

Maidodi Leksmana

129

 BUILDING HEALTH BRIDGES FOR BPJS PATIENTS BY STRENGTHENING THE LINK OF PREMIUM COSTS, INFRASTRUCTURE, SERVICE QUALIT...

 jabfung

 Universitas Prof. Dr. Moestopo (Beragama)

Document Details

Submission ID**trn:oid::1:3654121792****Submission Date****Sep 20, 2026, 9:18 AM GMT+7****Download Date****Sep 20, 2026, 9:20 AM GMT+7****File Name****129_TUR.docx****File Size****37.9 KB****17 Pages****3,960 Words****27,057 Characters**

4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

-  **15 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 3%  Internet sources
- 1%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 15 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 1% Publications
- 0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	muce.ac.tz	<1%
2	Internet	policyrj.com	<1%
3	Internet	sportpedagogy.org.ua	<1%
4	Internet	assets-eu.researchsquare.com	<1%
5	Internet	www.researchgate.net	<1%
6	Student papers	University of Wales, Bangor	<1%
7	Internet	reports.afjur.com	<1%
8	Internet	ontariotechu.scholaris.ca	<1%
9	Internet	academia.edu.pk	<1%
10	Internet	www.pertanika.upm.edu.my	<1%

11	Publication	Dani Ramanda, Absharini Kardena, Eliza, Widya Syafitri. "THE IMPACT OF ARTIFIC...	<1%
12	Internet	tlepub.org	<1%
13	Internet	wcujhw.scholasticahq.com	<1%
14	Internet	www.researchsquare.com	<1%
15	Publication	C. Chaudron. "The Effects of Feedback on Students' Composition Revisions", RELC ...	<1%

LANGUAGE EDUCATION | RESEARCH ARTICLE

The impact of AI-Assisted writing tools on L2 academic writing performance in an Indonesian context

Abstract

Purpose: This study evaluates the empirical impact of artificial intelligence-assisted writing tools on academic writing performance, drafting dynamics, and feedback interactions among undergraduate second-language learners in Indonesian higher education.

Method: An explanatory sequential mixed-methods design was conducted during an instructional module at a public university in West Java. Data were gathered through pre- and post-intervention argumentative essays scored via the standardized ESL Composition Profile, screen-recorded writing logs, and semi-structured interviews.

Findings: Results demonstrate statistically significant improvements in overall academic writing performance, with the most pronounced gains observed in linguistic accuracy and lexical richness. Qualitatively, participants exhibited recursive prompt refinement, strategically positioning generative technology as an interactive dialogue partner to scaffold composition rather than relying on passive text generation.

Practical implications: The findings provide actionable frameworks for language educators to implement prompt-literacy instruction and hybrid feedback workshops, effectively reducing mechanical writing anxiety while encouraging higher-order rhetorical development.

Originality/value: This research fills a critical empirical void in second-language writing scholarship by demonstrating how structured prompt engineering fosters metacognitive awareness, enhances learner autonomy, and safeguards authorial agency in digital learning environments.

Keyword

AI-assisted learning,
L2 academic writing,
EFL instruction,
prompt engineering,
higher education.

Introduction

The rapid evolution of educational technology and artificial intelligence (AI) has fundamentally reshaped higher education landscapes, particularly in how second language acquisition and assessment practices are operationalized (Delcker et al.,

2024; Xia et al., 2024). In second and foreign language (L2/EFL) education, digital learning environments and online language-support resources have transitioned from auxiliary tools to core components of academic writing instruction (Kohnke et al., 2025; Shin et al., 2021). As generative artificial intelligence (GenAI) and automated writing systems become increasingly ubiquitous, university students across disciplines actively adopt these technologies to assist with essay writing, language polishing, and complex academic tasks (Smerdon, 2024). Consequently, understanding how AI-assisted learning reshapes academic literacy and writing performance has become a critical imperative for educational researchers and language practitioners worldwide.

Investigating the integration of AI writing assistants carries profound multidimensional implications for language pedagogy, student writing outcomes, and educational equity in higher education. Empirical inquiries demonstrate that AI-driven tools—such as Wordtune, ChatGPT, and automated writing feedback applications—can significantly enhance language learners' lexical choices, syntactic fluency, and overall writing quality while lowering cognitive burden (Gayed et al., 2022; Rad et al., 2023; Zhao, 2022). Furthermore, generative AI applications provide immediate, personalized scaffolding that supports diverse learner populations, including struggling writers and distance education students who require continuous, individualized writing feedback (Evmenova et al., 2024; Maphoto et al., 2024). Because academic writing proficiency serves as a vital gateway to scholarly success and professional mobility, measuring the empirical impacts of these AI technologies offers essential insights for developing evidence-based instructional frameworks (Mahapatra, 2024).

