powered feedback on phonetics and intonation resulted in significant improvements in students' oral fluency and accuracy.

Despite these documented benefits, the integration of GenAI presents challenges. Much of the current discourse focuses on practical issues such as the factual accuracy of AI outputs and the potential for academic dishonesty ([Elkhatat et al., 2023](#); [Perkins, 2023](#); [Yusuf et al., 2024](#)). While these concerns are valid, a more nuanced research gap exists regarding the precise cognitive impact on learners. The existing literature tends to examine language proficiency or discrete cognitive abilities, such as critical thinking, in isolation. Consequently, there is a scarcity of empirical research that holistically investigates the simultaneous effects of ChatGPT on the crucial triad of critical thinking, creativity, and reflective thinking. It is this multifaceted gap in the literature that the present study aims to fill.

ChatGPT and Higher-Order Thinking Skills

The relationship between the use of ChatGPT and the development of critical thinking is a prominent theme in recent educational technology literature. Several studies suggest that ChatGPT can serve as a Socratic partner when guided by well-designed prompts and inquiry-based tasks. For example, [Janse van Rensburg \(2024\)](#) demonstrated that problem-based learning scenarios facilitated by ChatGPT prompted students to analyze arguments and evaluate different sources of evidence. However, the effectiveness of AI in fostering critical

thought is highly dependent on pedagogical intervention. Without explicit instruction from a teacher, students may simply accept AI-generated text as an authoritative source of facts (Nguyen & Tran, 2023). This indicates that ChatGPT is not an autonomous developer of critical thinking but rather a tool whose utility is mediated by the instructor's design.

In contrast to the role of ChatGPT in fostering critical thinking, the literature presents a more contentious picture regarding its impact on creative thinking. Most research indicates that conventional use of the tool does not lead to significant gains in originality or divergent thinking among students. An experimental study by Wang and Fan (2025) found no statistical difference in the creativity scores of students using ChatGPT for idea generation compared to a control group without AI assistance. The structured and pattern-based nature of the AI's output often encourages convergent thinking rather than genuine ideational exploration. Some scholars argue that creativity can be stimulated if the tool is used specifically for brainstorming unconventional perspectives as a starting point (Kartal, 2024; Rong et al., 2022). Nonetheless, the consensus suggests that fostering creativity with AI requires specialized tasks that push beyond simple question-and-answer interactions.

The development of reflective thinking through AI interaction is the least explored of the three higher-order thinking skills. Some research suggests that AI can prompt reflective processes if the instructional design explicitly requires students to evaluate their learning journey (Ibrahim et al., 2024; Jelodari et al., 2023). However, students often use the tool for immediate task completion and problem-solving, thereby bypassing valuable opportunities for metacognitive reflection on their learning process. This review of the literature reveals a clear pattern across all three domains. Scholarly work tends to investigate these cognitive skills—critical thinking, creativity, and reflective thinking—in isolation from one another. Consequently, a significant research gap exists regarding the simultaneous and interrelated impact of ChatGPT on this triad of skills, a holistic perspective that the present study aims to provide.

Student Perceptions of Learning with ChatGPT

Research on student perceptions of ChatGPT in language learning consistently reveals a predominantly positive attitude towards its practical utility. Students frequently praise the tool's 24/7 availability as a key advantage over traditional human resources, such as tutors or teachers (Nguyen & Barbieri, 2025; Waziana et al., 2024). The speed at which it provides answers and generates text is also highly valued for completing assignments efficiently. Furthermore, many learners perceive the AI as a non-judgmental learning companion. This

perception reduces communication anxiety and encourages students to ask basic questions they might be hesitant to pose to a human instructor (Cain, 2024; Mutanga et al., 2025). In the Indonesian context, Slamet (2024) found that students particularly appreciate its role in explaining complex English grammatical concepts in a simplified and accessible manner.

However, this positive reception is accompanied by a growing set of concerns expressed by the students themselves. A primary fear is the potential for over-reliance on the technology for academic tasks. Students worry that frequent use might lead to the atrophy of their own independent thinking and original writing skills (Athanasopoulos et al., 2023). The absence of genuine human interaction and empathetic feedback is another frequently cited drawback in perception studies. Learners are also becoming increasingly aware of the tool's potential to produce inaccurate or nonsensical information, known as 'hallucinations' (Alkaissi & McFarlane, 2023). This creates an internal conflict for students who often struggle to differentiate between using the tool as a legitimate aid and engaging in academic misconduct.

Beyond general attitudes, a deeper analysis reveals different student perceptions regarding ChatGPT's role in the cognitive learning process itself. Some students view the tool primarily as an efficient shortcut for task completion, a way to generate answers with minimal cognitive effort (Slamet, 2024; Wang & Fan, 2025). Conversely, other learners perceive it as a collaborative partner for brainstorming and exploring complex ideas (Van Horn, 2024; Widianingtyas et al., 2023). This distinction in perception is crucial, yet most existing studies stop at documenting general attitudes of favorability or concern. There is a clear lack of research that specifically investigates how students perceive the influence of their interactions with ChatGPT on the specific development of their higher-order thinking skills. Therefore, understanding the perceived mechanisms through which ChatGPT fosters or hinders the triad of critical thinking, creativity, and reflective thinking from the students' viewpoint remains a significant research gap that this study addresses.

Research Methods

Research Design

This study employed a mixed-methods sequential explanatory design, following the framework established by Creswell and Clark (2017). This model was selected because it allows for a comprehensive investigation of the research problem. It involves collecting and analyzing quantitative data first, followed by the collection and analysis of qualitative data to help explain, elaborate on, and enrich the initial quantitative findings. The initial quantitative

phase consisted of a pretest-posttest quasi-experimental design to compare the effects of two instructional conditions on students' higher-order thinking skills. The independent variable was the instructional method (ChatGPT-assisted learning vs. traditional lecture-based learning), while the dependent variables were the students' scores in critical thinking, creative thinking, and reflective thinking. The subsequent qualitative phase, involving semi-structured interviews, was designed to provide deeper insights into the student experience and the perceived mechanisms behind the quantitative results. By integrating quantitative scores with qualitative narratives, this design facilitates a robust methodological triangulation. This approach enhances the validity and credibility of the findings by using the qualitative data to explain how and why the observed outcomes occurred (Creswell & Clark, 2017).

Participants

The study's cohort consisted of 100 fourth-semester undergraduate students majoring in Information Systems at Institut Bakti Nusantara during the 2024–2025 academic year. This population was selected for its unique suitability, as it combines an inherent familiarity with technology with a diverse spectrum of English language competencies. Participants were recruited using a purposive sampling method based on several predefined criteria to ensure the integrity of the sample for the intervention. Key requirements for inclusion were confirmed enrollment in the mandatory English course, voluntary provision of informed consent, consistent access to technology, and a lack of substantial prior experience using AI for academic English tasks, which was verified through a screening questionnaire. The final sample of 100 students demonstrated academic homogeneity, as indicated by their mean Cumulative Weighted Average of 68.67 (SD = 6.81). Following the pretest, participants were assigned to either the experimental group (n = 50) for ChatGPT-assisted instruction or the control group (n = 50) for traditional instruction.

Instrument

Data for this study were collected using three validated quantitative scales and a semi-structured interview guide. The construct validity of all quantitative instruments was established through Confirmatory Factor Analysis (CFA) prior to their use in the main study (Harrington, 2009). The analysis confirmed a good model fit for each scale, with all models meeting established thresholds for academic research (RMSEA < 0.08; CFI > 0.90).

