JSEL
Journal of Studies in the English Language

From Prompting to Proficiency: A Mixed-Methods Analysis of Prompting with ChatGPT Versus Lecturer Interaction in an EFL Classroom

Widi Andewi[1]*
Winia Waziana[1]
Damar Wibisono[2]
Kristian Adi Putra[3]
Tommy Hastomo[4]
Irene Brainnita Oktarin[5]
Institut Bakti Nusantara, Indonesia [1]
Universitas Lampung, Indonesia [2]
Prince Sattam Bin Abdulaziz University, Saudi Arabia [3]
STKIP PGRI Bandar Lampung, Indonesia [4]
STIE Gentiaras, Indonesia [5]
*Corresponding author's email: widiandewi.91@gmail.com
Received July 26, 2025; revised August 27, 2025; accepted August 28, 2025

Abstract

The rapid advancement of artificial intelligence (AI) and extensive language models like ChatGPT have had profound implications for English as a Foreign Language (EFL) pedagogical practices. However, while its applications are widely explored, a notable gap persists in understanding the effectiveness of its core "prompting" feature for academic help-seeking compared to traditional lecturer interaction. This study aimed to address this gap by comparing the effectiveness of ChatGPT versus lecturers, exploring the advancement of student prompting strategies, and identifying associated challenges. A sequential explanatory mixed-methods design was employed over eight weeks with 60 Indonesian university students allocated to an experimental (ChatGPT) and a control (lecturer) group. Data collected via pre/post-tests (proficiency, writing, self-efficacy), interviews, and a Focus Group Discussion were analyzed using ANCOVA and thematic analysis. The quantitative findings conclusively demonstrated that the ChatGPT group significantly outperformed the control group in enhancing general English proficiency, writing competency, and self-efficacy ($p < .001$). Qualitatively, students' prompting strategies evolved from simple, single-turn queries to sophisticated, multi-turn dialogic interactions, which were a key determinant of deeper learning. The most critical challenge identified was pedagogy, specifically the tendency for cognitive offloading alongside difficulties in vetting the AI's accuracy and pervasive anxiety regarding academic integrity. These findings suggest that the effective use of Generative AI (GenAI) is not an innate skill but a learned competence,

necessitating a pedagogical shift from merely providing technological access to the explicit instruction of AI literacy and strategic prompting to harness its full potential.

Keywords: AI literacy, ChatGPT, EFL learning, prompt engineering, scaffolding inversion

The rapid development of artificial intelligence (AI) and large language models (LLMs) such as ChatGPT has significantly influenced instruction in English as a Foreign Language (EFL). Since ChatGPT was launched publicly by OpenAI in late 2022, many students worldwide have adopted it to support their learning (Belda-Medina & Kokošková, 2023). In Indonesia, learners use ChatGPT's conversational features to improve their understanding, expand vocabulary, refine sentence construction, and receive immediate feedback on writing and speaking tasks (Waziana et al., 2024). A key feature in this process is prompting, which refers to designing clear and contextually appropriate instructions to effectively direct the AI's output (Zhu et al., 2024).

Although Generative AI (GenAI) offers promising educational benefits, it is important to evaluate its effectiveness compared to conventional, in-person instruction. This is particularly relevant in higher education, where efficient learning approaches are essential. There are also concerns that dependence on AI might limit students' development of pragmatic and interpersonal skills typically nurtured through human interaction (Bai et al., 2023). The value of AI-based interaction relies heavily on the learner's prompting ability, an emerging skill requiring pedagogical guidance (Woo et al., 2025). A student's success in using AI is closely linked to their ability to generate purposeful and well-structured prompts. This study seeks to evaluate the effectiveness of prompting and explore how students develop prompting strategies and navigate related challenges in their interaction with ChatGPT.

Studies have shown that the quality of AI responses is primarily determined by how prompts are formulated. A prompt acts as a linguistic command that directs the AI to produce valuable and accurate content (Cain, 2024). According to Knoth et al. (2024), the ability to create effective prompts reflects a core component of AI literacy rather than just a technical skill. Students with strong AI literacy produce more straightforward prompts, which leads to more accurate and beneficial responses that support independent learning. Giray (2023) also emphasizes that this skill enables learners to enhance their academic tasks and research processes. Mohamed (2024) further reports that students view prompting as essential to maximizing ChatGPT's usefulness alongside traditional learning methods.

In EFL education, ChatGPT has been applied in a variety of ways, including assisting in writing essays (Kim et al., 2025), preparing lesson plans (Kerr & Kim, 2025), composing collaborative stories (Woo et al., 2025), supporting research (Nurchurifiani et al., 2025), boosting learner engagement (Hastomo et al., 2025), and facilitating teacher development (Zulianti et al., 2024). ChatGPT's ability to engage in dialogue offers flexibility and personalization that are often limited in conventional classrooms due to rigid curricula and restricted teacher availability (Slamet, 2024). Learners can ask questions on various subjects and receive immediate responses, making academic support accessible anytime and anywhere.

However, using ChatGPT also presents challenges. Mutanga et al. (2025) observed that many learners rely on basic prompting methods such as directly copying questions, often resulting in low-quality responses. In contrast, students who adopt more complex strategies like combining or rephrasing questions tend to receive better output. Understanding how students develop and refine these strategies over time is important. The language of the prompt is also important. Hwang et al. (2025) reported that prompts written in English produced more accurate results compared to those written in a student's first language. Clarity, purpose, and consistency further influence the AI's responses.

Recent scholarship highlights the importance of prompt engineering and AI literacy in higher education. Without targeted instruction, students often engage with AI at a surface level (Vera, 2024). A direct link has been observed between AI literacy and the complexity of prompt use (Kim et al., 2025). Skilled users engage in interactive, refined prompting, while others rely on generic inputs (An et al., 2023). In teacher training programs, the success of AI integration depends on the user's ability to manage varied responses and align them with instructional goals (Kerr & Kim, 2025). This shows that effective use of ChatGPT in language learning depends on access and user competence. Despite this growing body of research, few studies have examined the direct impact of ChatGPT on academic outcomes through students' help-seeking via prompting. This study aims to fill that gap by comparing the effectiveness of prompting and traditional lecturer support, while also analyzing how students frame prompts and the difficulties they face. Three key questions guide this research:

1. Is prompting with ChatGPT more effective than direct interaction with a lecturer in enhancing students' English language proficiency?

2. How are ChatGPT prompting strategies formulated and adapted by students to improve their English language proficiency?

3. What challenges are associated with prompting ChatGPT to enhance students' English language proficiency?

Literature Review

ChatGPT Interaction in EFL Learning

Recent developments have shown a growing integration of ChatGPT into EFL education. Its user-friendly interface and responsive features enable learners to engage in personalized, self-paced language practice. Studies have reported that EFL students utilize ChatGPT for various purposes such as vocabulary development, grammar assistance, writing support, dialogue simulation, learner engagement, research, and professional growth (Hastomo et al., 2025; Kerr & Kim, 2025; Nurchurifiani et al., 2025; Oktarin et al., 2024; Zulianti et al., 2024). GenAI provides opportunities for learners to experience authentic language use and simulate real-life communication. However, the effectiveness of such interactions depends on how learners initiate and sustain their prompts, highlighting the need for instructional guidance (Mutanga et al., 2025).

