eISSN 2721-5016 | pISSN 2721-5024
International Conference on English Language Teaching
http://conference.unisma.ac.id/index.php/ICON/ICON-ELT
Volume 6 | Number 1 | p. 31-39
Published on November 24th, 2025

Does ChatGPT Enhance or Hinder Critical Thinking? Non-English Department Students' Perception in EFL Classrooms

Faridatul Istighfaroh
Public Administration Department, Universitas Bojonegoro, Indonesia
(*) correspondence: faridatulistighfaroh@gmail.com

ABSTRACT

This experimental study investigates the impact of ChatGPT, an AI-powered language model, on the development of critical thinking skills in English language learning among Public Administration department students. Amid the growing integration of artificial intelligence in educational settings, the research aims to determine whether ChatGPT functions as a tool that enhances or hinders EFL learners' cognitive engagement and analytical reasoning in non-English major. A total of 186 undergraduate students participated in this study. The quasi-experimental design is used and data were collected through pre-test and post-test control group structure measuring critical thinking ability in EFL classroom. Across proficiency levels, the analyses revealed that the use of ChatGPT in EFL classrooms did not yield statistically significant improvements in students' critical thinking skills. These findings suggest that the integration of ChatGPT neither significantly enhanced nor hindered students' critical thinking performance within the context of this study. The study contributes to the ongoing discourse on AI in education and provides pedagogical recommendations for integrating AI tools to foster higher-order thinking in EFL contexts, especially in non-English departments.

Keywords: ChatGPT; critical thinking; artificial intelligence; English learning; non-English department students; EFL education

INTRODUCTION

The integration of artificial intelligence (AI) into educational settings has transformed traditional pedagogical approaches, particularly in language learning environments. Among these AI tools, ChatGPT, developed by OpenAI, has gained prominence for its ability to generate human-like text, offering potential benefits for English as a Foreign Language (EFL) learners. Its capabilities in providing instant feedback and facilitating language practice have positioned it as a valuable resource in enhancing students' writing skills and critical thinking abilities (Nguyen Minh, 2024).

However, the use of AI tools like ChatGPT also raises concerns regarding over-reliance and the potential erosion of independent critical thinking. Some researchers have highlighted the risk that students may become dependent on AI-generated content, leading to diminished engagement in the cognitive processes essential for critical analysis (Melisa et al., 2025). Moreover, inaccuracies in AI outputs and the lack of contextual understanding can pose challenges in educational settings(Zhang & Kim, 2024).

Critical thinking is a fundamental skill in higher education, enabling students to analyze information, evaluate arguments, and construct coherent reasoning. In the context of non English department like Public Administration in this study, where policy analysis and decision-making are paramount, fostering critical thinking is essential. The application of ChatGPT in EFL instruction has shown promise in supporting these skills. For instance, studies have indicated that ChatGPT can aid in developing students' argumentative writing by providing structured feedback and promoting analytical thinking (Suh et al., 2025).

Despite these concerns, the strategic implementation of ChatGPT, coupled with educator guidance, can mitigate potential drawbacks. Integrating AI tools into the curriculum with a focus on enhancing, rather than replacing, critical thinking processes is crucial. Educators play a vital role in

framing AI as a supplementary tool that encourages students to question, analyze, and reflect, thereby reinforcing their critical thinking skills (Harahap, 2024).

Furthermore, the students' perceptions of ChatGPT in a pedagogical context have been rising (Firdaus et al., 2025). Some students found it interesting and motivating (Shoufan, 2023), others benefit from ChatGPT for idea generation and language support, while also noting concerns about overreliance, accuracy, and ethical use (Lahby, 2024). The employment of ChatGPT in higher education settings have been massively explored in terms of learning satisfaction (Almulla, 2024), academic honesty and integrity (Nebieridze, 2024), potential and challenges in education (Ahmad et al., 2021), and value and convenience (Cavazos et al., 2025). In EFL context, ChatGPT has been used in various skills such as pragmatic development among EFL students (Tahir, 2025), the need for balanced, guided use to maximize benefits and minimize risks of ChatGPT in EFL contexts (Amin, 2023). On the other hand, the study regarding ChatGPT use focusing on critical thinking for non-English department students is still underexplored.