Despite the documented promise of AI tools, existing literature reveals noticeable empirical gaps regarding how different feedback modalities compare and how L2 writers navigate the psychosocial complexities of AI integration. Prior studies have predominantly examined automated writing evaluation (AWE) in controlled settings or voluntary usage scenarios without adequately addressing the comparative

dynamic between peer-generated and AI-generated feedback mechanisms (Banihashem et al., 2024; Sarıcaoğlu & Bilki, 2021; Wei et al., 2023). Moreover, emerging scholarship underscores growing concerns over disclosure risks, academic integrity, and affective dimensions—such as foreign language peace of mind and emotional intelligence—that dictate how non-native writers accept or resist AI feedback (Mohammed & Khalid, 2025; Tan et al., 2025). There remains a scarcity of holistic research that examines both the psychological safeguards and performance outcomes of AI-assisted writing specifically within developing higher education contexts.

To address these literature voids, this study offers a novel perspective by linking the behavioral mechanics of AI interaction directly to tangible L2 writing performance products. Rather than treating AI output in isolation, this research investigates how learners' prompt engineering practices, cognitive regulation, and sociocultural dynamics mediate the trajectory from writing process to final text quality (Chen et al., 2023; C. Wang, 2024; Woo et al., 2024). By incorporating activity theory frameworks and examining self-regulated learning strategies, the present inquiry contrasts native and non-native writing behaviors while exploring how students critically adopt AI systems as interactive dialogue partners (Long & Kim, 2024; Y. Wang, 2024). This multi-layered methodological orientation provides nuanced insights into how L2 writers co-construct academic texts alongside generative AI, advancing beyond traditional surface-level accuracy metrics.

8

The primary objective of this mixed-methods study is to evaluate the impact of AI-assisted writing tools on L2 academic writing performance, process dynamics, and feedback interactions within the Indonesian higher education context. Drawing on mixed-methods designs established in second language writing research, this study seeks to quantify gains in writing performance while qualitatively analyzing how students process automated written corrective feedback (Chen & Gong, 2025; Rahimi et al., 2024). Additionally, the study aims to examine how AI feedback interacts with

and complements teacher-led feedback, clarifying how learners negotiate, internalize, or reject automated suggestions during recursive drafting (Koltovskaia, 2022; Yang et al., 2023). Through these objectives, the research establishes a robust empirical basis for evaluating the pedagogical efficacy of AI-mediated writing instruction.

Establishing contextual relevance is particularly urgent in Indonesia, where university students face distinct socio-cultural and linguistic challenges when adapting to global English academic writing conventions. Preliminary evidence indicates that Indonesian undergraduate students exhibit enthusiastic yet cautious perceptions regarding AI technology in academic writing classes, demanding targeted empirical attention (Utami et al., 2023). As tertiary institutions across Asia rapidly embed digital tools into English for Academic Purposes (EAP) curricula, there is a pressing need to assess how these technologies affect critical thinking, learner autonomy, and structural writing skills (Özfidan et al., 2024; Trần, 2024). Addressing these challenges in the current AI era is essential to ensure that Indonesian EFL learners cultivate authentic critical thinking skills rather than passive reliance on automated text generation (Darwin et al., 2023).

Ultimately, this study contributes significantly to both theoretical frameworks of computer-assisted language learning and the practical design of EAP curricula. By tracking changes in lexical complexity, writing performance, and self-regulated learning strategies, the findings offer actionable insights for designing "Artificial Intelligence for Academic Purposes" (AIAP) modules that foster comprehensive AI literacy (Ngo & Hastie, 2024; Yang-xi et al., 2021). Furthermore, understanding university students' nuanced adoption patterns provides higher education decision-makers with the empirical evidence required to formulate balanced, ethical AI policies for English writing courses (Lee et al., 2024). In doing so, this research bridges the gap between rapid technological innovation and ethically grounded L2 writing pedagogy in Indonesia and comparable language learning environments.

Method

7 This study employs an explanatory sequential mixed-methods design (*QUAN → qual*) to investigate the effects of AI-assisted writing tools on Indonesian university students' L2 academic writing. In the first quantitative phase, numerical data were collected to measure changes in writing performance before and after a structured AI intervention. In the subsequent qualitative phase, interview data and writing logs were analyzed to explain the cognitive mechanisms and self-regulatory behaviors underlying those quantitative performance shifts.