This integration aligns with a sociocultural theoretical lens, particularly Vygotsky's Zone of Proximal Development, which underscores the importance of scaffolding in learning (Vygotsky, 1978). ChatGPT offers immediate responses as a simulated conversational partner that helps bridge learners' linguistic gaps. Although it does not replace teacher feedback, it can function as a form of distributed cognition that supports formal instruction (Slamet, 2024). Still, research has identified limitations, including shallow engagement and insufficient pragmatic depth, mainly when learners rely on simple prompts (Zheng & Stewart, 2024). These findings suggest that the quality of interaction is influenced by access to technology and learners' autonomy and metacognitive strategies.

Understanding such interactions involves dialogic learning, in which meaning emerges through active questioning and responsive dialogue. Open-ended prompting encourages students to construct meaning collaboratively and reflect on the feedback received (Käser & Schwartz, 2020). An et al. (2023) found that extended interactions with ChatGPT can enhance language awareness, though repetitive exchanges may compromise content relevance. This research positions AI as a valuable educational tool, raising concerns about learner control, authorship, and the authenticity of language production. Although ChatGPT holds significant promise for EFL instruction, its pedagogical value depends mainly on how students generate and refine their prompts. This aspect remains insufficiently examined and calls for further empirical exploration in EFL contexts.

Prompting Strategies and AI Literacy

Developing the ability to construct effective prompts has become essential in AI-assisted language learning. Knoth et al. (2024) describe prompting as the interface layer that connects human users to large language models, directly

shaping the relevance and quality of AI responses. In EFL contexts, effective prompting is linked to enhanced language performance (Kerr & Kim, 2025). Learners with higher levels of AI literacy are better able to produce structured, purposeful, and contextually appropriate prompts. In contrast, beginners often use vague or fragmented instructions, which result in limited feedback and reduced learning outcomes (Zamfirescu-Pereira et al., 2023).

This line of inquiry is grounded in constructivist learning theory, which emphasizes active knowledge construction through engagement and reflection. Vera (2024) found that students working on research methods tasks with ChatGPT demonstrated minimal progress without targeted guidance on prompt formulation. Similarly, Woo et al. (2025) showed that students who were prompted with a clear intention, such as resolving writer's block or expanding narrative content, produced more sophisticated language output. These findings indicate that prompting involves cognitive and metacognitive processes, requiring awareness of linguistic goals and how AI systems interpret human input.

Prompting is closely related to AI literacy, digital competence, and second language pragmatics. According to Leung (2024), prompt engineering includes teachable skills such as clarity, specificity, contextual awareness, and the ability to revise prompts iteratively. Sawalha et al. (2024) identified several prompting strategies, including single-copy, reformulated, and multi-question formats, with the latter two associated with better learning outcomes. These findings suggest that prompting is a dynamic skill that improves through structured practice and reflective feedback. As Kerr and Kim (2025) emphasized, integrating prompt engineering into EFL teacher training and curriculum is essential, since digital tools alone cannot guarantee meaningful learning without thoughtful human mediation. Therefore, prompt engineering must be regarded as a technical competence and a reflective cognitive process. However, this dimension remains underexplored in current empirical studies.

Self-efficacy has been widely recognized as a crucial psychological factor influencing second language learning outcomes. Fan and Cui (2024) define self-efficacy as an individual's belief in their ability to successfully perform specific tasks, which in the context of EFL refers to students' confidence in using English for communication and academic purposes. Research consistently shows that learners with higher self-efficacy are more likely to persist in challenging language tasks, employ effective strategies, and perform better in reading, writing, and speaking (Mahande et al., 2025). In this sense, self-efficacy reflects motivation and actively mediates the relationship between instructional input and language proficiency gains. The present study examined self-efficacy alongside proficiency and writing to capture how psychological readiness interacts with linguistic performance in AI-mediated learning.

Method

Research Design

This research employed a sequential explanatory mixed-methods design, an approach in which quantitative data provide a general understanding of a phenomenon, which is then explained in greater detail through qualitative data (Creswell & Clark, 2017). To answer the RQ1, a quasi-experimental methodology was employed for eight weeks to quantitatively assess the comparative effectiveness of the ChatGPT intervention against traditional lecturer-led interaction. The objective of the intervention was to measure its effect on students' English proficiency, writing competency, and self-efficacy through pre-tests and post-tests. The experimental group completed weekly English tasks, including essay writing, paragraph revision, and grammar and vocabulary practice using ChatGPT. They were guided to refine prompts, request revisions, and record their interactions with ChatGPT as part of the assignment. The control group completed the same tasks with direct guidance and feedback from the lecturer. The same lecturer taught both groups to ensure consistency in objectives and assessment. In the experimental group, the lecturer only introduced the weekly objectives, explained the task requirements, and monitored the ethical use of ChatGPT without giving language corrections. Students in this group were allowed to access ChatGPT during preparation, drafting, and revision stages, while students in the control group received feedback during class meetings and through written comments after submission. Subsequently, to address the RQ2 and RQ3, a qualitative inquiry was undertaken to offer a rich, contextualized interpretation of the initial findings by exploring the processes and challenges unique to the experimental group.

Participants

This study was conducted at Bakti Nusantara Institute, a private university in Indonesia specializing in technology and information systems. The participants were fourth-semester students from the Information Systems program, where English is required to prepare students for global professional environments. A total of 60 students participated, selected through purposive sampling based on specific inclusion criteria: active enrollment in the 2024/2025 academic year, participation in the mandatory English course regardless of proficiency level, and willingness to provide informed consent for the full eight-week study. Students with prior formal training in advanced prompt engineering were excluded to avoid potential bias.

Participants were then assigned to the experimental group (n = 30) and control group (n = 30) using stratified random sampling, with stratification based on English proficiency levels identified through an institutional placement test. Students were grouped into Beginner, Intermediate, and

Advanced categories to ensure proportional representation across both groups, strengthening internal validity (Cohen et al., 2018). As shown in Table 1, most participants were at the Beginner level (80.0%), and the sample was predominantly female (81.7%), reflecting the demographic composition of the program during the study.

