



The Role of ChatGPT in Enhancing Argumentative Writing Skills of EFL Students through Formative Feedback

VIRGIN SETIAWAN

Universitas 17 Agustus 1945 Surabaya

Email: virgansetiawan88@gmail.com**PARIYANTO**

Universitas 17 Agustus 1945 Surabaya

Email: priyanto@untag-sby.ac.id

Abstract. This research aims to analyze the role of ChatGPT's formative feedback in improving the quality of EFL Students' argumentative essay writing. The methodology involved a quantitative design with pre-feedback and post-feedback argumentative essays written by 31 participants. These essays were assessed by the argumentative writing rubric from SERP Institute which focusing in four aspects; argumentation, evidence, organization, and language. Scores were rated by two raters, with high interrater reliability confirmed by ICC value of .950. The result of this research revealed statistically improvement in the rating of argumentative essays after receiving formative feedback from ChatGPT, with the evidence aspect showing the highest mean increase (5.8387 points). This is based on the .Sig value of paired sample test of .000, it shows a significant improvement. Additionally, based on the interview responses, it is revealed that after utilizing the feedback from ChatGPT, the participants' overall showed positive perceptions of ChatGPT, highlighting its usefulness in identifying weaknesses, improving structure, and generating ideas, although students noted that clear prompts were necessary.

Keywords: *Argumentative Writing, ChatGPT, EFL Learners, Formative Feedback, Interrater Reliability*

INTRODUCTION

Writing proficiency in English as a Foreign Language (EFL) represents a critical academic skill, particularly in argumentative writing where students must present well-founded assertions supported by empirical evidence and logical reasoning to persuade readers toward specific viewpoints (Su, Lin, & Lai, 2023). Argumentative essays require clear claims supported by reasoning and evidence, acknowledgment of counter-arguments, and integration leading to substantiated conclusions (Latifi, Noroozi, Hatami, & Biemans, 2021). However, EFL students often struggle with developing these complex writing skills due to limited exposure to effective feedback mechanisms.

Traditional feedback approaches, while valuable, face significant limitations in providing immediate, consistent, and personalized responses to student writing. Teacher feedback, though expert-driven, is often delayed and may not address individual learning needs comprehensively. Peer feedback lacks consistency and may miss critical writing elements, while self-assessment is limited by students' inability to recognize their own

errors (Bozorgian & Yazdani, 2021). These constraints highlight the need for innovative feedback mechanisms that can provide immediate, objective, and customizable responses to support EFL argumentative writing development.

The emergence of Artificial Intelligence (AI) large language models, particularly ChatGPT, presents unprecedented opportunities for addressing these pedagogical challenges. However, the integration of AI in educational contexts requires careful consideration, as concerns about misuse; including academic dishonesty, over-reliance, and the potential for superficial learning, have emerged across educational institutions. When implemented thoughtfully with appropriate pedagogical frameworks, ChatGPT can serve as a powerful educational tool that enhances rather than replaces traditional learning processes (Bom, 2023).

Recent studies have demonstrated ChatGPT's potential in academic writing contexts when used judiciously. Mahapatra (2024) found positive impacts on ESL students' writing abilities and favorable attitudes toward AI-assisted learning, while Zebua and Katemba (2024) reported largely positive student perspectives on ChatGPT's effectiveness in enhancing writing skills. When employed as a formative feedback mechanism rather than a content generator, ChatGPT offers several pedagogical advantages including instant feedback, objective assessment, personalized responses, and the ability to guide students in identifying grammatical and structural improvements.

Despite these promising developments, research examining ChatGPT's specific effectiveness in improving argumentative writing skills among EFL students remains limited. While existing studies explore general writing improvement, there is insufficient evidence regarding which specific aspects of argumentative writing benefit most from AI-generated formative feedback, and how students perceive this technological intervention in their learning process.

This study addresses this research gap by investigating the effectiveness of ChatGPT-generated formative feedback in enhancing argumentative writing skills among English Literature students. Specifically, it examines: (1) whether ChatGPT feedback enhances overall argumentative writing performance, (2) which specific aspects of argumentative writing show the most significant improvement, and (3) students' perceptions and attitudes toward AI-generated feedback. By focusing on argumentative writing, a fundamental academic skill reflecting critical thinking and logical organization, this research contributes to understanding AI's role in supporting EFL writing pedagogy and offers practical insights for educators seeking to integrate technological tools in language instruction.

LITERATURE REVIEW

Argumentative Writing in EFL Context

Argumentative writing represents a complex cognitive process requiring students to present well-founded assertions supported by empirical evidence and logical reasoning (Su, Lin, & Lai, 2023). According to Hyland (2009), this genre demands coherent argument construction through logical organization, evidence utilization, and counter-

argument addressing. For EFL learners, argumentative writing poses additional challenges as it requires not only content generation but also mastery of linguistic and rhetorical forms in the target language (Connor, 1996).