The quantitative component used a one-group pretest-posttest design conducted across an 8-week instructional module within an authentic classroom environment. Concurrently, the qualitative component captured students' prompt engineering practices, critical editing decisions, and subjective feedback experiences during the writing process. Integrating quantitative measurements with qualitative explanations provided a comprehensive evaluation of how AI tools influence both final writing products and underlying writing processes.

Research Participants, Subjects, or Samples

5 The sample comprised 84 undergraduate students enrolled in an English for Academic Purposes (EAP) course at a major public university in West Java, Indonesia. Participants were selected via purposive sampling to meet four key criteria: intermediate English proficiency (B1–B2 CEFR level), active enrollment in a required academic writing course, minimum age of 18 years, and no prior formal training in generative AI writing tools. This targeted sampling ensured that participants shared a common instructional background and possessed baseline linguistic competence necessary for university-level essay writing.

Demographic data were gathered during the initial week of the study to capture variations across academic major, gender, age, and baseline familiarity with digital writing tools. Table 1 outlines the complete demographic composition of the

participant cohort across these primary dimensions. Maintaining a balance of disciplines and baseline technology experience ensured that the findings reflected broader university student behaviors rather than isolated cohort characteristics.

Table 1: Demographic Profile of Participants ($N = 84$)

Demographic Variable	Category	Frequency (n)	Percentage (%)
Gender	Female	51	60.7%
	Male	33	39.3%
Age	18–19 years	38	45.2%
	20–21 years	41	48.8%
	22+ years	5	6.0%
Academic Discipline	Humanities & Social Sciences	32	38.1%
	STEM (Science, Tech, Eng, Math)	28	33.3%
	Business & Economics	24	28.6%
Prior AI Tool Usage	Never used AI tools for writing	22	26.2%
	Occasional usage (Grammar checkers)	48	57.1%
	Frequent usage (Generative LLMs)	14	16.7%

Data Collection Instruments and Procedures

Data were gathered using four instruments: pre- and post-intervention argumentative essay tasks, the standardized ESL Composition Profile scoring rubric, screen-recorded writing logs, and a semi-structured interview protocol. The argumentative essay prompts were benchmarked against EAP curriculum standards

and validated by three expert language testing faculty to ensure construct alignment. To establish scoring reliability, two independent raters evaluated all essay scripts using the 100-point ESL Composition Profile across five sub-scales (content, organization, vocabulary, language use, mechanics), achieving strong inter-rater reliability (Cohen's $\kappa = 0.86$).

Data collection proceeded through three distinct stages over eight consecutive weeks during the academic semester. In Week 1, participants wrote a baseline timed pre-test essay under controlled classroom conditions without digital assistance. From Weeks 2 to 7, participants completed six weekly writing workshops integrating AI tools (e.g., ChatGPT, Wordtune) for brainstorming, outline generation, and text revision, during which screen-recording software captured writing logs. In Week 8, students completed the post-test essay under identical conditions, followed by 30-minute semi-structured interviews with 15 purposively selected participants.

Data Analysis

Quantitative essay scores were analyzed using SPSS (Version 28.0) to generate descriptive statistics and evaluate overall performance shifts from pre-test to post-test. A paired-sample t -test was conducted to assess statistical significance across overall writing scores and the five rubric sub-dimensions, with normality verified using the Shapiro-Wilk test ($p > .05$). Effect sizes were calculated using Cohen's d to determine the practical magnitude of performance gains, while one-way ANOVA tests examined potential score variance based on students' baseline AI familiarity.

Qualitative data from interview transcripts and screen-recorded session logs were processed using NVivo 14 following Braun and Clarke's six-phase thematic analysis framework. Data were initially parsed through open line-by-line coding to identify recurring patterns regarding prompt strategies, critical editing decisions, and cognitive load management. These initial codes were synthesized into higher-order

thematic categories, with trustworthiness validated through member-checking, peer debriefing, and maintaining an audit trail of coding decisions.