Table 1
The Participants' Demographic Information

Characteristic	Category	Experimental Group (n = 30)	Control Group (n = 30)	Total Sample (N = 60)
Age (years)	19	11 (36.7%)	12 (40.0%)	23 (38.3%)
	20	15 (50.0%)	14 (46.7%)	29 (48.3%)
	21	4 (13.3%)	4 (13.3%)	8 (13.3%)
	M (*SD*)	19.77 (0.68)	19.73 (0.70)	19.75 (0.69)
Gender	Male	5 (16.7%)	6 (20.0%)	11 (18.3%)
	Female	25 (83.3%)	24 (80.0%)	49 (81.7%)
English Proficiency	Beginner	24 (80.0%)	24 (80.0%)	48 (80.0%)
	Intermediate	5 (16.7%)	4 (13.3%)	9 (15.0%)
	Advanced	1 (3.3%)	2 (6.7%)	3 (5.0%)

Instruments

To address all research questions comprehensively, this study employed four primary instruments, consisting of two quantitative tools and two qualitative instruments. Each was selected and validated to fulfill a specific role within the mixed-methods framework. Instrument validation involved expert judgment from two senior TEFL lecturers, following established procedures (Lynn, 1986), and was further improved based on insights gained from a pilot study. To address RQ1, three instruments were employed to capture different dimensions of proficiency: the English Proficiency Test measured receptive linguistic ability, the essay task assessed productive competence, and the self-efficacy questionnaire reflected learners' confidence in using English for academic purposes. To answer RQ2, semi-structured interviews explored how students developed and applied prompting strategies with ChatGPT. To address RQ3, focus group discussions were carried out to identify students' challenges and perceptions when integrating ChatGPT into their learning. These instruments ensured that the study could evaluate measurable proficiency gains and the processes and experiences that shaped students' engagement with ChatGPT.

English Proficiency Tests

The English Proficiency Test was used to quantitatively assess student language proficiency changes (RQ1). Developed by the Language Center of

Bakti Nusantara Institute, the test comprises 50 multiple-choice items covering listening, grammar, and reading. It has been validated through institutional testing procedures and has demonstrated acceptable reliability in prior administrations, consistent with established practices in language test validation (Brookhart, 2018; Cohen et al., 2018). The second instrument was an argumentative essay, evaluated using a rubric grounded in practical writing assessment principles (Brookhart, 2018). The rubric measured five aspects of writing: content and argumentation, organization, vocabulary, language use, and mechanics. Two trained raters independently scored the essays to ensure consistency. Inter-rater reliability, calculated using Cohen's kappa, produced a coefficient of 0.81, which reflects substantial agreement (Landis & Koch, 1977). Any differences in scoring were resolved through discussion and consensus. The writing task was conducted in a controlled environment, with a 60-minute time limit to maintain standardized testing conditions.

Self-Efficacy Questionnaire

To capture the psychological dimension of the learning experience, a self-efficacy questionnaire, adapted from the work of Mahande et al. (2025), was administered before and after the intervention. This instrument consisted of 20 items on a 5-point Likert scale to measure students' perceived confidence in using English for various communicative purposes. The reliability analysis for this instrument yielded a high internal consistency, with Cronbach's alpha coefficient of 0.85, which is considered a good reliability level (George & Mallery, 2019).

Semi-Structured Interviews

To gather in-depth qualitative data for RQ2 and RQ3, semi-structured interviews were conducted with 15 purposively selected participants from the experimental group. An interview protocol was developed based on the guidelines proposed by Brinkmann and Kvale (2005). These interviews, conducted at the mid-point (week 4) and end (week 8) of the study, aimed to uncover the thought processes behind how students formulated and adapted their prompting strategies. Each interview lasted approximately 15-20 minutes, was audio-recorded with the participant's consent, and was transcribed verbatim to ensure data accuracy for thematic analysis.

Focus Group Discussion (FGD)

A Focus Group Discussion (FGD) was conducted post-intervention to provide a deeper, collective understanding of students' challenges (RQ3). The FGD involved 15 purposively selected participants from the experimental group,

chosen for their diverse experiences with ChatGPT. The session was guided by a semi-structured protocol and facilitated by a researcher to encourage dynamic interaction among participants, a method known to stimulate shared norms and richer insights than individual interviews alone (Krueger, 2014). The 90-minute discussion was audio- and video-recorded, and subsequently transcribed to capture the full spectrum of verbal and non-verbal cues during the interaction. In addition, chat logs from the experimental group were collected as supplementary data. These logs were submitted alongside weekly tasks to document students' prompting practices and to complement the qualitative analysis.

Data Collection

The data collection process was systematically structured over nine weeks and commenced after obtaining ethical approval from the Research and Community Service Institute (LPPM) of Bakti Nusantara Institute and securing informed consent from all participants. Initially, all 60 participants completed a pre-test battery in the first week, which consisted of the English Proficiency Test, an initial argumentative essay, and a self-efficacy questionnaire. This was followed by an eight-week intervention phase, during which participants in the experimental group used ChatGPT for their weekly English tasks, while the control group received parallel instruction from a lecturer. Semi-structured interviews were conducted at the mid-point (week 4) and end (week 8) of the study, aimed to uncover the thought processes behind how students formulated and adapted their prompting strategies. Each interview lasted approximately 15–20 minutes, was audio-recorded with the participant's consent, and was transcribed verbatim. All interviews and the FGD were conducted in students' L1 (*Bahasa Indonesia*) to ensure that participants could express their ideas freely and in detail. The transcripts were translated into English before the thematic analysis to maintain consistency and comparability across the data.

Data Analysis

Quantitative Data Analysis

Quantitative data were analyzed using SPSS version 27. To answer RQ1, descriptive statistics were calculated for all variables. A series of inferential statistical tests was then performed. To further enhance rigor, Analysis of Covariance (ANCOVA) was applied to control for any minor pre-existing differences in pre-test scores, with the pre-test score as the covariate (Pallant, 2016).

Qualitative Data Analysis

All qualitative data, including interview and FGD transcripts, were transcribed verbatim. To address RQ2, the data were analyzed using the six-phase recursive process outlined in the thematic analysis (Braun & Clarke, 2006). This involved familiarizing oneself with the data, generating initial codes, searching for themes, reviewing the themes, defining and naming them, and producing the final report, all facilitated by NVivo software. Themes were developed inductively by repeatedly reading the transcripts, identifying meaningful units of text, clustering similar codes, and refining them through iterative comparison. Two researchers independently reviewed the coding process and discussed discrepancies until agreement was reached to enhance the credibility and trustworthiness of the analysis. The categories reported in Table 4 were derived directly from this inductive thematic analysis and represent the main themes and subthemes identified for RQ2. To answer RQ3, a content analysis was conducted on data related to challenges (Krippendorff, 2018). This analysis focused on identifying recurrent patterns of difficulties rather than quantifying frequency, and the categories induced from the data were in line with an interpretive approach. The categories presented in Table 5 reflect the results of this inductive content analysis.

Ethical Considerations

This study followed strict ethical procedures to safeguard participants' rights and well-being. Informed consent was obtained before the study began, with participants informed that their participation was voluntary and that they could withdraw at any time without penalty. Confidentiality was maintained throughout the research by anonymizing data using pseudonyms and removing any identifying details prior to analysis. Additionally, all audio and video recordings from interviews and focus groups were stored on a secure, encrypted drive accessible only to the principal researchers, ensuring the protection and integrity of the collected data.

Findings and Discussion

The findings are organized thematically according to the study's three research questions, beginning with quantitative results on the comparative effectiveness of the two interventions (RQ1). This is followed by qualitative insights from interviews and focus group discussions that explore student strategies (RQ2) and challenges (RQ3), enriching the interpretation of the statistical results.