The complexity of argumentative writing necessitates systematic assessment approaches. The Strategic Education Research Partnership (SERP) Institute's rubric provides a comprehensive framework evaluating four key dimensions: Argumentation (claim clarity and reasoning), Evidence (sufficiency and relevance), Organization (logical structure and coherence), and Language (academic register and accuracy). This multidimensional approach enables detailed analysis of writing development across specific skill areas.

Second Language Acquisition and Feedback

Long's Interaction Hypothesis (1996) emphasizes that meaningful interaction, particularly involving negotiation of meaning and corrective feedback, facilitates language development. When learners encounter communication difficulties, feedback prompts clarification requests, language reformulation, and heightened linguistic awareness, contributing to comprehensible input intake and form-focused attention. This theoretical framework supports AI-generated feedback implementation, as it provides immediate, consistent corrective input that promotes noticing and revision processes essential for language acquisition.

Computer-Assisted Language Learning Framework

Chapelle's (2008) CALL framework emphasizes the importance of connecting technology use with SLA theory, particularly the interactionist perspective. Effective CALL applications should facilitate meaningful communication, provide responsive feedback, and create opportunities for language improvement through interactive tasks. ChatGPT's implementation as a formative feedback tool aligns with these principles by offering immediate, personalized responses that encourage active engagement with writing revision processes.

METHOD

This study employed a mixed method approach, which integrates both quantitative and qualitative research methods to gain a comprehensive understanding of the research problem. According to Creswell (2014), mixed method research is beneficial when a single data source is insufficient to explain a complex educational issue. Therefore, this approach was selected to examine not only the effectiveness of ChatGPT's formative feedback on students' argumentative writing but also their perceptions and experiences when interacting with AI-generated feedback.

The sources of data include the results of argumentative texts written by 31 participants from the English Literature Department at Universitas 17 Agustus 1945 Surabaya. All participants were required to have completed the Paragraph Writing, Essay Writing, and Academic Writing courses prior to participating in the study. The participants consisted

of 19 sixth-semester students and 12 eighth-semester students. Of the total, 24 students participated as part of their Creative Writing class, while the remaining 7 students took part outside the classroom setting. The research data were collected and analyzed from both the pre-feedback and post-feedback writing samples produced by the participants.

The data analysis process followed a systematic four-step approach to evaluate the effectiveness of ChatGPT feedback on argumentative writing skills. Initially, participants were instructed to compose approximately 250-word argumentative essays on the topic "The Impact of AI on Education." Following essay collection, the texts were submitted to ChatGPT using the standardized prompt: "Please provide formative feedback to improve the quality of argumentation, evidence, organization, and language in the following essay." Participants then revised their essays based on the AI-generated feedback. Subsequently, both pre-feedback and post-feedback essays were compiled and organized in an Excel spreadsheet for systematic analysis. Finally, all essays underwent comprehensive evaluation by three raters; the researcher, an additional human rater, and ChatGPT, using the argumentative essay rubric developed by The SERP Institute (2021), ensuring triangulation of assessment and reliability of scoring across the four key dimensions of argumentative writing: argumentation, evidence, organization, and language.

Intraclass Correlation Coefficient							
Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0				
	Lower Bound	Upper Bound	Value	df1	df2	Sig	
Single Measures	.905 ^a	.814	.953	20.626	30	30	.000
Average Measures	.950 ^c	.897	.976	20.626	30	30	.000

Table 1. Intraclass Correlation Coefficient

Inter-rater reliability was calculated using the Intraclass Correlation Coefficient (ICC) based on a two-way mixed-effects model with absolute agreement. The result showed an average measures ICC of 0.950, indicating excellent reliability between the two raters (Koo & Li, 2016). This suggests that the scores used for both pre- and post-feedback writing tasks were highly consistent and reliable for further analysis.

RESULT AND DISCUSSION

The analysis of the role of ChatGPT AI-generated formative feedback on the essay written by EFL English Literature students is conducted through a task of writing argumentative essays about the impact of AI on education. The results serve notable improvements in some aspects of their writing.

It has been found that, initially, based on the argumentative essay rubric, EFL Students of varying demographics demonstrated some expected common issues such as confusing claim, no evidence is presented, and use academic language forms incorrectly. Although, there is some improvement in the quality of argumentation, evidence, organization and language after revising the essay using ChatGPT's formative feedback. The rating process is done by the researcher and also assisted by the other rater. The raters refer to the argumentative writing rubric by SERP Institute (2021) for the several criteria for writing argumentative essays.