The first instrument, the Critical Thinking Scale (CTS), was employed to measure changes in students' critical thinking abilities. This 20-item instrument, adapted from Sosu's

(2013) work, assesses two key dimensions: 'analysis,' which evaluates the ability to deconstruct arguments, and 'reflective skepticism,' which measures the tendency to question assumptions and seek evidence. Participants responded to each item on a 5-point Likert scale. The scale demonstrated excellent internal consistency, yielding a Cronbach's alpha coefficient of 0.93.

To assess creative thinking, the study utilized the Creative Thinking Scale, an instrument adapted from Hidayat et al. (2018). This 20-item scale is designed to evaluate multiple facets of a creative disposition. It includes subscales for 'innovative thinking' to gauge originality, 'intellectual courage' to measure willingness to take risks with ideas, and 'flexibility' to assess the ability to consider diverse perspectives. The reliability analysis for this instrument indicated strong internal consistency, with a Cronbach's alpha of 0.89.

Students' reflective thinking skills were measured using the Reflective Thinking Scale (RTS), a 20-item instrument based on the model developed by Basol and Gencel (2013). The scale is structured to measure a student's progression through different levels of reflection. These levels range from 'habitual action' (thoughtless repetition) and 'understanding' to the higher levels of 'reflection' and 'critical reflection,' which involve a change in perspective. This instrument displayed outstanding reliability, as confirmed by a Cronbach's alpha coefficient of 0.96.

For the qualitative phase of the study, a semi-structured interview guide was developed to explore student perceptions in depth. Adapted from the protocol used by Xiao and Zhi (2023), the guide was designed to elicit rich, detailed narratives about the students' learning experiences with ChatGPT. It contained open-ended questions and prompts that encouraged participants to discuss the perceived benefits, challenges, and specific interactions they believed fostered or hindered their learning processes. This qualitative tool was essential for providing explanatory context to the quantitative findings gathered from the scales.

Data Collection

The data collection for this study was conducted meticulously in three distinct phases over 12 weeks. Each phase was designed to systematically gather the quantitative and qualitative data required to address the research questions. The initial phase of the procedure involved administering a pretest. This step was designed to establish a comprehensive baseline of the participants' existing higher-order thinking skills prior to the intervention commencing. All 100 enrolled participants were gathered in a controlled setting to complete the three quantitative scales: the Critical Thinking Scale, the Creative Thinking Scale, and the

Reflective Thinking Scale. The data gathered from this pretest served as a crucial covariate in the final statistical analysis, controlling for pre-existing differences between the groups.

Following the pretest, the 12-week instructional intervention was implemented. Both groups received two hours of English language instruction per week on identical topics. The experimental group ($n = 50$) began with a one-hour orientation session that trained them on the functionalities of ChatGPT and the ethical considerations for its academic use. Their weekly learning activities were centered on using the AI as a cognitive tool for tasks such as idea generation, text revision, and complex inquiry. These activities were always followed by instructor-facilitated discussions to evaluate the AI's output critically. In parallel, the control group ($n = 50$) was taught using conventional pedagogical methods, which included teacher-led lectures, textbook-based assignments, and standard classroom discussions.

The final phase of data collection occurred upon the conclusion of the 12-week intervention. All 100 participants from both groups undertook a posttest, which consisted of the same three quantitative instruments administered during the pretest to measure any changes in their cognitive skills. Subsequently, the qualitative data were gathered. A subset of 15 students from the experimental group was selected through purposive sampling based on their varied levels of engagement during the intervention to capture a wide range of perspectives. Each student participated in a semi-structured interview that lasted approximately 30 to 45 minutes. With prior consent, each session was audio-recorded to ensure accurate transcription and a thorough analysis of their experiences with the AI-assisted learning process.

Data Analysis

The quantitative data were analyzed using IBM SPSS Statistics, Version 26. To address the first three research questions concerning the differences in cognitive skill scores, an Analysis of Covariance (ANCOVA) was employed. This statistical procedure was specifically chosen for its ability to increase statistical power and reduce error variance. ANCOVA effectively isolated the true effect of the instructional intervention (ChatGPT use vs. traditional methods) by comparing the posttest scores of the two groups while statistically treating their pretest scores as a covariate. Prior to conducting the main analysis, all necessary assumptions for ANCOVA were rigorously tested. These included the normality of data distribution, homogeneity of variances (Levene's test), and the homogeneity of regression slopes. The threshold for statistical significance for all analyses was established at an alpha level of $p < 0.05$.

The qualitative data, derived from the verbatim transcripts of the semi-structured interviews, were analyzed using thematic analysis. This study adopted the rigorous six-phase procedural guide proposed by Braun and Clarke (2006), a framework selected for its systematic and flexible approach to identifying patterns within qualitative data. The analytical process began with data familiarization, where researchers repeatedly read the transcripts to immerse themselves in the content. This was followed by a systematic process of generating initial codes for interesting features across the entire dataset. These codes were then collated into potential themes, which were subsequently reviewed, refined, and clearly defined. To ensure the credibility and reliability of the analysis, two researchers independently coded the data. After an initial round of coding and discussion to refine the coding frame, the final inter-coder reliability was calculated, yielding a Cohen's Kappa coefficient of 0.87, which indicates substantial agreement. The finalized thematic structure was then used to interpret the data and provide rich, contextualized answers to the fourth research question.

Ethical Considerations

Ethical approval was obtained from the Research and Community Service Ethics Committee (LPPM) of Institut Bakti Nusantara. The protocol adhered to the principles of the Declaration of Helsinki. All participants provided written informed consent after being fully informed about the study's purpose, procedures, their right to withdraw at any time without penalty, and the measures in place to ensure data confidentiality and anonymity.

Results

RQ1: Is there a significant difference in the critical thinking scores of university students who use ChatGPT in English language learning compared to those taught using traditional methods?

To address the first research question regarding the impact on critical thinking, a pretest-posttest quasi-experimental design was used. An ANCOVA was conducted to compare the posttest critical thinking scores of the ChatGPT group and the Traditional group, while statistically controlling for their pretest scores.

Table 1. Descriptive Statistics for Critical Thinking Scores

Group	Time	Mean	Std. Deviation	N
ChatGPT	Pretest	75.8	6.3	50
	Posttest	85.4	5.2	50
Traditional	Pretest	76	6.5	50
	Posttest	76.2	6.1	50

Table 1 outlines the descriptive statistics for the pretest and posttest scores of the experimental (ChatGPT) group and the control (Traditional) group. Each group consisted of 50 participants ($N = 50$). At the pretest stage, both groups demonstrated a comparable baseline proficiency. The mean score for the ChatGPT group ($M = 75.8$, $SD = 6.3$) was nearly identical to that of the Traditional group ($M = 76.0$, $SD = 6.5$). Following the intervention period, the posttest results showed a substantial improvement in performance for the ChatGPT group. Their mean score increased significantly to 85.4 ($SD = 5.2$). Conversely, the Traditional group exhibited only a negligible change in their performance. Their posttest mean score remained almost unchanged at 76.2 ($SD = 6.1$). These findings strongly suggest that the ChatGPT-assisted intervention had a significant positive effect on participant scores, while the traditional instructional method did not result in any meaningful improvement.