Therefore, this study aims to investigate the impact of ChatGPT on the development of critical thinking skill among non-English department students engaged in English language learning. Through an experimental design, the research will assess whether incorporating ChatGPT enhances or hinders students' critical thinking abilities, providing insights into effective pedagogical strategies for integrating AI in EFL contexts.

THEORETICAL REVIEW

This study is guided by the theory of critical thinking. Critical thinking has long been considered a central goal in higher education, including in English as a Foreign Language (EFL) classrooms. According to (Facione, 1990), critical thinking is defined as "purposeful, self-regulatory judgment" that involves interpretation, analysis, evaluation, and inference. (Paul & Elder, 2014) emphasize that critical thinking requires questioning assumptions, considering multiple perspectives, and developing reasoned arguments. In this study, critical thinking is used to measure students' cognitive skills in the EFL learning within non-English department settings.

Furthermore, in EFL contexts, critical thinking is often intertwined with language skills because learners must analyze texts, evaluate arguments, and construct logical responses in a second language (Atkinson, 1997). Studies have shown that promoting critical thinking in EFL classrooms not only improves students' language proficiency but also enhances their problem-solving abilities and academic literacy (Shirkhani & Fahim, 2011). For non-English department students, however, the challenge is twofold: they must process complex ideas while simultaneously overcoming linguistic limitations.

In addition, this study emphasizes the use of Artificial Intelligence (AI), and more specifically Large Language Models (LLMs) like ChatGPT, has introduced new possibilities in language education. ChatGPT is designed to generate human-like text responses, assist with writing, and simulate conversational practice (Bansal et al., 2024). Prior studies suggest that AI-powered tools can facilitate vocabulary development, provide immediate feedback, and scaffold learning (Kohnke et al., 2023). In this study, ChatGPT can act as a conversational partner, writing assistant, or knowledge resource. It offers learners opportunities for practice beyond the classroom and can lower anxiety by providing a non-judgmental interlocutor. However, scholars also caution that overreliance on AI tools may limit learners' autonomy, reduce engagement in authentic communication, and risk fostering superficial learning (Kasneci et al., 2023).

In terms of students' perception, this study examines the learners' perception on ChatGPT use during their EFL classroom activities. Particularly, perceptions of technology strongly influence its educational effectiveness. Davis' (1989) Technology Acceptance Model (TAM) highlights that perceived usefulness and ease of use determine whether learners adopt new technologies. In EFL classrooms, students' attitudes toward AI tools affect their willingness to engage critically with them (Zou et al., 2025). Non-English department students may perceive ChatGPT differently compared to English major students. Their primary motivation for using AI may be to simplify tasks or overcome

linguistic barriers rather than to refine reasoning skills. Therefore, examining their perceptions can reveal whether ChatGPT is functioning as a cognitive amplifier or as a cognitive crutch.

METHODS

Research Context

This study was conducted in the context of English for Academic Purposes (EAP) courses at a private university in Indonesia. These courses are compulsory for non-English department students and aim to develop academic reading, writing, and discussion skills in English. Students typically have diverse language backgrounds and varying levels of English proficiency, which range from beginner to advanced.

The integration of ChatGPT was introduced in response to the increasing presence of generative AI tools in higher education. ChatGPT was purposely selected because of its accessibility, popularity among students, and its ability to generate instant, context-sensitive responses to prompts. The classroom intervention was designed not only to expose students to this emerging technology but also to examine whether its use would stimulate or hinder critical thinking during academic tasks.

Research Design

This study employed a quasi-experimental design with a pre-test and post-test control group structure. The purpose was to investigate whether the integration of ChatGPT in EFL classrooms enhanced or hindered the critical thinking skills of non-English department students. Since random assignment was not feasible in the educational setting, intact classes were used as groups, making the design quasi-experimental in nature.

Participants

The participants were 186 undergraduate students from a non-English department. Participation was voluntary, and informed consent was obtained prior to data collection. To protect anonymity, participants were not asked to provide personal identifiers. Instead, they self-reported their English proficiency level as *Beginner* (n = 108), *Intermediate* (n = 76), or *Advanced* (n = 2). The unequal distribution across levels reflects the natural composition of the classes.

