Humanitatis: Journal of Language and Literature
SK Dirjen DIKTI Nomor 36/E/KPT/2019
Vol. 12, No. 2, June 2026, pp. 191-202
ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

Artificial Intelligence Writing Assistants as Reflective Partners: A Quasi-Experimental Study in a Non-English Academic Context

Hikma Tansilo[1], Yuyun Setiawan Putra[1], Andini Septama Sari[*1], Tommy Hastomo[2]

[1]STKIP Muhammadiyah Pagaralam, Pagaralam, Indonesia
[2]STKIP PGRI Bandar Lampung, Bandar Lampung, Indonesia

Received: 16th October 2025 | Revised: 29th May 2026 | Accepted: 30th June 2026
[*]Corresponding author. E-mail: andiniseptama@gmail.com.

Abstract

Argumentative writing is an essential academic skill in higher education; however, students in non-English academic contexts frequently encounter difficulties in developing coherent arguments, organizing ideas logically, and integrating supporting evidence effectively. This study aimed to investigate the effectiveness of Artificial Intelligence Writing Assistants (AIWAs), particularly ChatGPT, in improving Indonesian undergraduate students' argumentative writing performance. This study employed a mixed-methods approach with a quasi-experimental pretest–posttest control group design. Forty fourth-semester undergraduate students were divided equally into an experimental group receiving AI-assisted feedback through ChatGPT and a control group receiving conventional teacher and peer feedback. Quantitative data were collected through argumentative writing tests and analyzed using paired-samples t-tests, independent-samples t-tests, and gain score analysis. Qualitative data were obtained through classroom observations and semi-structured interviews to provide complementary insights into students' learning experiences. The results revealed that both groups improved after the intervention; however, the experimental group achieved significantly higher posttest scores ($M = 82.70$) than the control group ($M = 76.90$), with a statistically significant difference ($p = .005$). Qualitative findings further indicated that AIWAs facilitated iterative revision, strengthened coherence and argumentative organization, and enhanced students' confidence in writing. The findings suggest that AIWAs can serve as effective reflective partners in argumentative writing instruction and offer considerable pedagogical potential to support academic literacy development and digital transformation in higher education, particularly in non-English academic contexts.

Keywords: Argumentative Writing; Artificial Intelligence Writing Assistants; Digital Pedagogy; Higher Education; Indonesia.

How to Cite:
Tansilo, H., Putra, Y. S., Sari, A. S., & Hastomo, T. (2026). Artificial Intelligence Writing Assistants as Reflective Partners: A Quasi-Experimental Study in a Non-English Academic Context. *Humanitatis : Journal of Language and Literature*, *12*(2), 191-202.

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

1. INTRODUCTION

Argumentative writing has long been recognized as a cornerstone of higher education, serving as a key vehicle for cultivating critical thinking, analytical reasoning, and persuasive communication among students. This genre of writing requires learners to articulate claims supported by evidence, engage with counter-arguments, and present coherent, logically structured texts that reflect both intellectual rigor and communicative clarity (Van De Poel & Gasiorek, 2024; Yeo, 2023). Beyond its role in academic assessment, the mastery of argumentative writing equips students with competencies essential for professional life, where the ability to articulate ideas persuasively is highly valued across disciplines and industries (Balraj, 2025).

In non-English contexts, particularly in Indonesia, the development of argumentative writing skills remains a persistent challenge. Studies indicate that students often struggle with grammatical accuracy, lexical range, and syntactic complexity when writing in English as a second language (Sanosi, 2022; Shahsavar et al., 2024). These linguistic barriers are compounded by limited familiarity with the rhetorical conventions of academic argumentation, resulting in essays that lack coherence, logical flow, and depth of analysis (Hamamah et al., 2023). Moreover, the prevalence of rote-learning traditions in Indonesian education has further hindered the development of critical and analytical thinking skills necessary for constructing persuasive arguments (Kim et al., 2025; Tuong & Tran, 2025). Such pedagogical practices prioritize memorization over reasoning, leaving students ill-prepared to produce nuanced argumentative texts that demand originality and critical engagement (El-Garawany, 2024).

These challenges highlight a pressing need for innovative pedagogical interventions that not only address linguistic shortcomings but also cultivate higher-order cognitive skills. In this regard, technology-enhanced writing support systems have emerged as a promising avenue for improving writing pedagogy. In particular, AIWAs have gained traction as tools that provide immediate, personalized feedback on grammar, coherence, vocabulary use, and rhetorical effectiveness. Applications such as Grammarly, QuillBot, and ChatGPT have been widely adopted in English as a Foreign Language (EFL) and English as a Second Language (ESL) contexts, with empirical evidence pointing to their positive influence on student writing performance, motivation, and self-efficacy (Imran & Almusharraf, 2023; Sanosi, 2022; Tuong & Tran, 2025).