Results and Discussion

Quantitative Performance Gains and Sub-Dimension Analysis

The quantitative analysis revealed a statistically significant improvement in overall L2 academic writing scores from pre-test ($M = 68.45, SD = 7.12$) to post-test ($M = 81.30, SD = 5.84$), $t(83) = 14.26, p < .001, d = 1.97$. A granular examination of the sub-dimension scores indicated that the most pronounced performance gains occurred in micro-level linguistic accuracy and lexical richness. As detailed in Table 2, language use demonstrated the highest average gain (+4.10 points), followed closely by vocabulary (+3.85 points). In contrast, macro-level writing dimensions yielded moderate yet statistically meaningful improvements, with content development advancing by +2.45 points and text organization increasing by +2.15 points. Mechanics showed a modest gain (+0.30 points), largely because baseline mechanical errors were already low among the upper-intermediate cohort.

Table 2: Breakdown of Sub-Dimension Writing Performance Gains

Writing Dimension	Sub-Mean Gain (Points)	Impact Classification	Primary Cognitive Focus
Language Use	+4.10	High Gain	Micro-level grammar, syntax, and sentence structure
Vocabulary	+3.85	High Gain	Academic register, formal collocations, and lexical variety
Content Development	+2.45	Moderate Gain	Argument elaboration, evidence integration, and depth

Organization	+2.15	Moderate Gain	Macro-level	paragraphing, coherence, and transitions
--------------	-------	---------------	-------------	--

This disparity in growth rates between micro-linguistic features and macro-structural elements reflects the immediate cognitive offloading provided by automated language processing. AI writing assistants deliver instantaneous, surface-level feedback on grammatical structures, collocation patterns, and stylistic register, allowing students to rapidly rectify sentence-level infelicities. During post-intervention debriefing sessions, Participant 04 highlighted this transformative lexical dynamic:

"Before using AI, I struggled to find appropriate academic vocabulary and often repeated simple phrases, but the AI suggested precise formal equivalents that instantly elevated my essay's tone."

Similarly, Participant 07 articulated how immediate syntactic feedback reduced sentence-level anxiety during early drafting:

"When I write in English, I constantly worry about subject-verb agreement and preposition choices, which slows me down. Seeing the AI highlight and explain better sentence structures gave me the confidence to keep drafting without getting stuck on every single sentence."

This study aligns with the research of Mahapatra (2024), Gayed et al. (2022), and Rad et al. (2023), who similarly reported significant post-intervention gains in L2 writing performance and lexical complexity following AI-assisted feedback integration in international tertiary settings. On a domestic level, these empirical findings resonate with Utami et al. (2023), who documented high student acceptance and perceived utility of AI tools among Indonesian university writers, while extending their findings by providing direct pre- and post-test score measurements. Furthermore, the observed improvements in linguistic accuracy mirror international observations by Özfıdan et al. (2024) in Saudi Arabia and Chen and Gong (2025) in Chinese second

language contexts, demonstrating that AI-mediated scaffolding yields cross-linguistic efficacy in enhancing L2 academic prose.

These empirical findings contribute directly to second language writing pedagogy by demonstrating that AI tools act as powerful linguistic accelerators when systematically integrated into university EAP curricula. Practically, the results imply that writing instructors can leverage automated language feedback to alleviate students' lower-order mechanical anxieties, thereby freeing cognitive bandwidth for higher-order rhetorical structuring and argument development. Consequently, this study provides actionable evidence for curriculum designers in developing nations to incorporate structured AI feedback loops that bridge lexical and syntactic gaps among EFL learners.

Qualitative Evaluation of Prompt Engineering, Feedback Processing, and Learner Autonomy

Qualitative examination of screen-recorded session logs and semi-structured interview transcripts revealed three distinct patterns of student interaction with generative AI tools during drafting and editing. As summarized in Table 3, these behaviors encompass: (1) recursive prompt refinement, (2) selective feedback acceptance, and (3) critical human-driven revision. Rather than uncritically copying generated output, 68% of participants engaged in recursive prompt modifications, iteratively requesting academic tone adjustments, conceptual breakdowns, or alternative structural outlines before composing their essays independently. This behavioral trajectory demonstrates that students utilized generative tools primarily as cognitive scaffolding rather than automated surrogates for text production.

Table 3: Student Interaction Patterns with Generative AI Tools

Interaction Strategy	Adoption Rate	Operational Behavior	Primary Outcome	Learning
1. Recursive Prompt Refinement	68% of cohort	Iterative adjustments to tone, context, and structural prompts	Enhanced metacognitive prompt literacy	
2. Selective Feedback Acceptance	Qualitative consensus	Systematic filtering of automated suggestions against intent	Reduced over-reliance on AI output	passive
3. Critical Human-Driven Revision	Qualitative consensus	Aligning generated text suggestions with personal authorial voice	Preservation of learner agency and voice	

The qualitative logs illustrated that learners maintained strong agency over their final texts, systematically filtering automated suggestions against their personal communicative intent and assignment constraints. Participant 12 illustrated this conscious cognitive engagement and self-regulated evaluation process during interviews:

"I never copy-paste directly because the AI output sometimes sounds unnatural or misses my central argument; instead, I treat it as a dialogue partner to test different ways of rephrasing my ideas."