Instruments

Two main instruments were used in this study**.** A 20-item critical thinking questionnaire was developed based on Facione's (1990) Delphi Report on critical thinking skills (analysis, evaluation, inference, and reasoning). Items were rated on a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The questionnaire was administered as both pre-test and post-test. The second is instructional intervention**.** The experimental treatment involved the integration of ChatGPT into classroom activities. Students in the treatment sessions were instructed to use ChatGPT in EFL classroom to brainstorm ideas, evaluate arguments, and generate solutions for class discussion topics. Lecturer guided the students in interacting with ChatGPT while encouraging them to critique, verify, and reflect on the generated responses.

Procedure

The study was conducted over a six-week instructional period. In the first week, students completed the pre-test questionnaire. In the following five weeks, the intervention was implemented during the EFL classes. In the final week, students completed the post-test questionnaire.

Data Collection

Pertest and post-test data were collected anonymously. Instead of tracking individuals across time using identifiers, the data were grouped based on self-rated English proficiency (Beginner, Intermediate, Advanced). This approach allowed for group-level comparison but did not allow for paired-sample analysis.

Data Analysis

Data were analyzed using SPSS version 26. The analysis proceeded in several steps:

1. Normality Test
   Kolmogorov-Smirnov and Shapiro-Wilk tests were conducted to assess whether the data were normally distributed for each proficiency level.
2. Inferential Statistics
   - For the Beginner group (data not normally distributed), a Mann-Whitney U test was used to compare pre-test and post-test scores.
   - For the Intermediate group (data normally distributed), an Independent Samples t-test was conducted.
   - For the Advanced group (n = 2), only descriptive statistics (mean, standard deviation, range) were reported due to the small sample size.
3. Effect Size
   For the t-test, effect sizes were calculated using Cohen's *d* to determine the magnitude of differences between groups. All tests used a significance threshold of $p < .05$.

RESULTS AND DISCUSSIONS

The data gathered were firstly tested for the normality. After the normality test was conducted, the data were calculated based on the groups (beginner, intermediate and advanced). The last step, the effect size was evaluated to see the differences between groups.

Test of Normality

The normality of students' critical thinking scores was tested across three self-rated English proficiency levels: Beginner, Intermediate, and Advanced (Table 1).

Table 1

Test of Normality of Critical Thinking Scores by English Level

English Level	Kolmogorov-Smirnov (p)	Shapiro-Wilk (p)	Conclusion
Beginner (n=108)	0.020	0.019	Not Normal
Intermediate (n=76)	0.200	0.847	Normal
Advanced (n=2)	–	–	Not Testable

As shown in Table 1, the scores of students at the Beginner level were not normally distributed ($p < .05$ for both Kolmogorov-Smirnov and Shapiro-Wilk tests). In contrast, the Intermediate group's scores were normally distributed ($p > .05$). The Advanced group, however, only consisted of two students, making it impossible to conduct a valid normality test. Based on these results, subsequent analyses were conducted using non-parametric tests for the Beginner group (Mann-Whitney U test), parametric tests for the Intermediate group (Independent Samples t-test), and descriptive statistics for the Advanced group.

Comparison of Pretest and Posttest Scores

Beginner level (Mann-Whitney U Test)

Table 2

Mann-Whitney Test for Beginner Group

Ranks				
	Test_Type	N	Mean Rank	Sum of Ranks
Score	Pretest	94	90.70	8525.50
	Posttest	92	96.36	8865.50

Test Statistics[a]	
	Score
Mann-Whitney U	4060.500
Wilcoxon W	8525.500
Z	-.718
Asymp. Sig. (2-tailed)	.473
a. Grouping Variable: Test_Type	

The Mann-Whitney U test was employed to examine differences in pre-test and post-test scores of students at the Beginner level. The table show**s** Mann-Whitney U is 4060.500, Z = -0.718 and significant value of 0.473 (higher than the threshold of >.05) which reveals there is no significant impact of ChatGPT integration on the critical thinking competence within EFL learning context, especially for beginner-level students in non-English department.

Intermediate level (independent samples t-test)

Table 3

Independent Samples Test

Independent Samples Test											
		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Score	Equal variances assumed	.456	.500	-.606	184	.273	.545	-1.002	1.654	-4.266	2.261
	Equal variances not assumed			-.605	179.435	.273	.546	-1.002	1.656	-4.271	2.266

The results of the independent samples t-test revealed that Levene's test for equality of variances was not significant (F = .456, p = .500), indicating that the assumption of equal variances was met. Therefore, the "equal variances assumed" row was used for interpretation. The t-test result showed

no statistically significant difference in scores between the pre-test and post-test (t = –.606, df = 184, p = .545, two-tailed). The mean difference between the groups was –1.002, with a 95% confidence interval ranging from –4.266 to 2.261. Since the confidence interval includes zero and the p-value is greater than .05, it can be concluded that the students' critical thinking did not differ significantly in their performance with or without ChatGPT for the Intermediate group. Although the post-test mean was slightly higher, the effect was negligible, indicating that the use of ChatGPT did not produce a statistically significant enhancement in students' critical thinking.