The integration of AIWAs into academic writing instruction has shown several tangible benefits. For instance, students using Grammarly have demonstrated significant improvements in grammatical accuracy and lexical precision (Pryma et al., 2025; Sanosi, 2022). Similarly, QuillBot has been reported to enhance student confidence and engagement during collaborative writing projects (El-Garawany, 2024). Moreover, AI-driven platforms such as ChatGPT offer not only corrective feedback but also interactive support that enables students to experiment with language, refine argumentative structures, and engage in reflective learning practices (Kim et al., 2025). These findings suggest that AIWAs can function as reflective partners in learning, complementing instructor feedback while promoting autonomy and active learning among students (Bašić et al., 2023; Schei et al., 2024).

However, despite these promising developments, current scholarship on AIWAs remains disproportionately focused on English-dominant contexts. The dominance of English as the lingua franca in academia has resulted in a concentration of research that prioritizes the experiences of students and instructors in Anglophone institutions or EFL/ESL settings with a strong orientation toward English-language proficiency (Heintz et al., 2022; Shahsavar et al., 2024). Consequently, little is known about how AIWAs function in non-English settings, such as Indonesia, where linguistic, cultural, and pedagogical dynamics may influence both the adoption and effectiveness of these tools. The limited availability of AIWA features tailored to Bahasa Indonesia further exacerbates this research gap, raising questions about the adaptability of existing platforms and the unique challenges faced by Indonesian learners (Hamamah et al., 2023).

This gap underscores the importance of investigating the role of AIWAs in contexts beyond English. In Indonesia, where students often face difficulties in constructing logically consistent and linguistically accurate argumentative texts, AIWAs hold the potential to mitigate these challenges by offering real-time feedback and scaffolding support. Yet, the extent to which these tools can effectively enhance argumentative writing in Indonesian

academic settings remains underexplored. Addressing this gap requires empirical evidence that evaluates both the benefits and limitations of AIWAs when applied to a non-English linguistic and educational environment (Mohammed et al., 2025; Nazari et al., 2021).

The present study seeks to respond to this lacuna by systematically examining the effectiveness of AIWAs in improving the argumentative writing skills of Indonesian university students. Specifically, it aims to assess whether students who receive AIWA-assisted instruction demonstrate greater improvements in the quality of their argumentative essays compared to peers taught through conventional methods. The study also explores the perceived advantages and limitations of AIWAs from the student perspective, providing insights into how these tools shape learners' writing processes, motivation, and confidence. By comparing AIWA-based interventions with traditional pedagogical approaches, this research endeavors to generate evidence that can inform both theory and practice in writing instruction.

The significance of this study is threefold. Theoretically, it contributes to the growing body of literature on technology-enhanced writing pedagogy by extending discussions to a non-English context, thereby enriching understandings of human–AI collaboration in academic literacy (Tuong & Tran, 2025; Yeo, 2023). Practically, the findings are expected to provide instructors with a model for integrating AIWAs into the teaching of argumentative writing, offering pedagogical strategies that balance technological affordances with the cultivation of critical thinking and independent writing skills (Chemaya & Martin, 2024; Van De Poel & Gasiorek, 2024). At the policy level, this study aligns with broader efforts to promote digital transformation in higher education, highlighting the role of AI tools in enhancing student learning outcomes and preparing graduates for the demands of the digital era (Balraj, 2025).

Several previous studies have investigated the role of artificial intelligence in writing instruction. Sanosi (2022) examined the impact of automated written corrective feedback on EFL students' academic writing accuracy using an experimental approach and found significant improvements in grammatical accuracy and revision quality. Imran and Almusharraf (2023), through a systematic literature review, reported that ChatGPT functions effectively as a writing assistant by enhancing learner motivation, engagement, and writing support in higher education contexts. El-Garawany (2024) employed a QuillBot-based intervention with English language majors and demonstrated positive effects on writing performance, self-efficacy, and reduced writing apprehension. Milton et al. (2024) investigated AI-powered writing tools among health science graduates using quantitative analysis and found measurable improvements in independent writing skills and feedback responsiveness. Furthermore, Kim et al. (2025) explored university students' perspectives on generative AI-assisted academic writing through qualitative inquiry and revealed that students perceived AI tools as valuable supports for idea development, organization, and revision processes.

Although these studies provide important evidence regarding the pedagogical benefits of AI-assisted writing, several limitations remain. First, much of the existing literature has focused on English-dominant, EFL, or highly resourced educational contexts. Second, prior studies predominantly emphasize grammatical accuracy, lexical enhancement, learner perceptions, or general writing assistance rather than argumentative writing as a genre requiring higher-order thinking, coherence, and counter-argumentation. Third, limited empirical attention has been given to how AI writing assistants function within non-English academic environments such as Indonesia, where linguistic, pedagogical, and technological conditions differ substantially.

The novelty of this study is the investigation of AI Writing Assistants (AIWAs), specifically ChatGPT, as reflective partners in improving argumentative writing performance among Indonesian university students through a quasi-experimental design integrating quantitative and qualitative evidence. Unlike previous studies that mainly focus on surface-level writing improvement or English-dominant contexts, this study emphasizes higher-order argumentative writing development within a non-English academic setting. In sum, this research not only addresses an urgent pedagogical challenge in Indonesian higher education but also contributes to the broader discourse on equitable access to AI-supported learning. By foregrounding the experiences of non-English-speaking students, it aims to generate insights that support more inclusive, context-sensitive approaches to integrating AI into academic

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

writing pedagogy. Ultimately, the study underscores the potential of AIWAs to transform the teaching and learning of argumentative writing, while also cautioning against overreliance on technology by emphasizing the need for pedagogical frameworks that nurture creativity, critical reasoning, and authentic expression.