Findings

Comparative Effectiveness of ChatGPT versus Lecturer Interaction

Inferential statistical tests were performed to address the first research question. This analysis compared the gain scores between the experimental and control groups across three primary measures: general English proficiency (English Proficiency Test), writing competency (argumentative essay), and self-efficacy. The descriptive statistics for both the experimental and control groups, detailing their performance on the three primary instruments during the pre-test and post-test phases, are presented in Table 1.

Table 2
Descriptive Statistics and Pre-test/Post-test Scores for All Measurements

Variable	Group	Time	*N*	*M*	*SD*
EPT Score	Experimental	Pre-test	30	70.15	6.1
		Post-test	30	88.4	5.95
	Control	Pre-test	30	70.3	6.25
		Post-test	30	75.5	6.15
Essay Score	Experimental	Pre-test	30	68.45	5.15
		Post-test	30	85.2	4.9
	Control	Pre-test	30	68.93	5.23
		Post-test	30	74.6	5.01
Self-Efficacy	Experimental	Pre-test	30	65.4	7.88
		Post-test	30	85.1	7.45
	Control	Pre-test	30	66.1	7.95
		Post-test	30	73.5	7.62

Table 2 provides the descriptive statistics for the three main variables measured in the study, comparing the experimental group and the control group at both the pre-test and post-test stages. The pre-test results show that the two groups had nearly identical baseline scores, confirming their equivalence before the intervention. For instance, the experimental group's mean score on the English Proficiency Test was 70.15 (SD = 6.10), closely matching the control group's 70.30 (SD = 6.25). Similar consistency was observed in the essay scores (M = 68.45 vs. 68.93) and self-efficacy scores (M = 65.40 vs. 66.10). However, post-test results reveal a notable outcome difference. Although both groups improved, the experimental group showed significantly greater gains, with mean post-test scores of 88.40 in proficiency, 85.20 in writing, and 85.10 in self-efficacy, compared to the control group's 75.50, 74.60, and 73.50, respectively.

Inferential statistical tests were applied to determine the significance of these learning gains. Paired-samples t-tests were used to examine within-group improvements, while an ANCOVA was conducted to compare between-group differences, using pre-test scores as covariates to control for initial differences. The detailed results are presented in Table 3.

Table 3

Summary of ANCOVA and t-test Results for Group Comparisons

Dependent Variable	Statistical Test	Value	*df*	Significance (*p*)	Effect Size (Partial $\eta 2$)
EPT Score (Gain)	ANCOVA (Between Groups)	F=22.14	1, 57	<0.001	0.28
	t-test Experimental	t=−14.82	29	<0.001	-
	t-test Control	t=−5.21	29	<0.001	-
Essay Score (Gain)	ANCOVA (Between Groups)	F=25.81	1, 57	<0.001	0.312
	t-test Experimental	t=−15.98	29	<0.001	-
	t-test Control	t=−6.89	29	<0.001	-
Self-Efficacy (Gain)	ANCOVA (Between Groups)	F=21.33	1, 57	<0.001	0.272
	t-test Experimental	t=−12.55	29	<0.001	-
	t-test Control	t=−5.67	29	<0.001	-

The ANCOVA results in Table 3 indicate statistically significant differences between the experimental and control groups across all measured outcome variables. Students in the ChatGPT group achieved greater post-test gains in general English proficiency ($F(1,57) = 22.14$, $p < .001$, $\eta^2 = 0.280$), writing performance ($F(1,57) = 25.81$, $p < .001$, $\eta^2 = 0.312$), and self-efficacy ($F(1,57) = 21.33$, $p < .001$, $\eta^2 = 0.272$). These effect sizes are considered significant, suggesting a strong association between the use of ChatGPT and positive learning outcomes. However, a cautious interpretation is required before attributing these gains solely to the instructional effectiveness of ChatGPT. The conditions under which the two groups interacted with their respective learning supports were not identical regarding responsiveness, accessibility, and degree of individualization.

During the 8-week intervention, students in the experimental group had unrestricted, asynchronous access to ChatGPT. They were encouraged to use the platform to complete weekly English tasks, seek grammar and vocabulary support, revise their drafts, and explore ideas through dialogic interaction. ChatGPT's interface allowed them to submit multiple prompts per session, iterate on responses, and receive instant feedback within seconds. This form of interaction enabled a high frequency of engagement and supported self-paced learning. Students could ask clarification questions, experiment with different

phrasings, and refine their work in real time. As a result, their learning process was marked by autonomy, immediacy, and flexible exploration of linguistic challenges, features that are rarely attainable in conventional learning environments.

In contrast, students in the control group relied on direct instruction and feedback from a single lecturer. Support was delivered during two weekly class meetings, totaling 90 minutes of synchronous contact. Additional feedback on written assignments was provided through asynchronous messages within 48 hours of submission. While the lecturer ensured task alignment and provided formative feedback per the course objectives, the interaction structure was necessarily more limited. Students had to wait for responses, and opportunities for spontaneous follow-up or iterative revision were restricted by the lecturer's availability and classroom time constraints. Furthermore, the mode of delivery tended to be unidirectional, with less emphasis on extended dialogic exchange.

Another important distinction lies in the level of personalization. ChatGPT responded to every student individually, adapting to the specificity of each prompt and enabling learners to scaffold their tasks in increasingly complex ways. In contrast, the lecturer needed to balance feedback across an entire class, often generalizing responses or addressing only the most salient issues due to time constraints. While the quality of lecturer feedback may have been higher regarding pedagogical intentionality, the volume and immediacy of support were not comparable.

These variations in availability, frequency of interaction, and feedback immediacy likely influenced the outcome differences. Under its high-access environment, the ChatGPT group was exposed to more cycles of practice, correction, and self-monitoring. This increased exposure to responsive feedback could account for the significantly higher improvements captured by the post-test scores. While the statistical analysis shows robust evidence of greater gains in the experimental group, these gains may reflect the functional affordances of the AI tool and the structural differences in support systems. Therefore, any interpretation of instructional effectiveness must be made considering these contextual dynamics.

The results suggest that ChatGPT functioned as an instructional tool and an accessible, responsive, and highly individualized learning environment. The experimental group's superior performance on proficiency, writing, and self-efficacy measures is likely the outcome of this complex interplay between technological affordances and interaction design. Thus, while the ANCOVA values confirm a significant differential impact, the underlying reasons extend beyond modality and include critical factors such as support availability, prompt-based learning control, and learner autonomy.

The Emergence and Evolution of Prompting Strategies

The analysis of interview data revealed that students did not uniformly use ChatGPT. Instead, they developed a range of prompting strategies that evolved. The analysis showed a clear developmental path in students' prompting practices. They moved from short and simple commands to longer prompts that included more context and detail. This change indicated that students became more aware of interacting effectively with ChatGPT. These strategies can be classified into two primary approaches. They are single, reformulated, and multiple-question strategies, as detailed in Table 4. These findings address the second research question.