Advanced Level

Table 3

Descriptive Analysis

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Score	186	19	90	60.60	11.259
Valid N (listwise)	186				

It should be noted that the Advanced group consisted of only two participants, and thus their scores were analyzed descriptively rather than inferentially. While the descriptive findings provide some indication of performance, the extremely limited sample size prevents any generalizable conclusion.

Table 4

Independent Samples Effect Sizes

Independent Samples Effect Sizes					
				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Score	Cohen's d	11.278	-.089	-.376	.199
	Hedges' correction	11.324	-.089	-.375	.198
	Glass's delta	12.041	-.083	-.371	.205

In addition to the t-test, effect sizes were calculated to assess the magnitude of the difference between English level groups. The results showed a very small effect, with Cohen's d = –.089 (95% CI [–.376, .199]), Hedges' g = –.089 (95% CI [–.375, .198]), and Glass's delta = –.083 (95% CI [–.371, .205]). These effect sizes are close to zero, indicating a negligible difference between the groups. Combined with the non-significant t-test results, this suggests that English level did not substantially influence the participants' scores.

These results suggest that the mere integration of ChatGPT into classroom activities does not automatically translate into measurable gains in critical thinking. Several possible explanations may account for these findings. First, students' level of familiarity with ChatGPT may have influenced the outcomes. For many participants, ChatGPT represented a novel tool, and part of the learning process was likely dedicated to exploring its functions rather than engaging deeply in reflective and evaluative thinking. Previous studies have shown that the introduction of new technologies in education often requires an adaptation period before their pedagogical benefits can be fully realized (Li & Ma, 2010; Lund & Wang, 2023).

Second, critical thinking requires active cognitive engagement, such as analyzing arguments, evaluating evidence, and synthesizing ideas. While ChatGPT can provide instant responses, students at the beginner and intermediate levels may have relied too heavily on the AI-generated answers rather than questioning or critically evaluating them. This reliance could have limited the opportunities for

deeper reasoning and intellectual struggle, which are central to the development of critical thinking (Ennis, 2018).

Third, the language barrier may also have played a role. Since participants were non-English department students, varying levels of linguistic competence may have constrained their ability to fully engage with ChatGPT's output in English. For lower-level learners, the cognitive load of understanding ChatGPT's responses in a foreign language might have overshadowed the cognitive effort required for critical evaluation. This finding resonates with earlier research suggesting that linguistic proficiency strongly mediates the development of higher-order thinking skills in EFL contexts (Kuhn, 2019).

It is also important to note that the absence of significant differences does not necessarily imply that ChatGPT has no potential in fostering critical thinking. Rather, the results highlight the importance of pedagogical scaffolding in AI integration. Simply exposing students to ChatGPT may not suffice; instructors need to design guided tasks that explicitly encourage students to question, challenge, and critique AI-generated responses. For instance, structured activities such as "fact-checking ChatGPT" or "debating with ChatGPT" could push students beyond passive consumption toward active critical engagement.

Finally, the findings also raise questions about the possibility that ChatGPT may not enhance but instead hinder critical thinking if used uncritically. Students who treat ChatGPT's responses as authoritative may become less inclined to engage in independent reasoning. This risk aligns with concerns raised in the literature that generative AI, while useful, can inadvertently promote intellectual dependency if not pedagogically moderated (Kasneci et al., 2023).

CONCLUSION

This study examined whether the use of ChatGPT in EFL classrooms enhanced or hindered the critical thinking skills of non-English department students across different proficiency levels. The findings indicated no statistically significant improvements in students' critical thinking performance from pre-test to post-test, both at the beginner and intermediate levels. For advanced-level students, the small number of participants prevented meaningful inferential analysis.

Although the results suggest that ChatGPT, as implemented in this study, did not significantly enhance students' critical thinking, they also highlight important pedagogical implications. The integration of generative AI tools such as ChatGPT requires careful instructional design and scaffolding to promote active engagement, rather than passive reliance on AI-generated content. Without such guidance, students may benefit from the convenience of AI but miss the opportunity to develop higher-order thinking skills.