2. RESEARCH METHOD

This study employed a mixed methods approach, integrating quantitative and qualitative data to obtain a comprehensive understanding of the effectiveness of Artificial Intelligence Writing Assistants (AIWAs) in improving students' argumentative writing skills. The study primarily adopted a quasi-experimental quantitative design using a pretest–posttest control group format to examine differences in writing performance between experimental and control groups. This method was selected because it enables researchers to investigate the impact of instructional interventions within naturally occurring educational environments where complete randomization is difficult or institutionally impractical (Ng et al., 2025; Schei et al., 2024). In addition, qualitative data from classroom observations and semi-structured interviews were incorporated to enrich interpretation of the quantitative findings and provide deeper insights into students' experiences, perceptions, and learning behaviors during the intervention process. The use of mixed methods is appropriate because it enables researchers to combine statistical evidence with contextual understanding, thereby enhancing the completeness and credibility of educational research findings (Creswell & Creswell, 2017). In educational research, particularly in classroom-based studies involving writing instruction, quasi-experimental designs are widely recognized as appropriate approaches for evaluating pedagogical innovations while maintaining ecological validity. The present study aimed to determine whether the integration of AI-supported writing assistance, specifically ChatGPT, could significantly influence students' argumentative writing performance compared with conventional instructional practices.

The population of this study consisted of fourth-semester undergraduate students enrolled in the Indonesian Language Education program at a public university. This population was selected because students at this academic level had already completed foundational coursework related to academic literacy, composition, and introductory writing practices, making them suitable participants for argumentative writing instruction. From this population, 40 students who met the predetermined inclusion criteria were selected as the research sample. The sampling process employed purposive sampling to ensure that participants possessed relatively comparable educational backgrounds, writing exposure, and academic readiness. The participants were divided equally into two groups: 20 students in the experimental group and 20 students in the control group.

Table 1. Overview of research design, instruments, and analysis methods

Component	Description
Research Design	Quasi-experimental pretest–posttest control group
Participants	40 students (20 experimental, 20 control)
Instruments	Argumentative writing rubric, classroom observation, semi-structured interview
Procedures	Pretest → Intervention (AIWA vs. conventional) → Posttest
Data Analysis	Normality test, homogeneity test, paired t-test, independent t-test

As summarized in Table 1, this study employed a quasi-experimental pretest–posttest control group design to investigate the effectiveness of Artificial Intelligence Writing Assistants (AIWAs) in improving students' argumentative writing performance. The study involved 40 undergraduate students who were equally assigned to experimental and control groups. Multiple instruments were utilized, including an argumentative writing rubric, classroom observations, and semi-structured interviews, to obtain both quantitative and qualitative data. Furthermore, the research procedure consisted of three stages: pretest, intervention, and posttest. To ensure the robustness of the findings, quantitative data were analyzed using normality and homogeneity tests, paired-samples t-tests, independent-samples t-tests, and gain score analysis. The integration of multiple instruments and analytical procedures enabled a comprehensive evaluation of the effectiveness of AI-assisted writing instruction.

DOI: 10.30812/humanitatis.v12i2.5836

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

Several inclusion criteria guided participant selection. First, participants had to be actively enrolled in the Indonesian Language Education program during the semester in which the study was conducted. Second, participants were required to have completed introductory academic writing courses to ensure a minimum level of familiarity with argumentative writing conventions, paragraph organization, and academic discourse practices. Third, participants needed to demonstrate regular classroom attendance and willingness to participate throughout the intervention period. These criteria were established to minimize variability associated with unequal writing preparation and inconsistent participation (Sanosi, 2022). The experimental group received AI-assisted feedback through ChatGPT during argumentative writing instruction, whereas the control group received conventional instruction involving teacher explanations, manual revision guidance, and peer feedback activities.

Data collection involved multiple techniques to obtain comprehensive evidence regarding students' writing development and learning experiences. The study utilized argumentative writing tests, classroom observations, and semi-structured interviews as primary data collection methods. The use of multiple instruments was intended to strengthen methodological rigor through data triangulation, allowing quantitative findings to be supported and enriched by qualitative insights. The argumentative writing test functioned as the principal research instrument and was administered during both pretest and posttest sessions. The pretest was conducted before the intervention to establish participants' baseline writing proficiency, while the posttest was administered after the intervention to measure changes in writing performance attributable to the instructional treatment. During both testing sessions, students were required to compose argumentative essays based on predetermined prompts addressing contemporary educational or social issues relevant to university students. The prompts were designed to stimulate critical thinking, evidence-based reasoning, and structured argument development.