Table 2. Tests of Between-Subjects Effects for Critical Thinking

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2450.75	2	1225.38	38.15	$p < .001$
Pretest Critical_Thinking	150.45	1	150.45	4.68	0.033
Group	2295.3	1	2295.3	71.46	$p < .001$
Error	3115.25	97	32.12		
Corrected Total	5566	99			

An ANCOVA test was conducted to evaluate the effect of the intervention on students' critical thinking scores. The analysis controlled for initial differences by using the pretest critical thinking scores as a covariate. The results of the ANCOVA are presented in Table 2. There was a statistically significant main effect for the instructional group, $F(1, 97) = 71.46$, $p < .001$. This result indicates a significant difference in posttest critical thinking scores between the experimental and control groups after adjusting for the pretest scores. Furthermore, the covariate (pretest critical thinking) was also a significant predictor of

posttest performance, $F(1, 97) = 4.68$, $p = .033$. However, the primary finding is the substantial effect of the group variable. This suggests the intervention itself was the principal factor responsible for the observed differences in critical thinking outcomes.

Table 3. Estimated Marginal Means for Posttest Critical Thinking

Group	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
ChatGPT	85.4	0.8	83.81	86.99
Traditional	76.2	0.8	74.61	77.79

Table 3 shows the estimated marginal means for posttest critical thinking scores with pretest differences controlled. The ChatGPT group obtained a higher mean score ($M = 85.4$, $SE = 0.8$) than the traditional group ($M = 76.2$, $SE = 0.8$). The 95% confidence interval for the ChatGPT group ranges from 83.81 to 86.99. The confidence interval for the traditional group ranges from 74.61 to 77.79. These results indicate that students in the ChatGPT group achieved better critical thinking performance than those in the traditional group.

RQ2: Is there a significant difference in the creative thinking scores between students using ChatGPT and those learning through conventional methods in the context of English language acquisition?

The second research question examined the difference in creative thinking scores between the two groups. The same pretest-posttest quasi-experimental design and ANCOVA statistical procedure were employed, with creative thinking scores as the dependent variable.

Table 4. Descriptive Statistics for Creative Thinking Scores

Group	Time	Mean	Std. Deviation	N
ChatGPT	Pretest	79.1	5.9	50
	Posttest	88.1	4.9	50
Traditional	Pretest	78.9	6	50
	Posttest	79.5	5.8	50

Table 4 presents the descriptive statistics for the creative thinking scores, detailing the performance of the experimental (ChatGPT) and control (Traditional) groups at the pretest and posttest stages. Each group consisted of 50 participants ($N = 50$). At the outset of the

study, both groups demonstrated nearly identical levels of creative thinking proficiency. The mean pretest score for the ChatGPT group ($M = 79.1$, $SD = 5.9$) was highly comparable to that of the Traditional group ($M = 78.9$, $SD = 6.0$). This confirms a uniform baseline between the two groups before the intervention. Following the intervention period, a notable divergence in performance was observed. The ChatGPT group exhibited a substantial increase in their creative thinking scores. Their mean score rose to 88.1 ($SD = 4.9$). In stark contrast, the Traditional group's scores remained relatively static. Their posttest mean score was 79.5 ($SD = 5.8$), indicating only minimal improvement from their baseline. These descriptive results strongly suggest that the intervention that utilized ChatGPT was highly effective in fostering creative thinking skills. Conversely, the traditional instructional method did not yield any meaningful development in this area.

Table 5. Tests of Between-Subjects Effects for Creative Thinking

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2180.5	2	1090.25	36.52	$p < .001$
Pretest_Creative_Thinking	125.6	1	125.6	4.21	0.043
Group	1849	1	1849	61.94	$p < .001$
Error	2895.5	97	29.85		
Corrected Total	5076	99			

An ANCOVA test was conducted to assess the impact of the instructional intervention on students' creative thinking scores. The pretest scores for creative thinking served as the covariate. This method statistically controlled for any initial differences in proficiency between the groups. Table 5 displays the results of this analysis. The findings reveal a highly significant main effect for the group variable, $F(1, 97) = 61.94$, $p < .001$. This result demonstrates that a significant difference existed in the posttest creative thinking scores between the experimental and control groups after accounting for their pretest performance. Additionally, the covariate itself, pretest creative thinking, was a statistically significant predictor of the posttest scores, $F(1, 97) = 4.21$, $p = .043$. Despite the influence of the baseline scores, the primary finding is the substantial effect of the instructional group. This suggests that the intervention was the primary factor driving the observed improvements in students' creative thinking skills.

Table 6. Estimated Marginal Means for Posttest Creative Thinking

Group	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
ChatGPT	88.1	0.77	86.57	89.63
Traditional	79.5	0.77	77.97	81.03

Table 6 presents the estimated marginal means for posttest creative thinking scores with pretest differences controlled. The ChatGPT group reached a higher mean score ($M = 88.1$, $SE = 0.77$) compared to the traditional group ($M = 79.5$, $SE = 0.77$). The 95% confidence interval for the ChatGPT group ranges from 86.57 to 89.63. The confidence interval for the traditional group ranges from 77.97 to 81.03. These findings show that students who learned with ChatGPT demonstrated stronger creative thinking abilities than those who learned through traditional instruction.

RQ3: How do the reflective thinking skill scores of students who utilize ChatGPT compare to those of students in a traditional learning setting?

The third research question compared the reflective thinking skill scores of students in the ChatGPT-assisted setting with those in the traditional setting. The analysis again utilized a pretest-posttest design with ANCOVA to control for initial differences.

Table 7. Descriptive Statistics for Reflective Thinking Scores

Group	Time	Mean	Std. Deviation	N
ChatGPT	Pretest	74.9	6.6	50
	Posttest	83.7	5.5	50
Traditional	Pretest	75.1	6.8	50
	Posttest	75.3	6.4	50

Table 7 presents the descriptive statistics for reflective thinking scores across two instructional approaches, namely ChatGPT-assisted learning and traditional methods. Both groups consisted of 50 participants each. The ChatGPT group demonstrated an increase in mean scores from 74.9 ($SD = 6.6$) in the pretest to 83.7 ($SD = 5.5$) in the posttest, indicating a substantial improvement in reflective thinking. In contrast, the traditional group showed minimal change, with a pretest mean of 75.1 ($SD = 6.8$) and a posttest mean of 75.3 ($SD = 6.4$). These findings suggest that while both groups began with relatively similar

baseline scores, the ChatGPT group exhibited notable gains in reflective thinking over time, whereas the traditional group remained nearly constant.

Table 8. Tests of Between-Subjects Effects for Reflective Thinking

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2050.18	2	1025.09	28.55	p < .001
Pretest_Reflective_Thinking	180.28	1	180.28	5.02	.027
Group	1764	1	1764.00	49.13	p < .001
Error	3482.82	97	35.91		
Corrected Total	5533	99			

Table 8 displays the results of the ANCOVA test conducted to examine between-subjects effects on reflective thinking. The analysis reveals that the overall model was statistically significant, $F(2, 97) = 28.55$, $p < .001$, indicating that the predictors collectively explained a meaningful portion of the variance in reflective thinking scores. The pretest scores of reflective thinking also showed a significant effect, $F(1, 97) = 5.02$, $p = .027$, suggesting that initial reflective thinking levels contributed to the posttest outcomes. More importantly, the effect of the group was highly significant, $F(1, 97) = 49.13$, $p < .001$, demonstrating that students in the ChatGPT condition significantly outperformed those in the traditional group when controlling for pretest scores. This finding highlights the substantial impact of ChatGPT-based learning on enhancing students' reflective thinking abilities.

Table 9. Estimated Marginal Means for Posttest Reflective Thinking

Group	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
ChatGPT	83.7	0.85	82.01	85.39
Traditional	75.3	0.85	73.61	76.99

Table 9 presents the estimated marginal means for posttest reflective thinking scores with pretest differences controlled. The ChatGPT group obtained a higher mean score ($M = 83.7$, $SE = 0.85$) than the traditional group ($M = 75.3$, $SE = 0.85$). The 95% confidence interval for the ChatGPT group ranges from 82.01 to 85.39. The confidence interval for the traditional group ranges from 73.61 to 76.99. These results show that students who used

ChatGPT achieved stronger reflective thinking abilities compared to those who learned through traditional instruction.