In light of these findings, educators are encouraged to adopt strategies that position ChatGPT as a tool for inquiry, critique, and debate rather than as a source of ready-made answers. Future research should employ longitudinal and task-based designs, explore variations in scaffolding strategies, and investigate how different levels of language proficiency mediate students' engagement with AI tools. Overall, this study contributes to the growing body of research on AI in education by showing that the role of ChatGPT in fostering critical thinking is not automatic, but contingent on pedagogical context and instructional implementation.

REFERENCES

Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M., & Hyder, S. I. (2021). Artificial intelligence and its role in education. *Sustainability*, *13*(22), 12902. https://doi.org/10.3390/su132212902

Amin, M. Y. M. (2023). AI and ChatGPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies*, *4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly*, *31*(1), 71. https://doi.org/10.2307/3587975

Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2024). Transforming conversations with AI—A Comprehensive study of ChatGPT. *Cognitive Computation*, *16*(5), 2487–2510. https://doi.org/10.1007/s12559-023-10236-2

Cavazos, J. T., Hauck, K. A., Baskin, H. M., & Bain, C. M. (2025). ChatGPT goes to college: Exploring student perspectives on artificial intelligence in the classroom. *Teaching of Psychology*, *52*(3), 357–368. https://doi.org/10.1177/00986283241268829

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, *37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Facione, P. A. (1990). *The California Critical Thinking Skills Test--College Level. Technical Report #1. Experimental Validation and Content Validity.*

Firdaus, T., Mufidah, R., Hamida, R. N., Febrianti, R. I., & Guivara, A. E. R. (2025). Research trends and the impact of ChatGPT on educational environments. *International Journal of Social Sciences and Education Research*, *11*(1), 45–62. https://doi.org/10.24289/ijsser.1609741

Harahap, D. S. (2024). Implementation of ChatGPT to improve students' critical thinking abilities. *Indonesian Journal of Education and Social Humanities*, *1*(2), 33–39. https://doi.org/10.62945/ijesh.v1i2.58

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, *5*, 100156. https://doi.org/10.1016/j.caeai.2023.100156

Kuhn, D. (2019). Critical thinking as discourse. *Human Development*, *62*(3), 146–164. https://doi.org/10.1159/000500171

Lahby, M. (2024). *Empowering Digital Education with ChatGPT*. Chapman and Hall/CRC. https://doi.org/10.1201/9781032716350

Li, Q., & Ma, X. (2010). A Meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, *22*(3), 215–243. https://doi.org/10.1007/s10648-010-9125-8

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/10.1108/LHTN-01-2023-0009

Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Ero, P. E. L., Marlini, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical Thinking in the age of AI: A systematic review of AI's effects on higher education. *Educational Process International Journal*, *14*(1). https://doi.org/10.22521/edupij.2025.14.31

Nguyen Minh, A. (2024). Leveraging ChatGPT for enhancing english writing skills and critical thinking in university freshmen. *Journal of Knowledge Learning and Science Technology ISSN: 2959-*

6386 (Online), *3*(2), 51–62. https://doi.org/10.60087/jklst.vol3.n2.p62

Paul, R., & Elder, L. (2014). Critical thinking: Intellectual standards essential to reasoning well within every domain of human thought, Part 4. *Journal of Developmental Education; Boone*, *37*(3), 34.

Shirkhani, S., & Fahim, M. (2011). Enhancing critical thinking in foreign language learners. *Procedia - Social and Behavioral Sciences*, *29*, 111–115. https://doi.org/10.1016/j.sbspro.2011.11.214

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access*, *11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

Suh, S., Bang, J., & Han, J. W. (2025). Developing critical thinking in second language learners: Exploring generative AI like ChatGPT as a tool for argumentative essay writing. *ArXiv Preprint*.

Zhang, Y., & Kim, J. (2024). Exploring the complexities of professional identity formation among English teachers at a Korean university. *Sage Open*, *14*(2). https://doi.org/10.1177/21582440241253569

Zou, D., Xie, H., & Kohnke, L. (2025). Navigating the future: Establishing a framework for educators' pedagogic artificial intelligence competence. *European Journal of Education*, *60*(2). https://doi.org/10.1111/ejed.70117