Table 4

Categories and Descriptions of Student Prompting Strategies

Strategy Category	Strategy Name	Description
Single-Reformulated Strategies	Keyword Replacement	Students provide text or code with a placeholder and ask the AI to replace that specific keyword with the correct information or completion.
	Code Generation from Scratch	Instead of asking the AI to finish an incomplete piece of code, students prompt it to create the entire code based on specified inputs and outputs.
	Minimalist Prompting	A snippet of code or text is provided with little to no instructions, requiring the AI to infer the user's intent and the desired task.
	Focusing on the Core Idea	The prompt is refined to highlight the central concept or idea within a larger question, directing the AI to focus its response on that specific element.
	Contextual Expansion	Students provide detailed, comprehensive descriptions and relevant contextual information to help the AI generate a more precise and tailored response.
Multiple-Question Strategies	Role-Playing	Students assign a persona to the AI to simulate a human-like conversation, breaking down a problem into a series of sequential questions to guide the interaction.
	Question Segmentation	The main query is broken down into smaller, distinct parts, which are then posed to the AI one at a time to build a comprehensive answer incrementally.
	Divide and Conquer	A complex problem or code is deconstructed into more manageable, clearer units, with each unit being presented to the AI separately for analysis.
	Corrective Feedback	Students actively guide the AI toward the desired output by providing constructive feedback on its responses and correcting its course.
	Top-Down Exploration	The interaction begins with a broad topic to establish a foundational understanding, then progressively narrows to examine more specific details of the original query.
	Guided Generation	Rather than asking the AI to complete a task, students prompt it to generate a solution from scratch and then ask for specific refinements or modifications.

Table 4
Categories and Descriptions of Student Prompting Strategies (Cont.)

Strategy Category	Strategy Name	Description
	Requesting Verification	Students repeatedly ask the AI to confirm its answers, using prompts like "Are you certain?" or by rephrasing a previous response as a new question to test consistency.
	Key Point Identification	The core concept of a problem is identified, and the prompt is reformulated to focus the AI's attention on this key point while omitting extraneous details.
	Solution-Oriented Follow-Up	After receiving an initial answer, students ask follow-up questions to explore related concepts, alternative methods, or deeper explanations of the solution.
	Progressive Constraint Provision	An interaction starts with an open-ended query, and constraints (like a specific vocabulary list or format) are gradually added to refine the output.

The qualitative data analysis from interviews and FGDs revealed a progressive development in how students formulated prompts when interacting with ChatGPT. Initially, most students approached the tool with little understanding of how their language shaped the feedback. As the study advanced, students demonstrated increased control, purpose, and linguistic precision in their prompting—evolving from vague commands to highly contextualized, metacognitive interactions. These findings align with the strategic categories outlined in Table 4, but are best understood through the actual linguistic behavior observed during the intervention.

In the early weeks, prompting was typically minimal and unspecific, reflecting students' limited awareness of communicating with AI. In the excerpts below, the non-italicised text represents the exact prompts that students entered into ChatGPT, while the italicised text presents translated excerpts from the interview and FGD data that provide context for how students described their prompting practices. Many relied on simple instructions such as:

> "Fix this."
> "Make this better."

Participant 5 explained, "I only wrote short prompts like 'check my paragraph' but I didn't get what I needed. The answer was too general." Such minimalist prompts often made generic grammar corrections without addressing the core issue. As students received unhelpful responses, they began to adjust their strategy.

The next phase saw the emergence of better-formed prompts that explicitly stated purpose, topic, or desired tone. For example:

> "Please improve this paragraph about climate change. Focus on formal tone and coherence."

Participant 9 shared, "When I added the topic and told it what I wanted, the answer became much more useful. It felt more like real feedback."

These more detailed prompts correspond to what we categorize as Contextual Expansion—students provided richer input that framed their needs clearly. This resulted in more targeted and pedagogically relevant responses from ChatGPT.

Eventually, some students demonstrated advanced prompting, using multi-turn dialogic interaction to refine the AI's response. One student began with:

> "Can you help me rewrite this thesis statement?"
> Follow-up: "That sounds too informal. Can you use more academic connectors like 'in addition' or 'consequently'?"

Participant 3 commented, "It's like having a tutor. I don't just accept the answer—I reply, and the next version gets better."

In terms of vocabulary refinement, many students moved from broad commands like:

> "Change this word." to "Replace 'important' with a more precise academic synonym, like 'crucial' or 'significant'."

Participant 4 noted, "ChatGPT helped me realize 'good' is too basic. I asked for better alternatives, and now I use 'beneficial' or 'advantageous'."

Some categories from prior prompting literature—like Code Generation from Scratch—were adapted to reflect EFL student practices better. Instead of coding, students used these prompts to ask for model paragraphs:

> "Write a paragraph explaining why students should learn time management. Use formal tone and transition signals."

Participant 10 reflected, "I asked for a full paragraph because I didn't know how to start. Then I rewrote it in my own way. It's like using a map when you're lost."

Another emerging strategy was Top-Down Exploration, where students started with a broad conceptual prompt and then narrowed the scope as their understanding developed. For example:

> "What is online learning?"
> Follow-up: "How does online learning affect student motivation?"
> Then: "What are the challenges for students in rural areas?"

Participant 6 noted, "I started big and asked more specific questions. It helped me understand the topic step by step."

In contrast, Requesting Verification reflected students' growing critical thinking and caution toward AI responses. Rather than accepting the first output, students probed deeper:

> "Are you sure this structure is formal enough?"
> "Double-check the grammar of this sentence, please."

Participant 2 explained, "Sometimes I just didn't trust the answer. I asked again to be sure. It made me more confident in my writing."

Another important development was the use of Key Point Identification, where students asked the AI to help highlight essential ideas or remove irrelevant content:

> "What is the main idea of this paragraph?"
> "Remove unnecessary details and keep the core message."

Participant 8 stated, "I often write too long. Asking ChatGPT to simplify helped me focus on what really matters in academic writing."

Corrective Feedback, typically from the AI to the student, was reversed as students began critiquing ChatGPT's responses and instructing it to revise. For example:

> "This version is too wordy. Make it more concise."
> "You repeated 'students'—try a synonym in the second sentence."

Participant 14 reflected, "I used to just accept the answer, but now I can say 'no, this needs to be better.' It feels like I'm teaching the AI."

Students also became increasingly strategic in structuring their prompts. In the Divide and Conquer strategy, they isolated different components of an essay for targeted revision:

> "Check only the introduction."
> "Now improve the argument in the second body paragraph."
> "Help me conclude with a strong final sentence."

Participant 11 explained, "It's easier to fix one section at a time. I feel more in control."

The use of Role Assignment added another layer of prompting sophistication. Students could elicit responses based on the tone and depth expected from academic feedback by assigning a role to ChatGPT. For instance:

> "Act as a thesis supervisor. Evaluate this abstract."

Participant 1 said, "When I write that way, the response sounds more serious, unlike a grammar app."