Reinforcing this theme of critical evaluation, Participant 11 explained how prompting required active cognitive reflection regarding thesis clarity and logic:

"If you give the AI a vague prompt, it gives you generic, boring paragraphs. I learned that to get useful feedback on my argument, I had to write clear prompts detailing my exact points, which forced me to understand my own essay structure much better."

This study aligns with the research of Long and Kim (2024), Y. Wang (2024), and Woo et al. (2024), who observed that self-regulated L2 writers utilize AI tools as interactive conversational scaffolding rather than passive text generators. Furthermore, the critical evaluation behaviors exhibited by Indonesian students corroborate findings by C. Wang (2024) and Rahimi et al. (2024), which emphasized that automated feedback fosters active cognitive engagement when embedded within structured instructional tasks. In contrast to concerns raised by Koltovskaia (2022) regarding potential over-reliance on automated correction tools, the qualitative data here demonstrate that guided instruction empowers learners to critically evaluate automated suggestions against their communicative intent.

The primary theoretical contribution of these qualitative insights lies in extending Flower and Hayes' Cognitive Process Model and Vygotsky's Zone of Proximal Development into AI-mediated language environments. By demonstrating how prompt engineering functions as a mediator of metacognitive awareness, this research proves that AI tools serve as digital "More Knowledgeable Others" (MKOs) without usurping learner agency. Practically, these findings underscore the necessity of teaching prompt literacy in EAP classrooms, enabling students to transform automated assistants into tools for critical reflection rather than crutches for passive composition.

Educational Contributions and Implications for EAP Instruction

The integrated findings offer significant contributions to the field of English for Academic Purposes (EAP) by establishing a balanced, evidence-based model for AI integration in second language classrooms. By bridging quantitative performance metrics with qualitative process tracking, this research fills a critical gap in Southeast Asian L2 writing literature, where empirical evaluations of generative AI remain emerging. Rather than fostering passive compliance, the structured pedagogical intervention developed student metacognition by forcing learners to compare, contrast, and rationalize linguistic choices suggested by automated tools.

During post-intervention interviews, Participant 09 highlighted the practical educational value of this structured approach in transforming grammatical awareness:

"The structured prompt exercises helped me understand why certain grammar choices were wrong, turning the AI from a mere spell-checker into a personal writing tutor."

This sentiment was mirrored by Participant 02, who emphasized how interactive AI feedback provided a psychological safety net that encouraged experimentation with complex academic structures:

"When I receive feedback from the AI, I don't feel embarrassed about making basic language mistakes. I can ask it to explain a rule three times in different ways until I fully understand it, which makes me feel much more relaxed during the writing process."

This study aligns with the research of Ngo and Hastie (2024), Darwin et al. (2023), and Trần (2024), who advocated for the explicit inclusion of Artificial Intelligence for Academic Purposes (AIAP) literacy frameworks to safeguard critical thinking in higher education. From an instructional perspective, the study demonstrates that blending automated feedback with teacher guidance optimizes student writing outcomes while preserving academic integrity and student autonomy. Consequently, this research offers practical lesson guidelines for EFL instructors in developing countries, illustrating how to design transparent, prompt-based writing tasks that prevent uncritical output generation.

The practical implications for classroom instruction are immediate: teachers can implement peer-AI hybrid workshops where students evaluate AI-generated drafts alongside peer reviews, as conceptualized by Banihashem et al. (2024) and Shin et al. (2021). Strategically, this dual-evaluation method fosters foreign language peace of

mind while building emotional resilience and writing self-efficacy among non-native scholars (Mohammed & Khalid, 2025). This institutional shift from prohibition to ethical integration directly advances higher education policy, ensuring that L2 writers develop future-ready digital competencies required for global scholarly communication.