Student essays were evaluated using a validated argumentative writing rubric adapted from established academic writing assessment frameworks. The rubric assessed multiple dimensions of argumentative writing quality, including thesis clarity, logical organization, coherence and cohesion, evidence integration, grammatical accuracy, lexical appropriateness, rhetorical effectiveness, and counter-argument incorporation. Each dimension was scored using standardized performance descriptors to ensure consistency and objectivity in evaluation. To enhance scoring reliability, essays were reviewed systematically according to predetermined assessment criteria. The rubric-based evaluation process enabled the researchers to capture both surface-level linguistic features and higher-order argumentative competencies. In addition to writing tests, classroom observations were conducted throughout the intervention period to document students' learning behaviors, interaction patterns, classroom participation, and engagement with instructional activities. Observation activities focused on how students interacted with writing tasks, responded to feedback, revised drafts, and utilized available learning resources. Particular attention was given to differences between the experimental and control groups regarding revision strategies, collaborative behaviors, and responsiveness to instructional support. Observation notes were recorded systematically during classroom sessions to provide contextual information regarding the implementation of the intervention.

Semi-structured interviews were administered after the completion of the intervention to obtain qualitative insights into students' experiences, perceptions, attitudes, and challenges related to AI-supported writing instruction (Milton et al., 2024). Interview participants were selected from both experimental and control groups to ensure balanced representation of perspectives. The interview protocol included questions concerning students' perceptions of argumentative writing difficulty, experiences with feedback processes, perceived usefulness of instructional support, confidence in writing development, and reflections on classroom learning experiences. For students in the experimental group, additional questions explored their experiences using ChatGPT, including perceived advantages, limitations, ethical concerns, and the extent to which AI assistance influenced their writing practices.

The instructional intervention was conducted over four weeks and consisted of structured argumentative writing activities integrated into regular classroom instruction. During this period, students in the experimental group used ChatGPT as an AI writing assistant for grammar correction, coherence enhancement, rhetorical refinement, idea generation, organizational support, and iterative revision activities. Students were encouraged to interact critically with AI-generated suggestions rather than accepting outputs uncritically. The instructional

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

approach emphasized reflective engagement with AI feedback, requiring students to evaluate, modify, and justify revisions made to their essays.

The experimental group participated in guided writing sessions involving multiple drafting cycles. Students initially produced preliminary essay drafts independently before consulting ChatGPT for feedback and revision suggestions. They then revised their essays based on a combination of personal judgment, instructor guidance, and AI-generated recommendations. This iterative process was intended to promote metacognitive awareness, critical reflection, and deeper engagement with argumentative writing conventions. Meanwhile, students in the control group completed comparable writing tasks using conventional instructional support without AI assistance. Instruction in the control group relied on teacher explanations, classroom discussions, peer review activities, and manual revision practices. Students received feedback from instructors and classmates regarding grammar, organization, clarity, and argumentative structure. By maintaining comparable writing objectives and instructional duration across both groups, the study sought to isolate the specific contribution of AI-supported writing assistance to students' argumentative writing development.

At the conclusion of the intervention period, both groups completed a posttest administered under conditions equivalent to the pretest. Equivalent testing conditions were maintained to reduce procedural bias and ensure comparability of results. Students completed the writing assessment individually within a predetermined timeframe and without unauthorized external assistance.

Data analysis involved both quantitative and qualitative procedures to provide a comprehensive understanding of the intervention's effectiveness. Quantitative data analysis began with descriptive statistical procedures to summarize participants' writing performance across pretest and posttest conditions. Measures such as mean scores, standard deviations, and score distributions were calculated to provide an overview of students' writing achievement patterns.

Before conducting inferential statistical analyses, normality and homogeneity tests were performed to determine whether the data satisfied assumptions required for parametric statistical procedures. The normality test examined whether score distributions approximated normal distribution patterns, while the homogeneity test assessed variance equivalence between groups. Following confirmation of statistical assumptions, paired-samples t-tests were used to examine within-group differences between pretest and posttest scores. These analyses enabled the researchers to determine whether significant improvements occurred within each instructional condition. Independent-samples t-tests were subsequently conducted to compare writing performance between the experimental and control groups. This analysis aimed to determine whether students receiving AI-assisted instruction demonstrated significantly different outcomes compared with students receiving conventional instruction. Gain score analysis was additionally employed to measure the magnitude of improvement attributable to the intervention by calculating differences between pretest and posttest performance levels. The use of multiple quantitative procedures strengthened the robustness of statistical interpretation and enhanced confidence in the study's findings.

Qualitative data obtained from classroom observations and semi-structured interviews were analyzed descriptively using thematic categorization procedures. Observation notes and interview responses were reviewed, coded, and organized into recurring themes related to student engagement, writing confidence, revision behavior, perceptions of feedback, and experiences with AI-supported learning. The qualitative analysis served to contextualize and explain quantitative outcomes by providing deeper insight into how students experienced the instructional intervention and how AI assistance influenced their writing processes.

Ethical considerations were maintained throughout the study to ensure responsible research conduct and participant protection. Prior to data collection, participants were informed about the objectives, procedures, and voluntary nature of the study. Informed consent was obtained from all participants before their involvement in research activities. Participant anonymity was preserved through the use of coded identifiers, and all collected data were handled confidentially for research purposes only. Furthermore, the study emphasized the ethical use of AI tools as reflective learning supports rather than instruments for plagiarism or academic misconduct (Yeo, 2023). Students in the experimental group received explicit guidance regarding responsible AI usage, including the

DOI: 10.30812/humanitatis.v12i2.5836

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

importance of maintaining authorship, exercising critical judgment toward AI-generated suggestions, and adhering to principles of academic integrity.