Pre-feedback Essay Scores

Table 2. Pre-feedback Scores

Pre-feedback	RATER 1	RATER 2
Essay 1	55	60
Essay 2	39	42
Essay 3	73	74
Essay 4	76	78
Essay 5	81	76
Essay 6	40	42
Essay 7	48	42
Essay 8	55	54
Essay 9	44	42
Essay 10	52	54
Essay 11	70	72
Essay 12	66	60
Essay 13	40	42
Essay 14	61	66
Essay 15	54	60
Essay 16	69	66
Essay 17	66	60
Essay 18	64	60
Essay 19	38	42
Essay 20	38	30
Essay 21	54	42
Essay 22	50	36
Essay 23	63	72
Essay 24	34	30
Essay 25	36	24
Essay 26	62	60
Essay 27	40	36
Essay 28	36	36
Essay 29	66	72
Essay 30	49	48
Essay 31	48	54
Mean	53.7742	52.6452
Total	1667	1632

Criteria	Rater 1	Rater 2
Argumentation	15.2581	15.2258
Evidence	12.8710	12.8387
Organization	12.4839	11.8710
Language	13.0323	12.2581
Mean	53.6452	52.1935
Total	1663	1618

Post-feedback Essay Scores

Table 3. Post-feedback Scores

Post Feedback	RATER 1	RATER 2
Essay1	78	73
Essay2	73	72
Essay3	78	75
Essay4	82	84
Essay5	84	78
Essay6	77	73
Essay7	75	75
Essay8	78	86
Essay9	72	77
Essay10	74	79
Essay11	77	83
Essay12	74	77
Essay13	77	78
Essay14	82	86
Essay15	76	78
Essay16	74	71
Essay17	84	86
Essay18	70	66
Essay19	72	76
Essay20	66	71
Essay21	75	77
Essay22	78	76
Essay23	74	76
Essay24	59	57
Essay25	74	71

Essay26	78	87
Essay27	69	70
Essay28	68	69
Essay29	70	66
Essay30	65	63
Essay31	64	66
Mean	74.0968	75.3871
Total	2297	2337
Criteria	Rater 1	Rater 2
Argumentation	19.2581	19.3871
Evidence	18.7097	19.3548
Organization	18.4839	18.6452
Language	17.6452	17.5161
Mean	74.0968	74.9032
Total	2297	2322

Pre-feedback and Post-feedback Paired Sample Test Results

The results of the students' essays, both before and after receiving formative feedback from ChatGPT, were analyzed using the Paired Samples functions in SPSS. The detailed findings are presented in the table 4 and 5.

Rater 1

Table 4. Paired Samples Test Rater 1

Paired Samples Test									
		Paired Differences			95% Confidence Interval of the Difference			Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Pair 1	PreFeedbackRater1 - PostFeedbackRater1	-20.45161	10.83155	1.94540	-24.42466	-16.47857	-10.513	30	.000

Rater 2

Table 5. Paired Samples Test Rater 2

Paired Samples Test									
		Paired Differences			95% Confidence Interval of the Difference			Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Pair 1	PreFeedbackRater2 - PostFeedbackRater2	-22.70968	13.38206	2.40349	-27.61826	-17.80110	-9.449	30	.000

Overall Impact of ChatGPT's Feedback

Table 6. Aspects Ranking

Aspect	Pre-Feedback Mean	Post-Feedback Mean	Increase
Argumentation	15.2581	19.129	3.8709
Evidence	12.871	18.7097	5.8387
Organization	12.4839	18.4194	5.9355
Language	13.0323	17.6452	4.6129

Among the four assessed aspects, Organization demonstrated the most significant improvement, with a gain of 5.9355 points in the mean score after receiving formative feedback from ChatGPT. This suggests that students benefited greatly from the feedback in terms of structuring their ideas more coherently and logically. The feedback appeared to help them manage the flow and clarity of their arguments more effectively, which is a key component of strong argumentative writing. The next highest improvement was seen in the Evidence aspect, followed by Language, and finally Argumentation. This ranking indicates that while all aspects improved, structural coherence was where ChatGPT's formative feedback had the strongest impact.

Participants' Perception to ChatGPT's Formative Feedback

This section presents the perceptions of two participants regarding the use of ChatGPT as a formative feedback tool in writing argumentative texts. One participant represents a higher English proficiency level, and the other represents a lower level. Their responses help illustrate how students at different proficiency levels interact with and benefit from AI-generated feedback.

Higher proficiency participant demonstrated a strong appreciation for ChatGPT's assistance in organizing ideas and identifying weaknesses in the writing structure. However, they also acknowledged the limitations of the AI, particularly in misinterpreting prompts. They emphasized that human feedback is still irreplaceable due to the deeper contextual understanding that teachers provide.