Another higher-order technique was Solution-Oriented Follow-Up, in which students requested variations or alternatives rather than one definitive answer:

> "Give me two different ways to say this."
> "What's a better version of this sentence using passive voice?"

Participant 15 stated, "I don't just ask once—I explore different ways to express the same thing. It helps me learn style and structure."

Language choice emerged as a crucial factor. Although many students initially relied on Bahasa Indonesia for prompting, the majority shifted to English by mid-intervention. For example:

> Early prompt: "Tolong perbaiki kalimat ini supaya lebih akademik."
> Later prompt: "Please revise this sentence to meet formal academic standards."

Participant 6 explained, "When I used Indonesian, the answers were sometimes off. Switching to English made the response clearer and closer to what I need in academic writing." This shift occurred not through explicit instruction but due to student discovery. Prompting in English provided more accurate, context-appropriate, and stylistically aligned feedback. It also reinforced students' exposure to academic register, collocations, and coherence devices.

Overall, the emergence and evolution of prompting strategies reflected a trajectory from vague, L1-based, one-shot commands to structured, dialogic, and English-medium prompting with clear academic purposes. Students began to view ChatGPT not as a tool for quick fixes but as a responsive partner in their learning process. The ability to engage in effective prompting became both a linguistic practice and a metacognitive literacy—enabling learners to not only write better but also think better about writing.

Challenges in Prompting

The FGD and final interviews revealed challenges students faced when using ChatGPT's prompting. These challenges, which answer the third research question, were categorized into three dimensions: pedagogical, technical, and psychological, as summarized in Table 5.

Table 5
Categorization of Challenges in Prompting with ChatGPT

Category	Specific Challenge	Brief Description
Pedagogical	Over-reliance & Cognitive Offloading	Students use the AI as a shortcut, which hinders the development of their independent thinking and learning skills.
	Difficulty Evaluating Accuracy & Bias	Students struggle to discern whether the information provided by the AI is correct, biased, or factually sound.
	Receiving Generic/Shallow Feedback	Vague or general prompts from students often result in unhelpful and superficial responses from the AI.
Technical	Server Issues & Slow Responses	Unreliable server performance and slow response times make the tool frustrating and undependable for time-sensitive tasks.
	Inaccurate/Outdated Information (Hallucinations)	The AI can generate incorrect or entirely fabricated information, which erodes user trust and can lead to misinformation.
Psychological	Frustration from AI Misunderstanding	Students experience frustration when the AI fails to understand the nuance or intent of their complex questions.
	Lack of Interpersonal Connection	The interaction lacks the emotional support, encouragement, and relational nuance provided by a human teacher.
	Anxiety about Academic Integrity	Students feel persistent worry and uncertainty about whether their use of the AI constitutes cheating or plagiarism.

The most significant challenge identified was pedagogy, specifically the issues of over-reliance and cognitive offloading. Students frequently described ChatGPT as a quick fix rather than a learning tool, thereby outsourcing their cognitive effort. While efficient for completing assignments, this behavior undermines the development of foundational skills. One student confessed, “I stopped trying to remember grammar rules. I just paste the sentence and let ChatGPT fix it. It's faster, but I’m not sure I’m actually learning it.” This sentiment was common, highlighting a tendency to prioritize task completion over deep learning, the phenomenon researchers have termed “metacognitive laziness.”

Another major pedagogical hurdle was the Difficulty in Evaluating Accuracy and Bias. Students expressed a lack of confidence in the reliability of the AI’s output, recognizing that it could be flawed. “Sometimes it gives suggestions that sound good but turn out to be wrong when I check other sources,” one participant shared. “I can’t trust it 100%, especially for important information.” This points to a critical gap in students’ digital literacy skills, as they struggle to vet the information provided by the AI, leaving them vulnerable to incorporating inaccuracies or biases into their work. This was compounded by the challenge of Receiving Generic/Shallow Feedback, where

students found that the AI's advice was often too general to be useful without particular prompts.

On the technical front, while practical issues like Server Issues and Slow Responses were a source of frustration, a more profound problem was the AI's tendency to produce Inaccurate or Outdated Information, often referred to as "hallucinations." This unreliability was a significant concern. One student recounted a particularly jarring experience: "I once asked it to summarize a recent news article, and it made up an event that never happened. That was scary. I have never used it for facts since then." Such instances severely eroded trust in the tool for any fact-based queries and confined its use to more creative or opinion-based tasks for many students.

Finally, the psychological challenges revealed the limitations of human-AI interaction. A recurring theme was the frustration caused by AI Misunderstandings. Students felt that the AI often failed to grasp the nuance or context of their questions, leading to a communication breakdown. "It feels like talking to a wall. It doesn't really 'get' me," lamented one student. This was closely linked to a perceived Lack of Interpersonal Connection. The emotional and relational aspects of learning, such as encouragement and empathy, were notably absent. As one student said, "Getting feedback from a lecturer feels different. There's encouragement, there's understanding. With ChatGPT, it's just text. There's no emotion, no human connection." This emotional void was a significant drawback for many. Topping off the psychological burdens was a pervasive Anxiety about Academic Integrity. Students existed in a state of ethical uncertainty, constantly questioning where the line between assistance and cheating lay. "I'm always worried," a student admitted. "Is this considered cheating? Where is the line? I'm scared of using it too much and getting in trouble." This anxiety highlights the urgent need for clear institutional policies on the ethical use of AI in education.

Discussion

The quantitative results demonstrate that using ChatGPT for prompting significantly enhances students' English proficiency and academic self-efficacy. The experimental group outperformed the control group in essay writing, supporting earlier research showing that AI can improve writing quality (Bacon & Kraus, 2025; Oktarin et al., 2024). This improvement is primarily attributed to ChatGPT's capacity to deliver immediate, personalized, and iterative feedback on grammar, vocabulary, and mechanics, as noted by Waziana et al. (2024). Furthermore, the notable increase in self-efficacy among the experimental group aligns with findings that AI's consistent, non-judgmental availability creates a supportive learning environment that boosts student confidence and motivation (Hastomo et al., 2025).

The results indicate important benefits, but they need to be interpreted cautiously since the study focused only on proficiency, writing, and self-efficacy, and did not measure higher-order thinking skills such as argument development and critical reasoning. Future studies should examine these aspects to provide a more complete picture of AI-mediated learning. This limitation reflects a broader issue with current AI tools because students may depend on surface-level corrections that reduce opportunities for deeper engagement. Gerlich (2025) notes the risk of an "illusion of competence," where students appear skilled without achieving real understanding. It is therefore important to distinguish roles, as AI can function as a reliable proofreader while teachers remain necessary to foster critical thinking through interactive dialogue. Pragmatic and interpersonal skills, which were introduced earlier in the paper as areas where GenAI is limited, were also not assessed in this study. Their absence creates an important gap that future research should address to gain a fuller understanding of the educational impact of GenAI. Future studies should also examine whether differences in proficiency levels influence students' preferred language when interacting with ChatGPT, as this may provide further insight into the relationship between learner background and effective prompting practices.