3. FINDINGS AND DISCUSSION

The quasi-experimental investigation yielded convergent quantitative and qualitative evidence that the AI Writing Assistant (AIWA; ChatGPT) improved undergraduate students' argumentative writing more than conventional instruction. Prior to the intervention, the experimental and control groups exhibited comparable proficiency, thereby supporting the internal validity of subsequent contrasts. After four weeks of instruction, the experimental group outperformed the control group on the posttest, with statistically significant within-group and between-group differences and substantively meaningful gains (Milton et al., 2024; Sanosi, 2022).

3.1. Pretest Equivalence

Baseline comparability was established with the same validated Argumentative Writing Rubric applied across groups. Descriptive statistics indicated a mean of 56.81 ($SD = 9.69$) in the control group and 57.50 ($SD = 8.94$) in the experimental group. Normality (Shapiro–Wilk) held for both pretest distributions (control: $W = .959$, $p = .526$; experimental: $W = .946$, $p = .305$). As a result, parametric tests were appropriate, and an independent-samples t-test confirmed no significant pretest difference, indicating initial equivalence of writing ability and attenuating the threat of selection bias. This set-up is consistent with quasi-experimental best practice for isolating treatment effects in authentic classrooms (Ng et al., 2025; Schei et al., 2024).

3.2. Posttest Improvements

Following the intervention (four instructional weeks, with reflective, iterative drafting supported by AI feedback in the experimental group), both groups improved, but the experimental group's posttest mean was higher ($M = 82.70,\ SD = 5.03$) than the control group's ($M = 76.90,\ SD = 6.91$). Normality assumptions held (experimental: $W = .918,\ p = .091$; control: $W = .945, p = .264$), and homogeneity of variances was supported (Levene's $F = 0.189,\ p = .666$). A paired-samples t-test within the experimental group showed a significant pre–post gain ($M\triangle = 25.20,\ t(19) = 12.961,\ p < .001$), while an independent-samples t-test on posttest scores revealed a significant between-group difference favoring the AIWA condition ($t(38) = 2.972,\ p = .005$). Complementarily, gain-score analyses indicated high effectiveness for the experimental group relative to control, corroborating the hypothesis that AIWA support confers measurable advantages for argumentative writing (Milton et al., 2024).

Table 2. Descriptive statistics and assumption checks

Measure	Control Pretest	Experimental Pretest	Control Posttest	Experimental Posttest
N	20	20	20	20
Mean	56.81	57.50	76.90	82.70
SD	9.69	8.94	6.91	5.03
Shapiro–Wilk p	.526	.305	.264	.091
Levene's p (posttest)	-	-	-	.666

The descriptive and assumption-testing results presented in Table 2 provide preliminary evidence supporting the effectiveness of the intervention. As shown in Table 2, both groups demonstrated comparable pretest performance, with the control group obtaining a mean score of 56.81 ($SD = 9.69$) and the experimental group achieving a mean score of 57.50 ($SD = 8.94$), indicating relatively similar initial argumentative writing abilities. Following the intervention, the experimental group showed a higher posttest mean score ($M = 82.70$, $SD = 5.03$) than the control group ($M = 76.90$, $SD = 6.91$), suggesting stronger improvement among students receiving AIWA-supported instruction. Furthermore, the normality and homogeneity test results reported in Table 2 confirmed that the data met the assumptions required for parametric analysis, thereby supporting the validity of subsequent inferential statistical testing.

DOI: 10.30812/humanitatis.v12i2.5836

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

Table 3. Descriptive statistics and assumption checks

Test	Contrast	Statistic	df	p-value	Interpretation
Paired t-test	Experimental: Post - Pre	12.961	19	<.001	Significant within-group improvement
Independent t-test	Posttest: Experimental vs. Control	2.972	38	.005	Significant between-group difference
Gain score	Experimental vs. Control	-	-	<.05	Higher effectiveness for AIWA group

The inferential statistics presented in Table 3 provide additional support for this finding. The paired-samples t-test revealed a statistically significant improvement within the experimental group between pretest and posttest scores ($t(19) = 12.961$, $p < .001$), indicating substantial learning gains after exposure to AIWA-supported instruction. Moreover, the independent-samples t-test reported in Table 3 demonstrated a significant difference between the experimental and control groups at posttest ($t(38) = 2.972$, $p = .005$), suggesting that the AIWA intervention produced a stronger instructional effect than conventional teaching methods. Gain score analysis in Table 3 further corroborated the superior effectiveness of the experimental treatment condition.

3.3. Qualitative Corroboration

Classroom observations and interviews indicated that students in the AIWA condition engaged more frequently in revision cycles, invoked feedback to restructure claims and evidence, and reported higher confidence in articulating counter-arguments. Students characterized the AI as a reflective partner that facilitated noticing of coherence lapses and syntactic issues, echoing prior reports that AI-mediated feedback can enhance motivation and self-efficacy (Milton et al., 2024; Sanosi, 2022). These perceptions align with the study's theoretical positioning of AIWA as a scaffold for metacognitive regulation in writing rather than a mere correctness checker (Schei et al., 2024).