The participant with lower proficiency also found the feedback helpful, especially in making sentences and paragraphs more organized. However, they sometimes felt confused by the large amount of information provided by ChatGPT. Overall, both participants saw ChatGPT as a helpful tool to support their learning, but they believed that guidance from teachers is still needed to explain feedback more clearly and reduce misunderstandings.

CONCLUSION

This study investigated the effectiveness of ChatGPT-generated formative feedback in enhancing argumentative writing skills among EFL English Literature students. The findings provide compelling evidence that AI-assisted feedback can significantly improve multiple dimensions of argumentative writing performance. Quantitative analysis revealed statistically significant improvements in writing quality, with mean scores increasing from 69.69 to 74.75 (Rater 1) and 65.00 to 70.81 (Rater 2). Paired samples t-test results ($p < .001$ for both raters) confirm genuine learning gains rather than random variation. Detailed component analysis showed evidence improvement as the highest

(5.84 points), followed by organization (5.94 points), language use (4.61 points), and argumentation (3.87 points), indicating ChatGPT's particular effectiveness in strengthening argument support and structural coherence.

Qualitative findings revealed generally positive student perceptions across proficiency levels, with participants reporting valuable assistance in mistake identification, sentence structure improvement, and idea development. However, students also identified challenges including the need for precise prompt formulation and occasionally overwhelming feedback volume. These findings highlight both the potential and limitations of AI-assisted writing instruction, emphasizing that optimal implementation requires combining ChatGPT feedback with human supervision and guided interaction strategies.

The study contributes to understanding AI's role in EFL writing pedagogy, demonstrating that ChatGPT can serve as an effective supplementary tool for argumentative writing instruction while maintaining the essential human elements of effective teaching. Future research should examine long-term effects of sustained AI feedback use across broader student populations and explore optimal prompt design strategies. This research supports thoughtful integration of AI technologies in language education, showing that when implemented appropriately, tools like ChatGPT can meaningfully enhance traditional pedagogical approaches without replacing human instruction.

REFERENCES

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00455-4>

Bom, H. S. H. (2023). Exploring the opportunities and challenges of ChatGPT in academic writing: a roundtable discussion. *Nuclear medicine and molecular imaging*, 57(4), 165-167.

Bozorgian, H., & Yazdani, A. (2021). Direct written corrective feedback with metalinguistic explanation: Investigating language analytic ability. *Iranian Journal of Language Teaching Research*, 9(1), 65-85.

Chapelle, C. A. (2009). *The Relationship Between Second Language Acquisition Theory and Computer-Assisted Language Learning*. *The Modern Language Journal*, 93(s1), 741–753

Connor, U. (1996). *Contrastive rhetoric*. Cambridge University Press.

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The modern language journal*, 92(2), 232-249.

Gupta, M. R. (2024). ChatGPT-A Generative Pre-Trained Transformer. DOI: 10.48175/IJARSCT-15087

Hidayat, M. Z. A. T., & Pariyanto, P. (2024, July). The Impact of ChatGPT Formative Feedback on EFL Learner. In Proceeding of Undergraduate Conference on Literature, Linguistic, and Cultural Studies (Vol. 3, No. 1, pp. 143-153).

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of second language writing*, 12(1), 17-29.

Khasinah, S. (2014). Factors influencing second language acquisition. *Englisia: Journal of Language, Education, and Humanities*, 1(2).

Krashen, S. (1982). Principles and practice in second language acquisition. New York: Prentice-Hall International.

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. (2021). How does online peer feedback improve argumentative essay writing and learning?. *Innovations in Education and Teaching International*, 58(2), 195-206.

Long, M. H. (1996). *The Role of the Linguistic Environment in Second Language Acquisition*. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). San Diego: Academic Press.

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, 11(1), 9.

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific Region. *TESOL quarterly*, 37(4), 589-613.

Pariyanto, P., & Pradipta, B. (2020). FACTORS INFLUENCING AN EFL LEARNER'S PROFICIENCY: AN ENGLISH TEACHER'S PERSPECTIVE. *Anaphora : Journal of Language, Literary, and Cultural Studies*, 2(2), 89- 97. <https://doi.org/10.30996/anaphora.v2i2.3369>

Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1), 10.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752.

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *Journal of effective teaching*, 7(1), 3-14.

Zebua, J. A. Z., & Katemba, C. V. (2024). Students' Perceptions of Using the OpenAI ChatGPT Application in Improving Writing Skills. *Journal of Language and Literature Studies*, 4(1), 110–123. <https://doi.org/10.36312/jolls.v4i1.1805>