RQ4: From the students' perspective, what specific interactions with ChatGPT are perceived to foster or hinder the development of their critical thinking, creativity, and reflective thinking?

The fourth research question explored students' perspectives on their interactions with ChatGPT. The goal was to identify specific mechanisms that students perceived as either fostering or hindering their cognitive skills. Thematic analysis of qualitative data from interviews and reflections was conducted. This methodological approach aligns with similar recent studies investigating student perceptions of AI in education.

Table 10. Thematic Analysis of Student Perceptions on ChatGPT Interactions

Category	Sub-Theme	Description	Student Statement
Fostering Mechanisms	Idea Generation & Brainstorming	Students use ChatGPT to overcome mental blocks and generate initial ideas for writing and speaking tasks.	<i>"ChatGPT is a valuable aid in brainstorming. It gives prompts and suggestions that help me tackle mental blocks when I don't know where to start."</i>
	Perspective Broadening & Divergent Thinking	The tool exposes students to multiple viewpoints and alternative solutions, encouraging them to think beyond their inherent patterns of thought.	<i>"AI exposed me to several different viewpoints. This guidance helped me break out of my usual thinking and discover more innovative possibilities for my essay."</i>
	Argument Refinement & Convergent Thinking	ChatGPT assists in evaluating the strengths and weaknesses of different arguments, helping students to focus and structure their thoughts logically.	<i>"When I had several potential arguments, AI helped me weigh the pros and cons of each. It even suggested which one had the strongest evidence base, making my writing more focused and persuasive."</i>

Category	Sub-Theme	Description	Student Statement
	Language Scaffolding & Lowering the Affective Filter	The tool provides a non-judgmental space for language practice, offering instant feedback on grammar and vocabulary, which reduces anxiety.	<i>"It's a safe space to practice without the fear of judgment. I can try out new sentences and get immediate feedback, which makes me more confident."</i>
	Metacognitive Support & Reflective thinking	The AI acts as a "task manager" or thinking partner, helping students set goals, monitor progress, and reflect on their learning process.	<i>"Before writing, AI helps me establish clear objectives. It makes me think critically about my intentions from the start and keeps our team focused on the same goals."</i>
Hindering Mechanisms	Cognitive Offloading & Intellectual Laziness	Students express concern that relying on AI for ready-made answers allows them to bypass the difficult but necessary cognitive work of learning.	<i>"It's a replacement for any thinking whatsoever. If you use it to do your homework for you, you aren't trying to understand anything. It can make you intellectually lazy."</i>
	Accuracy Concerns & Misinformation	Students are worried about the reliability of AI-generated information, including plausible-sounding but incorrect facts ("hallucinations") and fabricated sources.	<i>"I am concerned about the reliability of information generated by ChatGPT. It can produce incorrect information or even make up sources, so you always have to fact-check it."</i>
	Risk of Over-reliance & Skill Atrophy	There is a concern that excessive reliance on AI will lead to the erosion of fundamental skills, such as independent research, analysis, and writing.	<i>"Over-reliance is a real problem. It can make students rely only on ChatGPT's responses, and they fail to develop their critical thinking. Your skills can atrophy."</i>

Category	Sub-Theme	Description	Student Statement
	Bias Perpetuation	Students recognize that AI models can reflect and amplify biases present in their training data, leading to a distorted understanding of complex issues.	<i>"AI systems inherit biases from their training data. This can limit your exposure to different perspectives and deepen existing misunderstandings if you are not careful."</i>
	Information Overload	The sheer volume of information that AI can generate quickly can be overwhelming, making it difficult for students to filter key information and focus.	<i>"Using ChatGPT sometimes leads to information overload. When it gives you so much at once, it's hard to filter out what is important, and it increases my cognitive load."</i>

The thematic analysis of student perceptions, as detailed in Table 10, indicates a significant dichotomy in their experiences with ChatGPT. These experiences are classified under two primary themes: 'Fostering Mechanisms' and 'Hindering Mechanisms'. On one hand, students reported that ChatGPT is a valuable instrument for promoting higher-order thinking skills. They used the tool for brainstorming and idea generation, a practice that helped them to overcome creative impasses. The AI also exposed them to diverse viewpoints. This exposure broadened their perspectives and encouraged divergent thinking. Furthermore, the tool assisted with argument refinement. It allowed students to evaluate the strengths and weaknesses of different positions, a process that supports convergent thinking. Students also valued ChatGPT's function as a language scaffold. The platform offered a non-judgmental environment for language practice, which effectively lowered the affective filter and reduced student anxiety. Finally, it served as a metacognitive support tool, aiding students in goal setting and reflective thinking on their learning journey.

On the other hand, students were also highly conscious of the tool's potential disadvantages. A principal concern was the risk of cognitive offloading and intellectual laziness. The ease of obtaining answers could encourage students to circumvent critical thought. Concerns about accuracy and misinformation were also prevalent. Students

understood that the AI could generate 'hallucinations,' which are plausible but factually incorrect statements. A significant fear involved over-reliance and subsequent skill atrophy. Excessive dependence on the AI could potentially erode their fundamental academic abilities. Participants also identified the risk of bias perpetuation. They acknowledged the possibility that the AI could reflect and even amplify biases present in its training data. Lastly, the large volume of AI-generated text could result in information overload. This situation increases cognitive load and makes it difficult for students to focus on essential information.

Discussion

The study's findings present an apparent contradiction regarding critical thinking. The quantitative data clearly show that the group using ChatGPT achieved significantly higher critical thinking scores. This suggests a positive causal relationship. The qualitative data, however, reveal that students are acutely aware of and concerned about the potential for AI to hinder critical thinking. They fear it can promote intellectual laziness and superficial engagement. This finding is supported by previous research; for example, [Gerlich \(2025\)](#) noted similar concerns among students. This dichotomy is not a methodological flaw. It is the central tension that defines the role of generative AI in education today. The resolution to this paradox lies in understanding that the tool's impact is not monolithic. Instead, its effect is mediated by the user's mode of engagement.

The positive quantitative results can be explained when students use ChatGPT as a "cognitive co-pilot," a concept explored by [Chen and Chang \(2024\)](#). In this collaborative mode, the AI acts as a scaffold for higher-order cognitive processes, a finding consistent with research by [Nückles et al. \(2020\)](#). Students in this study reported using the tool to deconstruct complex problems, explore topics from multiple perspectives, and evaluate the logic of arguments. This form of interaction does not replace human thought; it augments it. By offloading some of the initial, lower-level cognitive tasks, such as information gathering or initial structuring, the tool frees up the student's cognitive resources ([Schnotz & Kürschner, 2007](#)). This allows the learner to allocate more mental energy to the more demanding tasks of deep analysis, synthesis, and evaluation, which are the hallmarks of critical thinking ([Piolat et al., 2005](#)).

Conversely, the students' legitimate concerns about the negative effects of AI align with the well-documented phenomenon of "cognitive offloading" ([Chen & Chang, 2024](#)). This occurs when a user delegates essential cognitive tasks to an external tool, reducing their engagement in deep, reflective thinking. When students use ChatGPT to find direct answers

or generate complete texts without critical engagement, they bypass the very mental processes that build and strengthen critical thinking skills, a risk highlighted by [Correia et al. \(2024\)](#). This can lead to a superficial understanding of the material and a gradual erosion of independent analytical abilities ([Abbas et al., 2024](#)). Furthermore, research has shown a significant negative correlation between high confidence in AI and the user's application of critical thinking, which suggests that uncritical trust in the tool is intellectually detrimental ([Lee & Tseng, 2025](#)).