The qualitative data reveal that the effectiveness of ChatGPT is shaped by how students formulate prompts. Participants progressed from using simple queries to engaging in complex, multi-turn interactions. This shift reflects prompting taxonomies discussed in the literature, particularly the move from basic "single-copy" prompts to more advanced "multiple-question prompting" strategies (Hwang et al., 2025). Students who employed sophisticated techniques, such as assigning ChatGPT a role or refining prompts iteratively, achieved deeper learning outcomes. This development reflects a transition in metacognitive engagement. Basic prompts indicate a passive learning approach, whereas complex prompts require planning, self-regulation, and critical thinking (Giray, 2023; Knoth et al., 2024). Thus, prompting should be understood as a technical function and an essential aspect of AI literacy. It is closely connected to learners' ability to manage their learning through reflection and strategic interaction (Cain, 2024; Mutanga et al., 2025). However, not all students acquired these skills independently, indicating that effective use of AI requires explicit instruction and guided practice.

This progression is further explained through the framework proposed by Bowen and Todd (2025), which emphasizes specificity, iteration, and validity in prompting. Initially, prompts lacked clarity and yielded vague responses. Over time, students generated more focused and purposeful prompts, leading to better outcomes. Some learners crafted prompts aligned with academic goals, while others showed over-reliance on AI, producing well-structured but unoriginal responses—a risk identified as "overfitting" within the framework.

In some cases, grammatically correct prompts still failed to achieve intended outcomes, raising concerns about prompt validity. These observations support the concept of "scaffolding inversion," where learners gradually take control of the learning process by refining AI-generated input. Prompting becomes a reflective academic skill that should be intentionally taught in EFL instruction.

The emergence of scaffolding inversion indicates a shift in learner roles, where students assume instructional responsibilities such as evaluating, adjusting, and directing AI-generated support. This dynamic aligns with Vygotsky's Zone of Proximal Development (ZPD), where learners achieve autonomy by internalizing support traditionally provided by more capable others (Vygotsky, 1978). Within AI-supported learning, students manage their scaffolding, transforming the teacher-student relationship. This process also aligns with constructivist principles, as learners co-construct knowledge through iterative interaction with the AI, promoting learner independence (Benson, 2011). However, this shift requires students to develop competencies in planning, evaluating, and regulating their learning skills, which are central to self-regulated learning and not automatically acquired (Zimmerman, 2002). These findings emphasize that AI literacy involves more than tool access; it requires developing pedagogical and metacognitive skills to manage learning support structures effectively (Ng et al., 2021). Framing prompting as a pedagogical responsibility signals the need to integrate AI training into educational design, enabling students to critically engage with AI in their learning (Mutanga et al., 2025).

A central challenge identified in the study is cognitive offloading, where students delegate thinking tasks to AI to minimize effort (Chen & Chang, 2024). This behavior, described as "metacognitive laziness" by Fan et al. (2025), reflects an overdependence on AI that undermines deep learning. Interviews revealed that some students stopped practicing grammar rules, relying entirely on the AI for corrections. While effective for short-term performance, this approach risks long-term skill deterioration due to a lack of mental engagement.

Additional concerns include academic integrity and the lack of emotional connection in AI-mediated learning. Students expressed uncertainty about where AI support ends and academic dishonesty begins, a dilemma that Perkins (2023) noted. Moreover, they reported that AI lacks the emotional encouragement and relational support human instructors provide. These concerns highlight the importance of maintaining human elements in education. Integrating AI must therefore be intentional, with strategies to reduce cognitive offloading and preserve the relational core of teaching and learning (Yusuf et al., 2024).

The study's findings align with posthumanist perspectives, which question clear divisions between human and machine contributions to learning. The collaborative writing observed between students and ChatGPT challenges

traditional ideas of authorship and authenticity (Luther et al., 2024). Some scholars argue that ChatGPT cannot be considered an author due to its lack of intentionality or accountability (van Woudenberg et al., 2024). The findings support a more nuanced interpretation, viewing student-AI interaction as a form of co-authorship.

Building on this, the study introduces the concept of scaffolding inversion to describe the altered pedagogical relationship in AI-mediated learning. Traditionally, instructors provide macro-scaffolding by setting goals and structuring tasks, while students carry out micro-tasks. In AI-supported environments, this structure reverses. The AI provides immediate micro-support, such as grammar correction, while the student assumes responsibility for the overall learning direction through prompt design. The student's ability to perform this macro-level role determines whether learning remains meaningful or becomes shallow and dependent. This concept offers a useful theoretical lens for understanding why prompt literacy is not a peripheral skill but a central component of effective learning in the AI era.

Conclusion

The findings suggest that ChatGPT may offer advantages over lecturer feedback in supporting students' writing development, particularly regarding availability and responsiveness. However, this should be interpreted with caution, as the results are context-specific and may not generalize beyond the sample, task type, or duration of the intervention. Second, the qualitative findings for the second research question revealed a clear developmental path in students' prompting abilities; learners progressed from simple, single-turn queries to more sophisticated, multi-turn dialogic strategies such as Role-Playing and Solution-Oriented Follow-Up, with this evolution being a key determinant of deeper learning. Finally, addressing the third research question, students encountered significant pedagogical, technical, and psychological challenges, with cognitive offloading emerging as the most critical pedagogical hurdle, alongside difficulties in vetting the AI's accuracy and pervasive anxiety regarding academic integrity.

These findings have several significant implications for pedagogy, policy, and future research. The most critical pedagogical implication is that the effective use of GenAI is not an innate skill but a learned competence. This necessitates a pedagogical shift from merely providing access to AI tools towards the explicit instruction of AI literacy and strategic prompting, as conceptualized in our scaffolding inversion model, where students must be taught to manage the macro-scaffolding of their learning. More specifically, prompt literacy instruction should guide students in formulating context-rich prompts, practicing iterative refinement through multi-turn interactions, and

critically evaluating AI-generated responses for accuracy and appropriateness. Institutionally, there is an urgent need for clear policies on the ethical use of AI to alleviate student anxiety. Nevertheless, the study's findings should be interpreted in light of its limitations, including its single-institution context and a sample predominantly composed of beginner-level learners, which may affect the generalizability of the results. Therefore, future research is warranted. Longitudinal studies are needed to track long-term skill retention, alongside comparative research on pedagogical interventions for teaching prompt literacy. Further investigation with more diverse student populations across various disciplines would also enhance the external validity of these findings.

Acknowledgements

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).

Biodata

Widi Andewi is a lecturer at Institut Bakti Nusantara in Lampung, Indonesia, focusing her research on English language education and educational technology. For inquiries, she can be reached via widiandewi.91@gmail.com.

Winia Waziana is a lecturer at Institut Bakti Nusantara in Lampung, Indonesia. Her research interests include ESP and educational technology. She can be contacted at email: winiawaziana@gmail.com.

Damar Wibisono is is a faculty member at Universitas Lampung. He earned his Bachelor's from Universitas Lampung and Master's from Universitas Gadjah Mada. Currently, he serves as the Head of the Sociology Department for the 2025–2029 term. His email is damar.wibisono@fisip.unila.ac.id.