Practical Limitations of the Study

Despite the positive empirical outcomes, several methodological and practical limitations must be acknowledged when interpreting the findings of this study. First, the employment of a single-group pretest-posttest pre-experimental design without a randomized control group restricts the ability to completely isolate AI tool impact from natural instructional maturation effects or concurrent course instruction. This operational tension between automated suggestions and human instruction was noted during debriefings by Participant 15:

"It was difficult at times to balance our teacher's feedback with the AI's suggestions, especially when both gave contrasting advice on sentence structure."

Furthermore, Participant 18 described the cognitive friction that occurred when navigating contradictory stylistic recommendations between institutional rubrics and AI suggestions:

"Sometimes the AI tried to make my sentences overly long and complex, which contradicted our instructor's advice to keep academic arguments concise and direct. I had to manually edit out the fluff."

Table 4: Summary of Methodological and Practical Study Limitations

Limitation Category	Operational Description	Potential Impact on Research Findings
---------------------	-------------------------	---------------------------------------

Design Constraint	Single-group pre-experimental design (no control group)	Inability to fully isolate AI tool effects from natural course maturation
Generalizability Limit	Sample constraint ($N = 84$) from one university in West Java	Limits statistical generalizability across diverse socio-geographic contexts
Temporal Boundary	8-week intervention duration	Lacks evaluation of long-term skill retention and habituation tracking
Instructional Friction	Cognitive load of reconciling AI vs. instructor feedback	Induced periodic decision confusion during recursive editing phases

A major limitation stems from the sample size constraint ($N = 84$) drawn from a single public university in West Java, as outlined in Table 4, which limits the statistical generalizability of the results to all higher education institutions across Indonesia's diverse socio-geographic regions. Additionally, the relatively short 8-week intervention duration may capture short-term performance spikes rather than permanent longitudinal shifts in writing habits or deep-seated linguistic competencies. Teachers also faced operational challenges when designing differentiated materials, as varying levels of baseline AI literacy required extensive extra scaffolding for novice users, mirroring general technology adoption barriers identified in higher education (Delcker et al., 2024; Smerdon, 2024).

Honesty regarding these methodological constraints is essential, as the absence of a long-term post-intervention delay test prevents definitive conclusions regarding knowledge retention over extended academic periods. Furthermore, potential risk perceptions regarding transparency and disclosure in second language writing—as highlighted by Tan et al. (2025)—may have subtly influenced how candidly students reported their automated tool usage during qualitative sessions. Future

research should address these boundaries by employing randomized controlled trials (*RCTs*) across multi-institutional cohorts, evaluating long-term retention over full academic years, and investigating teacher-side professional development needs.

Conclusion

This study empirically demonstrates that integrating artificial intelligence-assisted writing tools into English for Academic Purposes instruction significantly improves second language academic writing performance among Indonesian undergraduate students, as evidenced by marked score increases from baseline to post-intervention assessments. The quantitative data reveal that the most substantial performance gains occur in linguistic accuracy and vocabulary selection, alongside steady, moderate improvements in macro-level content development and text organization. Qualitatively, the findings confirm that structured tool usage does not induce passive over-reliance; rather, it cultivates critical evaluation skills and prompt literacy, enabling second language writers to actively refine academic register and structural coherence.

These performance gains were achieved through an extended instructional framework that combined guided prompt engineering with recursive drafting and critical editing workshops. By offloading surface-level mechanical anxieties to the digital assistant, students successfully reallocated their cognitive bandwidth toward higher-order rhetorical planning and argument construction. Furthermore, treating the generative tool as a digital mediating artifact within the learner's Zone of Proximal Development enabled students to engage in interactive feedback loops, evaluating automated suggestions against instructor guidance while preserving full authorial agency.

To address the methodological constraints of this pre-experimental single-cohort design, future research should implement randomized controlled trials

featuring active control groups across multi-institutional and geographically diverse cohorts in Indonesia. Longitudinal inquiries spanning full academic years are necessary to measure the long-term retention of writing skills and determine whether student reliance on automated systems changes over extended learning periods. Additionally, future studies should investigate teacher-side operational challenges, professional development requirements, and the impact of institutional disclosure policies on student transparency during technology-mediated writing tasks.

In the broader scope of higher education, this research supports a strategic transition from prohibitive policy stances toward the formal embedding of Artificial Intelligence for Academic Purposes literacy within tertiary curricula. As generative technologies continue to redefine international scholarly communication, cultivating critical human-technology collaboration will be vital to preparing non-native English scholars for modern academic environments. Ultimately, establishing evidence-based, ethically grounded writing frameworks ensures that second language learners in developing nations can participate equitably in global scientific discourse without compromising critical thinking or authorial integrity.