This analysis reveals that the key variable determining whether ChatGPT fosters or hinders critical thinking is the pedagogical framework within which it is deployed. The positive quantitative scores achieved by the experimental group are likely the result of a structured learning environment. In this environment, the instructor guided students on how to use the tool critically and strategically. The negative perceptions voiced by students likely reflect their experiences with, or fears of, unguided use, where the default tendency may be to seek the path of least cognitive resistance. This has profound implications for educators. The instructor's role is evolving. It is no longer sufficient to teach a subject like English simply. Educators must now also teach students how to learn effectively with AI, a point emphasized by [Dong et al. \(2025\)](#). This transforms the teacher's primary function from that of a knowledge transmitter to that of a process facilitator, a metacognitive guide, and an instructor of critical AI literacy.

ChatGPT's influence on creativity presents a notable dichotomy. While quantitative data confirmed a significant improvement in creative thinking scores for the experimental group, the qualitative findings highlighted a critical tension. This tension exists between using the AI for creative augmentation and the risk of it promoting creative automation. Differentiating between these two functions is essential for effective pedagogical strategies in AI-integrated learning environments. Creative enhancement occurs when AI is utilized to augment, rather than replace, human cognitive processes ([O'Toole & Horvát, 2024](#)). The results of this study indicate that ChatGPT effectively stimulates both divergent and convergent thinking, which are core components of creativity ([Koivisto & Grassini, 2023](#)). For divergent thinking, students used the AI to generate a wide range of ideas. This practice helped them move beyond conventional thought patterns. For convergent thinking, the tool facilitated the analysis and refinement of these ideas. This process enabled students to construct more coherent and persuasive arguments. In this augmentation framework, the AI serves as a catalyst, providing novel pathways for students to apply their own intellectual and creative abilities ([Xu, 2025](#)).

The primary risk to creativity materializes when the role of AI shifts from augmentation to full automation. Concerns exist that excessive reliance on AI to generate complete works may hinder the development of a student's unique authorial voice, a point noted by [Werdiningsih et al. \(2024\)](#). In such cases, students outsource the entire intellectual labor of the creative process. This includes idea formulation, argument construction, and personal stylistic expression. This practice can lead to a homogenization of student work. The output may be technically proficient but will likely be devoid of the personal insight and intellectual ownership that define genuine creativity. Resolving this tension requires a fundamental pedagogical shift in assessment. The focus should shift from evaluating the final product to assessing the student's creative process ([Zhan et al., 2022](#)). Evidence suggests AI tools are most beneficial during the preliminary and refinement stages of work, such as brainstorming or checking for consistency. The core act of generation and synthesis should remain a fundamentally human task. This principle necessitates the redesign of academic assignments to make a student's intellectual journey both visible and assessable. For instance, educators can design multi-stage tasks that require students to use AI for initial options and then justify how they will adapt and personalize a chosen option ([Nguyen et al., 2024](#)). This pedagogical model values a student's critical engagement with AI as a tool, rather than just the final output that the AI can produce independently ([Hastomo et al., 2025](#)).

A combination of metacognitive theory and Krashen's Affective Filter Hypothesis offers a robust explanation for the observed increase in students' reflective thinking scores ([Krashen, 2003](#)). The interactive nature of ChatGPT appears to create a unique learning environment. This environment promotes both metacognitive development and a reduction in language learning anxiety. Specifically, the AI can operate as a "metacognitive mirror" for the EFL student. The process of formulating a detailed prompt compels a student to engage in reflective thinking. They must first clarify their own ideas and learning objectives to receive a useful response, a mechanism described by [Levine et al. \(2025\)](#). The AI's output provides immediate, externalized feedback on the student's initial query. This initiates an iterative cycle of evaluation, reflection, and refinement of their thinking. Therefore, the AI is not a simple information source but a "thinking partner" that scaffolds reflective interaction in the language learning process ([Lee & Palmer, 2025](#)). This metacognitive cycle is significantly enhanced by the AI's ability to lower the EFL learner's affective filter. Krashen's hypothesis posits that negative emotions like anxiety impede second language acquisition ([Dulay & Burt, 1980](#)). In a traditional classroom, this anxiety often stems from the fear of making errors in front of peers and instructors ([Botes et al., 2020](#)). This study's qualitative data reveal

that students perceived ChatGPT as a "safe" and non-judgmental environment for language practice. The AI provides personalized feedback on linguistic elements, such as grammar and vocabulary. It delivers this feedback without the social risks inherent in human evaluation. By reducing anxiety, the tool frees up students' cognitive resources. This allows them to engage more deeply in the reflective and metacognitive processes necessary for language development.

Lowering the affective filter fosters a supportive atmosphere where EFL students feel empowered to experiment with complex language and take necessary learning risks. Consequently, students become more receptive to linguistic input and more willing to engage in reflective thinking, which is crucial for skill development. This analysis indicates a significant evolution for the role of the language teacher. As AI manages mechanical feedback within a low-stakes environment, the educator's role is liberated. It shifts from a "corrector of errors" to a "facilitator of meaning" who can concentrate on uniquely human, high-impact instruction (Guo, 2024). This new focus involves cultivating a collaborative community and orchestrating nuanced discussions, aligning with pedagogical philosophies like those of Freire (1984) and Nieto (2003). The future of TEFL pedagogy is a hybrid model of human-AI collaboration (Nguyen et al., 2024). In this framework, the AI delivers personalized practice at scale. The human teacher, in turn, designs overarching learning experiences, teaches critical engagement with technology, and cultivates essential interpersonal skills such as communication and empathy.

The findings of this study argue against institutional bans on generative AI, a strategy that is both impractical and pedagogically unsound (de Fine Licht, 2024). Instead, a proactive pedagogical shift is necessary to integrate this technology into EFL curricula thoughtfully. The cornerstone of this shift is the cultivation of Critical AI Literacy. This competence requires educating students on the operational principles and inherent limitations of Large Language Models, including their potential for bias and factual errors (Perkins, 2023). A key component of this literacy is prompt engineering. The process of crafting effective prompts enhances students' metacognitive skills by forcing them to clarify their objectives and structure their thoughts deliberately (Cain, 2024). This pedagogical integration also demands a comprehensive overhaul of assessment practices. Traditional assignments vulnerable to automation must be replaced with AI-resilient tasks that evaluate the learning process over the final product (Cotton et al., 2024). These new assessments prioritize higher-order human skills. They require students to critically evaluate AI-generated content, defend their reasoning in real-time, or document their intellectual workflow. Through these strategic

changes, educators can transform generative AI from a perceived threat into a powerful facilitator for developing the critical and adaptive competencies that EFL learners need (Hastomo et al., 2024).

Conclusion

This study's findings provide a multifaceted view of ChatGPT's impact on higher-order thinking skills in an EFL context. The quantitative results from the quasi-experiment were conclusive. The instructional model using ChatGPT produced statistically significant improvements in students' critical thinking, creative thinking, and reflective thinking scores when compared to traditional methods. The qualitative data from student interviews complemented these findings and revealed a significant duality in their perceptions. Students identified ChatGPT as a valuable mechanism for fostering cognitive skills through idea generation, perspective broadening, and safe, anxiety-free language practice. However, they also expressed strong concerns about the risks of cognitive offloading, skill atrophy, and the potential for misinformation. The primary implication of this study is that the pedagogical framework significantly influences the effectiveness of the tool. ChatGPT acts as a powerful cognitive enhancer when used strategically for augmentation. Its use requires a shift in the educator's role toward facilitating critical AI literacy and process-oriented learning.