Kristian Adi Putra is an Assistant Professor of English at Prince Sattam Bin Abdulaziz University, Saudi Arabia, where he teaches undergraduate EAP and ESP courses. Previously, he trained pre-service English teachers at Universitas Sebelas Maret, Indonesia, and taught Indonesian language and culture at the University of Montana and the University of Arizona, USA. He earned his Ph.D. in Second Language Acquisition and Teaching from the University of Arizona. His research focuses on Indigenous youth and multilingualism, language teacher education, and language planning and policy. Contact: kristianadiputra@gmail.com and k.putra@psau.edu.sa.

Tommy Hastomo is a lecturer at STKIP PGRI Bandar Lampung, Lampung. He is a doctoral student in the English Language Department of Universitas Negeri Malang. His research interest is ICT in ELT. He can be contacted at email: tomhas182@gmail.com.

Irene Brainnita Oktarin is a lecturer at STIE Gentiaras in Bandar Lampung, Indonesia. Her research interests include English language education and educational technology. She can be contacted at email: irenebrainnita.ib@gmail.com.

References

An, R., Yang, Y., Yang, F., & Wang, S. (2023). Use prompt to differentiate text generated by ChatGPT and humans. *Machine Learning with Applications*, *14*, Article 100497. https://doi.org/10.1016/j.mlwa.2023.100497

Bacon, E. D., & Kraus, H. (2025). Improving academic writing proficiency for EFL students: Leveraging ChatGPT using data-driven learning principles. *REFLections, 32*(1), 550–575. https://doi.org/10.61508/refl.v32i1.280410

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X, 1*(3), 1–9. https://doi.org/10.1002/brx2.30

Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education*, *20*(1), 1–20. https://doi.org/10.1186/s41239-023-00432-3

Benson, P. (2011). *Teaching and researching autonomy in language learning*. Pearson Education.

Bowen, N. E. J. A., & Todd, R. W. (2025). Enhancing ChatGPT-based writing research through effective prompt use. *Teaching English with Technology, 2025*(1), 26–40. https://doi.org/10.56297/vaca6841/OZFH1876/SFKW4394

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

Brinkmann, S., & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology*, *18*(2), 157–181. https://doi.org/10.1080/10720530590914789

Brookhart, S. M. (2018). *How to create and use rubrics for formative assessment and grading*. ASCD.

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

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

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Fan, L., & Cui, F. (2024). Mindfulness, self-efficacy, and self-regulation as predictors of psychological well-being in EFL learners. *Frontiers in Psychology*, *15*, Article 1332002. https://doi.org/10.3389/fpsyg.2024.1332002

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

George, D., & Mallery, M. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

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

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, *51*(12), 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Hastomo, T., Sari, A. S., Widiati, U., Ivone, F. M., Zen, E. L., & Andianto, A. (2025). Exploring EFL teachers' strategies in employing AI chatbots in writing instruction to enhance student engagement. *World Journal of English Language*, *15*(7), 93–102. https://doi.org/10.5430/wjel.v15n7p93

Hwang, M., Lee, K.-H., & Lee, H.-K. (2025). A word to the wise: Crafting impactful prompts for ChatGPT. *System*, *133*, Article 103756. https://doi.org/10.1016/j.system.2025.103756

Käser, T., & Schwartz, D. L. (2020). Modeling and analyzing inquiry strategies in open-ended learning environments. *International Journal of Artificial Intelligence in Education, 30*(3), 504–535. https://doi.org/10.1007/s40593-020-00199-y

Kerr, R. C., & Kim, H. (2025). From prompts to plans: A case study of pre-service EFL Teachers' use of generative ai for lesson planning. *English Teaching*, *80*(1), 95–118. https://doi.org/10.15858/engtea.80.1.202503.95

Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, *57*, 1–18. https://doi.org/10.1080/15391523.2025.2456043

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage.

Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Leung, C. H. (2024). Promoting optimal learning with ChatGPT: A comprehensive exploration of prompt engineering in education. *Asian Journal of Contemporary Education*, *8*(2), 104–114. https://eric.ed.gov/?id=EJ1461939

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

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Mahande, R. D., Fakhri, M. M., Suwahyu, I., & Sulaiman, D. R. A. (2025). Unveiling the impact of ChatGPT: Investigating self-efficacy, anxiety and motivation on student performance in blended learning environments. *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-07-2024-0372

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

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2,* Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nurchurifiani, E., Maximilian, A., Ajeng, G. D., Wiratno, P., Hastomo, T., & Wicaksono, A. (2025). Leveraging AI-powered tools in academic writing and research: Insights from English faculty members in Indonesia. *International Journal of Information and Education Technology*, *15*(2), 312–322. https://doi.org/10.18178/ijiet.2025.15.2.2244

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

Pallant, J. (2016). *SPSS Survival Manual*. Open University Press.

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), Article 07. https://doi.org/10.53761/1.20.02.07

Sawalha, G., Taj, I., & Shoufan, A. (2024). Analyzing student prompts and their effect on ChatGPT's performance. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2397200

Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence, 4*(1), Article 46. https://doi.org/10.1007/s44163-024-00143-2

van Woudenberg, R., Ranalli, C., & Bracker, D. (2024). Authorship and ChatGPT: A conservative view. *Philosophy & Technology*, *37*(1), Article 34. https://doi.org/10.1007/s13347-024-00715-1

Vera, F. (2024). Student performance in writing prompts for text-based GenAI tools in a research methodology course. *Revista Electrónica Transformar*, *5*(2), 71–75. https://revistatransformar.cl/index.php/transformar/article/view/129

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

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

Woo, D. J., Guo, K., & Susanto, H. (2025). Exploring EFL students' prompt engineering in human–AI story writing: An activity theory perspective. *Interactive Learning Environments*, *33*(1), 863–882. https://doi.org/10.1080/10494820.2024.2361381

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), Article 21. https://doi.org/10.1186/s41239-024-00453-6

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In A. Schmidt, K. Vaananen, T. Goyal, P. O. Kristensson, A. Peter, s. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581388

Zheng, Y., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Computers and Education: Artificial Intelligence*, *6*, Article 100223. https://doi.org/10.1016/j.caeai.2024.100223

Zhu, S., Wang, Z., Zhuang, Y., Jiang, Y., Guo, M., Zhang, X., & Gao, Z. (2024). Exploring the impact of ChatGPT on art creation and collaboration: Benefits, challenges and ethical implications. *Telematics and Informatics Reports*, *14*, Article 100138. https://doi.org/10.1016/j.teler.2024.100138

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zulianti, H., Hastuti, H., Nurchurifiani, E., Hastomo, T., Maximilian, A., & Ajeng, G. D. (2024). Enhancing novice EFL teachers' competency in AI-powered tools through a TPACK-based professional development program. *World Journal of English Language*, *15*(3), 117. https://doi.org/10.5430/wjel.v15n3p117