3.4. Interpretation of Findings

The combined evidence supports the conclusion that AIWAs significantly improve argumentative writing. The largest effect was observed in coherence, logical structuring, and argumentative force—dimensions explicitly targeted by iterative feedback on organization and support. The experimental group's substantial mean increase (≈25 points) and reduced dispersion (lower posttest SD) suggest both improvement and stabilization of quality, consistent with targeted scaffolding effects. Importantly, the initial equivalence mitigates rival explanations related to prior ability. These outcomes are theoretically consistent with designs that leverage formative, timely feedback to strengthen genre-specific moves such as claim formulation, evidence selection, and rebuttal integration under constrained classroom conditions (Kim et al., 2025; Schei et al., 2024). In addition, recent studies have shown that AI-supported instructional environments can significantly strengthen student engagement and active participation during writing activities, particularly when AI tools are integrated through reflective and interactive pedagogical strategies (Hastomo, Sari, Widiati, Ivone, Zen, & Kholid, 2025). The iterative interaction between learners and AI feedback observed in this study also reflects broader findings that AI-assisted learning environments can facilitate more autonomous and self-regulated learning behaviors among students (Sari et al., 2026).

3.5. Comparison with Previous Studies

The present pattern aligns with EFL/ESL research on AI-enabled writing support tools (e.g., Grammarly, QuillBot, ChatGPT), which has documented gains in grammatical accuracy, lexical selection, and perceived competence (Milton et al., 2024; Sanosi, 2022). Similarly, previous studies have reported that AI chatbot integration can positively influence learners' engagement, cognitive participation, and language development through personalized and adaptive feedback mechanisms (Hastomo, Sari, Widiati, Ivone, Zen, & Andianto, 2025; Hastomo, Sari, Widiati, Ivone, Zen, & Kholid, 2025). At the same time, the current study extends the evidence base to a non-English context by demonstrating improvements in discourse-level features such as coherence and argumentative structure when AIWA is framed as a reflective partner in revision. This nuance is salient in light of cautions that AI tools may under-serve higher-order thinking without pedagogical mediation (Tuong & Tran, 2025). The present findings are also consistent with studies emphasizing that AI-based instructional support becomes more effective when

DOI: 10.30812/humanitatis.v12i2.5836

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

learners actively engage in reflective revision and self-monitoring processes (Sari et al., 2026). Furthermore, the non-English setting and rubric-based validation contribute new empirical grounding for contextualized AI integration, particularly in Indonesian higher education contexts where AI-based language learning research remains relatively limited (Putra et al., 2025).

3.6. Pedagogical Implications

First, positioning AIWA as a reflective partner—not solely a surface-level corrector—appears to optimize its utility. Instructors can engineer cycles of plan–draft–AI feedback–revise to encourage structural reorganization and argument strengthening, consistent with evidence that structured, technology-supported feedback promotes sustained engagement and self-efficacy (Milton et al., 2024; Sanosi, 2022). Recent evidence also suggests that interactive AI-supported learning environments can increase students' affective, behavioral, and cognitive engagement during writing instruction when teachers strategically integrate personalized feedback and collaborative activities (Hastomo, Sari, Widiati, Ivone, Zen, & Andianto, 2025). Second, instructor mediation remains pivotal: task prompts, exemplars, and rubric-aligned feedback channel AIWA suggestions toward genre expectations, supporting personalized scaffolding (Kim et al., 2025; Schei et al., 2024). This aligns with findings emphasizing that learner autonomy and self-regulated learning develop more effectively when reflective activities and teacher-guided monitoring are explicitly embedded within instructional practices (Sari et al., 2026). Third, adoption in non-English contexts benefits from explicit guidance on translating AI feedback into local rhetorical norms, reinforcing transfer without eroding authentic voice. Similar concerns regarding contextual adaptation and localization of AI-assisted language learning have also been highlighted in studies investigating AI implementation beyond English-dominant educational settings (Putra et al., 2025)).

3.7. Theoretical Contribution

The findings support a synthesis of genre-based pedagogy with AI affordances: AIWA can surface genre moves such as thesis clarity, evidence warranting, and counter-argument integration while learners iteratively align drafts to rubric criteria. This operationalizes human–AI collaboration in academic literacy beyond English-dominant settings, expanding theoretical accounts of how feedback timing, specificity, and dialogic interaction mediate uptake in revision (Milton et al., 2024; Ng et al., 2025). Furthermore, the findings reinforce theoretical perspectives emphasizing that technology-enhanced learning environments can promote learner engagement, self-regulation, and reflective learning behaviors when appropriately scaffolded by instructors (Hastomo, Sari, Widiati, Ivone, Zen, & Andianto, 2025; Sari et al., 2026). The study thus extends discussions of equitable AI-enabled writing development to contexts where tool capabilities and linguistic resources differ from Anglophone defaults (Milton et al., 2024), particularly in relation to Indonesian learners navigating AI-assisted academic writing environments.