It is essential to acknowledge the limitations of this study in order to contextualize the findings. The quasi-experimental design establishes a strong correlation but not definitive causation. Furthermore, the specific cohort of university-level Indonesian students may limit the generalizability of the results to other educational contexts. These limitations highlight several important avenues for future research. There is a need for longitudinal studies to assess the long-term cognitive effects of AI engagement. Future work should also explore how its impact varies across different academic disciplines. Further investigation is also required to understand the precise mechanisms of effective human-AI collaboration and its nuanced impact on affective factors, such as student motivation. Finally, a critical and practical area for future inquiry involves designing and evaluating professional development programs. These programs are needed to equip educators with the competencies for an ethical and effective AI-integrated pedagogy.

Acknowledgement

The authors gratefully acknowledge the financial support for this research from the Directorate General of Higher Education, Ministry of Education and Culture of the Republic

of Indonesia. This funding was awarded under Contract Nos. 123/C3/DT.05.00/PL/2025 (dated May 28, 2025) and 107/LL2/DT.05.00/PL/2025 (dated June 2, 2025).

References

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 1–22. <https://doi.org/10.1186/s41239-024-00444-7>
- Algaraady, J., & Mahyoob, M. (2025). Exploring ChatGPT's potential for augmenting post-editing in machine translation across multiple domains: challenges and opportunities. *Frontiers in Artificial Intelligence*, 8, 1–11. <https://doi.org/10.3389/frai.2025.1526293>
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), 1–4. <https://doi.org/10.7759/cureus.35179>
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/AMLER.2023.02.009>
- Baskara, F. R. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning*, 5(1), 44–55. <https://doi.org/10.31763/ijele.v5i1.858>
- Basol, G., & Gencel, I. E. (2013). Reflective thinking scale: A validity and reliability study. *Educational Sciences: Theory and Practice*, 13(2), 941–946. <https://eric.ed.gov/?id=EJ1017318>
- Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative ai as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning*, 19(4), 479–532. <https://doi.org/10.1007/s11412-024-09437-0>
- Botes, E., Dewaele, J. M., & Greiff, S. (2020). The power to improve: Effects of multilingualism and perceived proficiency on enjoyment and anxiety in foreign language learning. *European Journal of Applied Linguistics*, 8(2), 279–306. <https://doi.org/10.1515/eujal-2020-0003>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

- Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68(1), 47–57. <https://doi.org/10.1007/s11528-023-00896-0>
- Chen, C. H., & Chang, C. L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*, 29, 18621-18642. <https://doi.org/10.1007/s10639-024-12553-x>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 72(4), 2401–2416. <https://doi.org/10.1007/s11423-024-10366-w>
- Combrinck, C., & Loubser, N. (2025). Student self-reflection as a tool for managing GenAI use in large class assessment. *Discover Education*, 4(1), 1–19. <https://doi.org/10.1007/s44217-025-00461-2>
- Correia, A. P., Hickey, S., & Xu, F. (2024). Beyond the virtual classroom: Integrating artificial intelligence in online learning. *Distance Education*, 45(3), 481–491. <https://doi.org/10.1080/01587919.2024.2338706>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.
- de Fine Licht, K. (2024). Generative artificial intelligence in higher education: Why the “banning approach” to student use is sometimes morally justified. *Philosophy & Technology*, 37(3), 1–17. <https://doi.org/10.1007/s13347-024-00799-9>
- Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students’ academic achievement in classroom: A meta-analysis. *Computers and Education: Artificial Intelligence*, 8, 1–10. <https://doi.org/10.1016/j.caeai.2025.100400>
- Dulay, H., & Burt, M. (1980). The relative proficiency of limited English proficient students. *NABE Journal*, 4(3), 1–24. <https://doi.org/10.1080/08855072.1980.10668381>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International*

- Journal for Educational Integrity*, 19(1), 17. <https://doi.org/10.1007/s40979-023-00140-5>
- Freire, P. (1984). *Pedagogy of the oppressed*. The Continuum Publishing Corporation.
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 1–28. <https://doi.org/10.3390/soc15010006>
- Guo, X. (2024). Facilitator or thinking inhibitor: Understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, 32, 1–19. <https://doi.org/10.1080/10494820.2024.2445177>
- Harrington, D. (2009). *Confirmatory factor analysis*. Oxford University Press.
- Hastomo, T., Mandasari, B., & Widiati, U. (2024). Scrutinizing Indonesian pre-service teachers' technological knowledge in utilizing AI-powered tools. *Journal of Education and Learning (EduLearn)*, 18(4), 1572–1581. <https://doi.org/10.11591/edulearn.v18i4.21644>
- Hastomo, T., Sari, A. S., Widiati, U., Ivone, F. M., Zen, E. L., & Kholid, M. F. N. (2025). Does student engagement with chatbots enhance english proficiency?. *ELOPE: English Language Overseas Perspectives and Enquiries*, 22(1), 93–109. <https://doi.org/10.4312/elope.22.1.93-109>
- Hidayat, T., Susilaningsih, E., & Kurniawan, C. (2018). The effectiveness of enrichment test instruments design to measure students' creative thinking skills and problem-solving. *Thinking Skills and Creativity*, 29, 161–169. <https://doi.org/10.1016/j.tsc.2018.02.011>
- Ibrahim, K. A. A. A., Kassem, M. A. M., & Lami, D. (2024). Intelligent Computer-Assisted Language Assessment (ICALA) in philosophy-based language instruction: Unraveling the effects on critical thinking, self-evaluation, academic resilience, and speaking development. *Language Testing in Asia*, 14(1), 1–17. <https://doi.org/10.1186/s40468-024-00320-1>
- Janse van Rensburg, J. (2024). Artificial human thinking: ChatGPT's capacity to be a model for critical thinking when prompted with problem-based writing activities. *Discover Education*, 3(1), 1–12. <https://doi.org/10.1007/s44217-024-00113-x>
- Jelodari, M., Amirhosseini, M. H., & Giraldez-Hayes, A. (2023). An AI powered system to enhance self-reflection practice in coaching. *Cognitive Computation and Systems*, 5(4), 243–254. <https://doi.org/10.1049/ccs2.12087>
- Kartal, G. (2024). The influence of ChatGPT on thinking skills and creativity of EFL student teachers: A narrative inquiry. *Journal of Education for Teaching*, 50(4), 627–642. <https://doi.org/10.1080/02607476.2024.2326502>

- Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific Reports*, *13*(1), 1–10. <https://doi.org/10.1038/s41598-023-40858-3>
- Krashen, S. D. (2003). *Explorations in Language Acquisition and Use*. Heinemann.
- Lee, C. C., & Low, M. Y. H. (2024). Using genAI in education: The case for critical thinking. *Frontiers in Artificial Intelligence*, *7*, 1–3. <https://doi.org/10.3389/frai.2024.1452131>
- Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, *22*(1), 1–22. <https://doi.org/10.1186/s41239-025-00503-7>
- Lee, K. W., & Tseng, Y. F. (2025). When ChatGPT meets classical management theories: The role and impact of AI Chatbots on management learning. *Interactive Learning Environments*, *33*, 1–24. <https://doi.org/10.1080/10494820.2025.2482589>
- Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2025). How do students use ChatGPT as a writing support?. *Journal of Adolescent & Adult Literacy*, *68*(5), 445–457. <https://doi.org/10.1002/jaal.1373>
- Lin, H., & Chen, Q. (2024). Artificial intelligence (AI) -integrated educational applications and college students' creativity and academic emotions: Students and teachers' perceptions and attitudes. *BMC Psychology*, *12*(1), 1–16. <https://doi.org/10.1186/s40359-024-01979-0>
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries?. *Library Hi Tech News*, *40*(3), 26–29. <https://doi.org/10.2139/ssrn.4333415>
- Luther, T., Kimmerle, J., & Cress, U. (2024). Teaming up with an AI: Exploring human–AI collaboration in a writing scenario with ChatGPT. *AI*, *5*(3), 1357–1376. <https://doi.org/10.3390/ai5030065>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, *10*(2), 1–17. <https://doi.org/10.1080/2331186X.2023.2236469>
- Mohamed, A. M. (2024). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, *29*(3), 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>

- Mutanga, M. B., Msane, J., Mndaweni, T. N., Hlongwane, B. B., & Ngcobo, N. Z. (2025). Exploring the Impact of LLM Prompting on Students' Learning. *Trends in Higher Education*, 4(3), 31. <https://doi.org/10.3390/higheredu4030031>
- Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5), 847–864. <https://doi.org/10.1080/03075079.2024.2323593>
- Nguyen, N. N., & Barbieri, W. (2025). Mentorship in the age of generative AI: ChatGPT to support self-regulated learning of pre-service teachers before and during placements. *Education Sciences*, 15(6), 642. <https://doi.org/10.3390/educsci15060642>
- Nguyen, T. H. B., & Tran, T. D. H. (2023). Exploring the efficacy of ChatGPT in language teaching. *AsiaCALL Online Journal*, 14(2), 156–167. <https://doi.org/10.54855/acoj.2314210>
- Nieto, S. (2003). *What keeps teachers going?*. Teachers College Press.
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, 32(4), 1089–1126. <https://doi.org/10.1007/s10648-020-09541-1>
- Nugroho, A., Andriyanti, E., Widodo, P., & Mutiaraningrum, I. (2025). Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing. *Innovations in Education and Teaching International*, 62(2), 499–511. <https://doi.org/10.1080/14703297.2024.2319184>
- Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, 8(4), 2306–2319. <https://doi.org/10.55214/25768484.v8i4.1600>
- O'Toole, K., & Horvát, E. Á. (2024). Extending human creativity with AI. *Journal of Creativity*, 34(2), 1–8. <https://doi.org/10.1016/j.yjoc.2024.100080>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), 07. <https://doi.org/10.53761/1.20.02.07>
- Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19(3), 291–312. <https://doi.org/10.1002/acp.1086>
- Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges,

- and future research directions. *Journal of Applied Learning and Teaching*, 6(1), 41–56. <https://doi.org/10.37074/JALT.2023.6.1.29>
- Rong, Q., Lian, Q., & Tang, T. (2022). Research on the influence of AI and VR technology for students' concentration and creativity. *Frontiers in Psychology*, 13, 1–9. <https://doi.org/10.3389/fpsyg.2022.767689>
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19, 469–508. <https://doi.org/10.1007/s10648-007-9053-4>
- Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence*, 4(1), 46. <https://doi.org/10.1007/s44163-024-00143-2>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1–14. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity*, 9, 107–119. <https://doi.org/10.1016/j.tsc.2012.09.002>
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., ... & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, 11(3), 1–32. <https://doi.org/10.3390/jintelligence11030054>
- Tseng, Y. C., & Lin, Y. H. (2024). Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, 22(2), 78–97. <https://doi.org/10.34190/ejel.21.5.3329>
- Tu, J. (2020). Learn to speak like a native: AI-powered chatbot simulating natural conversation for language tutoring. *Journal of Physics: Conference Series*, 1693(1), 012216. <https://doi.org/10.1088/1742-6596/1693/1/012216>
- Van Horn, K. R. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *Teaching English as a Second or Foreign Language-TESL-EJ*, 28(1), 1–26. <https://doi.org/10.55593/ej.28109a8>
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education.

- International Journal of Educational Technology in Higher Education*, 21(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>
- Wang, C. (2024). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*, 30, 1825-1846. <https://doi.org/10.1007/s10758-024-09744-3>
- Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(1), 1–11. <https://doi.org/10.1057/s41599-025-04787-y>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Waziana, W., Andewi, W., Hastomo, T., & Hasbi, M. (2024). Students' perceptions about the impact of AI chatbots on their vocabulary and grammar in EFL writing. *Register Journal*, 17(2), 328–362. <https://doi.org/10.18326/register.v17i2.352-382>
- Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities*, 11(1), 1–15. <https://doi.org/10.1080/23311983.2024.2392388>
- Widianingtyas, N., Mukti, T. W. P., & Silalahi, R. M. P. (2023). ChatGPT in language education: Perceptions of teachers - a beneficial tool or potential threat?. *VELES (Voices of English Language Education Society)*, 7(2), 279–290. <https://doi.org/10.29408/veles.v7i2.20326>
- Wulyani, A. N., Widiati, U., Muniroh, S., Rachmadhany, C. D., Nurlaila, N., Hanifiyah, L., & Sharif, T. I. S. T. (2024). Patterns of utilizing AI-assisted tools among EFL students: Need surveys for assessment model development. *LLT Journal: A Journal on Language and Language Teaching*, 27(1), 157–173. <https://doi.org/10.24071/llt.v27i1.7966>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/LANGUAGES8030212>
- Xu, M. (2025). Interaction between students and artificial intelligence in the context of creative potential development. *Interactive Learning Environments*, 33, 1–16. <https://doi.org/10.1080/10494820.2025.2465439>

- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 327–343. <https://doi.org/10.1017/S0958344022000039>
- Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, *21*(1), 21. <https://doi.org/10.1186/s41239-024-00453-6>
- Zhan, Z., Shen, W., & Lin, W. (2022). Effect of product-based pedagogy on students' project management skills, learning achievement, creativity, and innovative thinking in a high-school artificial intelligence course. *Frontiers in Psychology*, *13*, 1–16. <https://doi.org/10.3389/fpsyg.2022.849842>

Appendices

Instrument I: Critical Thinking Scale (CTS)

Participant Instructions

Please indicate your level of agreement with the following statements on a 5-point scale, where:

1 = Strongly Disagree

2 = Disagree

3 = Neutral

4 = Agree

5 = Strongly Agree

Table A1. Critical Thinking Scale (CTS) Items

Subscale	Item No.	Statement
Analysis	1	In English class discussions, I try to think about the bigger picture, not just the details.
	2	I often use new concepts from my English course to change the way I approach my assignments.
	3	For my English course assignments, I make it a point to consult multiple sources (e.g., academic journals, reputable websites, books) to get a balanced view.
	4	I am often on the lookout for new ideas or perspectives when reading academic texts in English.
	5	I appreciate when a reading in my English class presents a strong argument that challenges my existing beliefs.
	6	I believe it is important to understand different cultural viewpoints presented in English-language materials.
	7	In my academic writing, I know it is important to provide strong justification for the claims I make.
	8	I am willing to consider different interpretations of a text, even if they conflict with my own initial reading.
	9	I am open to receiving critical feedback from my instructor or peers on my English writing drafts.
	10	Before deciding on a final argument for an essay, I like to explore several alternative viewpoints.

Subscale	Item No.	Statement
	11	I am genuinely curious about the complex topics we discuss in my English course.
	12	I try to connect ideas from my English course with concepts I've learned in my Information Systems courses.
Reflective Skepticism	13	I often re-evaluate my understanding of a topic after a class discussion or lecture.
	14	When researching for an English essay, I critically evaluate the credibility of my sources (e.g., author's expertise, publication bias).
	15	Before taking a strong stance in a debate or paper, I think about the wider implications of my position.
	16	After completing an English assignment, I often think about my process to see how I could have done it better.
	17	When an academic text presents a conclusion, I look for the evidence the author provides to support it.
	18	I try to identify potential biases or logical weaknesses in the arguments I read for my courses.
	19	I make an effort to question my own assumptions before starting a research paper in English.
	20	I make sure to cross-reference information from online sources with academic texts to ensure its accuracy for my assignments.