3.8. Challenges and Limitations

First, the limited feature coverage for Bahasa Indonesia in prevailing AI tools may constrain feedback quality on rhetorical nuance and idiomatic usage, a limitation echoed in prior critiques of tool localization (Tuong & Tran, 2025). Similar challenges concerning contextual interpretation, linguistic naturalness, and technological adaptation in non-English learning environments have also been identified in AI-assisted language learning studies conducted in Indonesian contexts (Putra et al., 2025). Second, some students initially resisted AI-supported revision due to unfamiliarity and concerns about over-reliance, requiring orientation and ethical framing to maintain academic integrity and authorship (Yeo, 2023). Previous studies have likewise noted that while students generally respond positively to AI integration, concerns related to ethical usage, dependency, and the balance between human instruction and AI assistance remain important pedagogical considerations (Hastomo, Sari, Widiati, Ivone, Zen, & Kholid, 2025). Third, the study's sample size (N = 40) and four-week duration limit generalizability and detection of longer-term transfer; future research should examine sustained effects, delayed posttests, and broader cohorts. Finally, given the quasi-experimental design, residual confounds cannot be entirely excluded despite pretest equivalence and assumption checks (Kim et al., 2025; Schei et al., 2024).

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

3.9. Overall Synthesis

Overall, the findings of this study both align with and extend prior research on the use of artificial intelligence in writing instruction. Consistent with previous studies (Milton et al., 2024; Sanosi, 2022), the results confirm that AI-assisted feedback contributes to measurable improvements in students' writing performance, particularly in grammatical accuracy and overall writing quality. Similarly, the observed increase in students' confidence and engagement supports earlier findings that AI tools can enhance motivation, engagement, and self-efficacy in academic writing contexts (Hastomo, Sari, Widiati, Ivone, Zen, & Kholid, 2025; Imran & Almusharraf, 2023; Schei et al., 2024).

However, this study also reveals notable distinctions. While much of the existing literature has emphasized surface-level improvements such as grammar correction and lexical enhancement, this study demonstrates that AIWAs can significantly support higher-order writing skills, including coherence, logical organization, and argumentative strength. This suggests that when AI tools are used iteratively and reflectively, they function not merely as corrective mechanisms but as cognitive scaffolds that facilitate deeper engagement with the writing process. Such findings are also compatible with perspectives emphasizing the importance of reflective and self-regulated learning processes in technology-enhanced language instruction (Sari et al., 2026).

Furthermore, unlike many prior studies conducted in English-dominant or highly resourced EFL contexts, this research situates AIWA use within a non-English academic environment. The findings indicate that despite limitations in linguistic localization, AI tools remain effective in supporting writing development, thereby extending the applicability of AI-assisted pedagogy to underrepresented contexts. This contrasts with concerns raised in previous research regarding the limited adaptability of AI systems in non-English settings (Tuong & Tran, 2025). Similar evidence has also emerged from recent Indonesian studies showing that AI-supported language learning technologies can effectively improve learner outcomes when integrated through context-sensitive pedagogical approaches (Putra et al., 2025).

Taken together, these findings suggest that the effectiveness of AI writing assistants is not solely determined by their technical capabilities, but also by how they are pedagogically integrated. When positioned as reflective partners within structured instructional frameworks, AIWAs can meaningfully enhance both the quality of student writing and the development of higher-order thinking skills. Overall, the findings of this study largely support previous research on the pedagogical value of artificial intelligence in writing instruction while also extending existing scholarship into non-English higher education contexts (Hastomo, Sari, Widiati, Ivone, Zen, & Kholid, 2025; Putra et al., 2025).

4. CONCLUSION

This study demonstrated that Artificial Intelligence Writing Assistants (AIWAs), particularly ChatGPT, can significantly improve undergraduate students' argumentative writing performance. Using a mixed-method quasi-experimental design, the findings revealed that students who received AI-assisted feedback achieved significantly higher posttest scores than those who received conventional instruction. The greatest improvements were observed in coherence, logical organization, and argumentative strength. Qualitative findings further indicated that AIWAs functioned as reflective partners by supporting iterative revision, enhancing confidence, and facilitating students' engagement with the writing process. These findings suggest that AI-assisted writing instruction can serve as an effective pedagogical approach for developing academic writing skills in non-English higher education contexts.

Future research is recommended to examine the long-term effects of AI-assisted writing instruction across larger and more diverse populations. Further studies may also investigate the effectiveness of different AI tools, explore their application in other writing genres and language skills, and evaluate how pedagogical factors such as teacher mediation, learner autonomy, and contextual adaptation influence the effectiveness of AI-supported learning environments. Such investigations would contribute to a deeper understanding of the sustainable and responsible integration of artificial intelligence in language education.