Scoring Protocol

Sum the responses for all 20 items. The total score will range from 20 to 100.

Instrument II: Creative Thinking Scale

Participant Instructions

Please indicate your level of agreement with the following statements on a 5-point scale, where:

1 = Strongly Disagree

2 = Disagree

3 = Neutral

4 = Agree

5 = Strongly Agree

Table A2. Creative Thinking Scale Items

Subscale	Item No.	Statement
Innovative Thinking	1	In my English assignments, I enjoy developing a unique argument or thesis that my peers might not have considered.
	2	I am naturally curious about the underlying meanings or alternative interpretations of the texts we study in English.
	3	I often come up with unconventional or original ideas when brainstorming for an English presentation.
	4	I prefer English assignments that allow me to generate my own original concepts rather than just summarize information.
	5	I find it easy to connect ideas from different English texts or modules in a way that creates a new perspective.
	6	When working on a project, I try to find a completely new way to present my information.
	7 (R)	When writing an essay for my English class, I prefer to use a standard, safe structure rather than experimenting with a more creative format.
Intellectual Courage	8	I am not afraid to propose an interpretation of a text in class that might seem unusual to others.
	9	I am willing to defend my ideas in an academic debate, even if they are unpopular.
	10	I see critical feedback on my creative ideas as a valuable learning opportunity, not a failure.
	11	I enjoy tackling challenging English assignments that push my intellectual limits.
	12	I am willing to take intellectual risks (e.g., argue a controversial point) to produce a more original essay.
	13 (R)	I worry about what my instructor or peers will think if I suggest a strange or different idea for a project.
14 (R)	I avoid expressing opinions in class discussions that go against the general consensus of the group.	
Flexibility	15	I can easily see an academic problem or topic from multiple different viewpoints.
	16	When my first approach to an essay isn't working, I quickly think of alternative structures or arguments.
	17	I enjoy considering academic perspectives in my readings that are very different from my own.
	18	I am good at adapting my thinking about a topic when presented with new information in class.
	19	I am open to changing my thesis statement for an essay if I encounter a strong counter-argument during my research.
	20 (R)	I find it difficult to move away from my initial idea about how to approach an assignment.

(R) indicates a reverse-scored item.

Scoring Protocol

- Reverse Score Items:** For items marked with (R) (7, 13, 14, 20), reverse the score: 1→5, 2→4, 3→3, 4→2, 5→1.
- Calculate Total Score:** Sum the scores for all 20 items (using the reversed scores for the appropriate items). The total score will range from 20 to 100.

Instrument III: Reflective Thinking Scale (RTS)

Participant Instructions

Please indicate your level of agreement with the following statements on a 5-point scale, where:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neutral
- 4 = Agree
- 5 = Strongly Agree

Table A3. Reflective Thinking Scale (RTS) Items

Subscale	Item No.	Statement
Habitual Action	1 (R)	For my English assignments, I tend to use the same approach and vocabulary without thinking much about how to improve.
	2 (R)	When I study for my English course, I often just memorize the facts as they are given in the textbook.
	3 (R)	I prefer to follow the assignment instructions exactly without questioning them or trying a different approach.
	4 (R)	Once I learn a way to write an essay or do a presentation, I don't see a reason to change my method.
	5 (R)	I tend to complete my English grammar exercises quickly without thinking deeply about the rules behind them.
Understanding	6	When I learn a new English vocabulary word, I try to connect it to other words I already know.
	7	To learn a grammar rule well, I need to understand the logic behind it, not just memorize it.
	8	I try to find my own examples to help me understand difficult concepts from our English course readings.
	9	I work to make sure I fully comprehend the academic articles we read, not just find the answers for a quiz.
	10	I try to break down complex academic arguments from our readings into smaller, more understandable parts.

Subscale	Item No.	Statement
Reflection	11	After receiving feedback on an English essay, I analyze the comments to understand my mistakes and plan how to improve my next assignment.
	12	I re-examine my understanding of a topic when a class discussion presents new information that contradicts my view.
	13	My experiences in my English class often lead me to reconsider my own ideas about language and communication.
	14	I try to analyze my own performance after giving a presentation to identify areas for improvement.
	15	Comparing my own writing to examples of good academic essays helps me see my own strengths and weaknesses.
Critical Reflection	16	I sometimes question the fundamental assumptions or cultural biases behind the academic texts we are assigned to read.
	17	Learning about different perspectives in my English course has sometimes led me to change my long-held beliefs or values.
	18	I think it's important to analyze my own biases when evaluating an argument in an academic article.
	19	I try to understand the broader societal implications of the topics we discuss in my English course.
	20	My English course sometimes makes me think about how language shapes the way we understand technology and society.

(R) indicates a reverse-scored item.

Scoring Protocol

- Reverse Score Items:** For items marked with (R) (1, 2, 3, 4, 5), reverse the score: 1→5, 2→4, 3→3, 4→2, 5→1.
- Calculate Total Score:** Sum the scores for all 20 items (using the reversed scores for the appropriate items). The total score will range from 20 to 100.

Instrument IV: Semi-Structured Interview Guide

Phase 1: Opening & General Experience (Approx. 5 minutes)

- "To start, thank you again for agreeing to talk with me. Could you tell me a bit about your general experience using ChatGPT in your English class this semester? What kinds of tasks did you use it for?"
- "Overall, what was your first impression of using ChatGPT as a learning tool?"

Phase 2: Core Questions on Fostering Mechanisms (Approx. 15-20 minutes)

3. "Let's talk about critical thinking—things like analyzing arguments, evaluating evidence, or questioning information. Can you recall a specific time when an interaction with ChatGPT actually helped you think more critically about a topic for your English class?"
4. "Now, let's think about creativity—coming up with original ideas, being imaginative, or finding new ways to express yourself. Can you describe an experience where using ChatGPT helped you be more creative in your English work?"
5. "Finally, let's consider reflective thinking—thinking about your own learning process, understanding your strengths and weaknesses, or evaluating your own work. Was there any way that using ChatGPT prompted you to reflect more on your own learning in English?"

Phase 3: Core Questions on Hindering Mechanisms (Approx. 10-15 minutes)

6. "We've talked about the helpful aspects. Now I'd like to ask about the other side. Have there been times when you felt using ChatGPT might have been a disadvantage for your thinking skills?"
7. "Thinking again about critical thinking, have you ever felt that ChatGPT made it too easy to find an answer, possibly preventing you from thinking through a problem yourself? Can you describe a situation like that?"
8. "What about creativity? Have you ever felt that using ChatGPT limited your own original ideas or made your work feel less personal or unique?"
9. "And for reflective thinking, do you think using ChatGPT for quick answers or corrections might cause students to miss opportunities to reflect on their own mistakes and learn from them?"

Phase 4: Conclusion & Synthesis (Approx. 5 minutes)

10. "Based on all your experiences, what is the most important difference between using ChatGPT as a helpful tool versus using it as a crutch that stops you from thinking?"
11. "If you were to give advice to a new student on how to use ChatGPT to genuinely improve their thinking skills in English, what would you tell them?"
12. "Is there anything else you'd like to share about your experience with ChatGPT that we haven't talked about?"

"Thank you so much for your time and for sharing your valuable insights. This has been very helpful."