DOI: 10.30812/humanitatis.v12i2.5836

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

REFERENCES

Balraj, B. M. (2025). Exploring the Use of ChatGPT in Academic Writing: A Systematic Literature Review on Undergraduates' Perceptions. *Arab World English Journal*, (1), 348–357. https://doi.org/10.24093/awej/AI.20

Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities and Social Sciences Communications*, *10*(1), 750. https://doi.org/10.1057/s41599-023-02269-7

Chemaya, N., & Martin, D. (2024). Perceptions and detection of AI use in manuscript preparation for academic journals (J. Blake, Ed.). *PLOS ONE*, *19*(7), e0304807. https://doi.org/10.1371/journal.pone.0304807

Creswell, J. W., & Creswell, J. D. (2017, December 12). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

El-Garawany, M. S. M. (2024). The Effects of a QuillBot-Based Intervention on English Language Majors' EFL Writing Performance, Apprehension, and Self-Efficacy. *Language Teaching Research Quarterly*, *43*, 167–189. https://doi.org/10.32038/ltrq.2024.43.10

Hamamah, Emaliana, I., Hapsari, Y., Degeng, P. D. D., & Fadillah, A. C. (2023). Using Nominal Group Technique to Explore Publication Challenges and the Usefulness of AI-Based Writing Technologies: Insights From Indonesian Scholars. *Theory and Practice in Language Studies*, *13*(8), 2038–2047. https://doi.org/10.17507/tpls.1308.20

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. https://doi.org/10.5430/wjel.v15n7p93

Hastomo, T., Sari, A. S., Widiati, U., Ivone, F. M., Zen, E. L., & Kholid, M. F. N. (2025). Does Student Engagement with Chatbots Enhance English Proficiency? *ELOPE: English Language Overseas Perspectives and Enquiries*, *22*(1), 93–109. https://doi.org/10.4312/elope.22.1.93-109

Heintz, K., Roh, Y., & Lee, J. (2022). Comparing the accuracy and effectiveness of Wordvice AI Proofreader to two automated editing tools and human editors. *Science Editing*, *9*(1), 37–45. https://doi.org/10.6087/kcse.261

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, *30*(1), 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Milton, C., Vidhya, L., & Thiruvengadam, G. (2024). Examining the Impact of AI-Powered Writing Tools on Independent Writing Skills of Health Science Graduates. *Advanced Education*, *12*(25), 143–161. https://doi.org/10.20535/2410-8286.315068

Mohammed, S. I., Jabbar Kadhim, D., Zuheir Al-Metwali, B., Mudher Mikhael, E., & Mohammed Hassan, N. (2025). Academic Staff Perspectives on the Impact of Artificial Intelligence on Pharmaceutical Sciences Research and Writing: A Qualitative Study. *Iraqi Journal of Pharmaceutical Sciences*, *33*, 12–19. https://doi.org/10.31351/vol33iss(4SI)pp12-19

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Ng, H. Y., Hsu, T.-Y., Min, J., Kim, S., Rossi, R. A., Yu, T., Jung, H., & Huang, T.-H. K. (2025). Understanding How Paper Writers Use AI-Generated Captions in Figure Caption Writing. In Q. Wang, W. Yin, A. Aich,

ISSN (Print) : 2338-9362
ISSN (Online) : 2477-2267

Y. Suh, & K.-C. Peng (Eds.), *AI for Research and Scalable, Efficient Systems* (pp. 173–192, Vol. 2533). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-8912-5_8

Pryma, V., Pelivan, O., Teletska, T., Tsobenko, O., & Zagrebelna, N. (2025). AI Writing Assistants and Student Competence: A Linguistic Aspect. *Arab World English Journal*, (1), 319–329. https://doi.org/10.24093/awej/AI.18

Putra, Y. S., Tansilo, H., Hastomo, T., Sari, A. S., & Aguilar, M. G. W. (2025). Efficacy of AI-based Text-to-Speech in Indonesian pronunciation training for foreign speakers (BIPA): A mixed-method analysis. *Journal of Educational Management and Instruction (JEMIN)*, *5*(2), 479–492. https://doi.org/10.22515/jemin.v5i2.12403

Sanosi, A. B. (2022). The Impact of Automated Written Corrective Feedback on EFL Learners' Academic Writing Accuracy. *Journal of Teaching English for Specific and Academic Purposes*, *10*(2), 301–317. https://doi.org/10.22190/JTESAP2202301S

Sari, A. S., Widiati, U., Zein, E. L., & Suharyadi, S. (2026). From classroom to autonomy: A systematic literature review of self-regulated learning strategies in English language teaching. *Journal of Research on English and Language Learning (J-REaLL)*, *7*(1), 319–343. https://doi.org/10.33474/j-reall.v7i1.24739

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies. *Education Sciences*, *14*(8), 922. https://doi.org/10.3390/educsci14080922

Shahsavar, Z., Kafipour, R., Khojasteh, L., & Pakdel, F. (2024). Is artificial intelligence for everyone? Analyzing the role of ChatGPT as a writing assistant for medical students. *Frontiers in Education*, *9*, 1457744. https://doi.org/10.3389/feduc.2024.1457744

Tuong, T. P. L., & Tran, T.-T. (2025). The Differential Impact of AI Tools Among EFL University Learners: A Process Writing Approach. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 452–471. https://doi.org/10.26803/ijlter.24.5.24

Van De Poel, K., & Gasiorek, J. (2024). Using AI to expand the "Toolbox" for EAP writing instruction: Student experiences and perceptions of ChatGPT's instructional potential. *AILA Review*. https://doi.org/10.1075/aila.24029.van

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, *14*(3), e716. https://doi.org/10.1002/tesj.716