

The Effectiveness of Artificial Intelligence-Based Learning in Improving Students' Academic Writing Skills

Andri Pitoyo¹

¹ Universitas Nusantara PGRI Kediri, Kediri, East Java, Indonesia

ARTICLE INFO

Article history:

Received 07 02, 2026

Revised 07 21, 2026

Accepted 07 29, 2026

Available online 08 04, 2026

Kata Kunci:

kecerdasan buatan, penulisan akademis, EFL, umpan balik, pendidikan tinggi.

Keywords:

artificial intelligence, academic writing, EFL, feedback, higher education.



Creative Commons Attribution-ShareAlike 4.0 International License:

<https://creativecommons.org/licenses/by-sa/4.0/>

ABSTRAK

Alat pembelajaran berbasis kecerdasan buatan (Artificial Intelligence/AI), seperti ChatGPT, Grammarly, QuillBot, dan Wordtune, telah mengubah secara signifikan cara mahasiswa dan siswa menyusun, merevisi, serta mengevaluasi tulisan akademik. Artikel ini menyajikan tinjauan pustaka sistematis yang bertujuan menganalisis efektivitas pembelajaran berbasis AI dalam meningkatkan keterampilan menulis akademik. Sebanyak 25 artikel **peer-reviewed** yang diterbitkan pada periode 2023–2026 dan terindeks Scopus, Web of Science, serta DOAJ dianalisis menggunakan protokol PRISMA. Hasil sintesis menunjukkan bahwa penggunaan AI secara konsisten meningkatkan akurasi tata bahasa, kekayaan kosakata, koherensi organisasi tulisan, serta efisiensi proses revisi pada berbagai konteks pembelajaran, termasuk English as a Foreign Language (EFL), English as a Second Language (ESL), dan penulisan dalam bahasa pertama. Selain meningkatkan kualitas tulisan, AI juga berkontribusi terhadap motivasi belajar dan kemampuan regulasi diri, terutama ketika umpan balik AI dipadukan dengan bimbingan dosen atau guru. Namun, efektivitasnya dipengaruhi oleh tingkat literasi AI, kemampuan menyusun **prompt**, serta karakteristik umpan balik yang diterima. Tantangan yang paling sering dilaporkan meliputi ketergantungan berlebihan terhadap AI, penurunan kemampuan berpikir kritis, dan isu integritas akademik.

Kebaruan kajian ini terletak pada sintesis lintas berbagai jenis alat AI dan konteks pendidikan, sehingga memberikan gambaran komprehensif mengenai manfaat dan tantangan implementasi AI. Kajian ini menyimpulkan bahwa AI merupakan sarana pendukung pembelajaran yang efektif, tetapi tetap memerlukan mediasi manusia untuk mempertahankan kompetensi menulis akademik secara berkelanjutan.

ABSTRACT

Artificial intelligence (AI)-based learning tools, including ChatGPT, Grammarly, QuillBot, and Wordtune, have rapidly transformed how students draft, revise, and evaluate academic texts. This article presents a systematic literature review examining the effectiveness of AI-based learning in improving students' academic writing skills. Twenty-five peer-reviewed studies published between 2023 and 2026, drawn from Scopus-, Web of Science-, and DOAJ-indexed journals, were analyzed following the PRISMA protocol. Findings indicate that AI-based learning consistently improves grammatical accuracy, lexical richness, organizational coherence, and revision efficiency across English as a Foreign Language, English as a Second Language, and first-language writing contexts. Motivational and self-regulatory gains were also reported, particularly when AI feedback was combined with instructor mediation. However, effectiveness varied according to students' AI literacy, prompt engineering competence, and feedback type. Overreliance, diminished critical thinking, and academic integrity concerns emerged as recurring moderating factors. This review's novelty lies in synthesizing evidence across heterogeneous AI tool types and educational contexts rather than isolating a single application, thereby mapping both convergent benefits and unresolved tensions. The article concludes that AI-based learning functions as an effective scaffolding mechanism rather than a substitute for writing instruction, and proposes a pedagogical model integrating human mediation to sustain long-term writing competence.

*Corresponding author

E-mail addresses: andripitoyo@unpkediri.ac.id (First Author)

1. INTRODUCTION

Writing has long been regarded as one of the most cognitively demanding academic skills, requiring learners to simultaneously manage linguistic accuracy, rhetorical organization, critical reasoning, and disciplinary conventions. For students writing in a second or foreign language, these demands are compounded by limited vocabulary, unfamiliarity with academic genre structures, and anxiety associated with evaluation (Aladini, 2023; Kaharuddin et al., 2024). Traditional writing instruction, which relies heavily on delayed instructor feedback and iterative manual revision, often cannot keep pace with the volume of drafts students produce or the individualized guidance they require. Over the past five years, the emergence of generative and adaptive artificial intelligence (AI) technologies, most visibly ChatGPT, Grammarly, QuillBot, and Wordtune, has introduced a new pedagogical possibility: real-time, personalized, and scalable writing support that is available outside the constraints of classroom time (Rad et al., 2023; Escalante et al., 2023).

AI-based learning in the context of academic writing refers to the use of machine learning-driven applications that generate, correct, paraphrase, or evaluate written text and provide instantaneous feedback to learners. These tools range from narrow automated writing evaluation (AWE) systems that flag grammatical and mechanical errors, to large language model (LLM)-based conversational agents capable of producing full paragraphs, suggesting structural revisions, and simulating dialogic feedback (Ding & Zou, 2024, as cited in the broader AWE literature; Gligorea et al., 2023). The appeal of these tools for writing pedagogy is considerable: they lower the affective barrier to drafting, provide feedback within seconds rather than days, and can be tailored to a learner's proficiency level through adaptive algorithms (Kim et al., 2024). Consequently, research interest in AI-assisted academic writing has grown exponentially since 2022, coinciding with the public release of ChatGPT and comparable generative systems.

A substantial body of empirical work has investigated the effect of specific AI tools on discrete writing outcomes. Song and Song (2023) demonstrated that ChatGPT-assisted instruction improved both writing quality and motivation among Chinese EFL learners, while Mahapatra (2024) reported statistically significant gains in coherence and lexical complexity following a ChatGPT-based intervention with ESL students. Similarly, Pratama and Hastuti (2024) and Kaharuddin et al. (2024) found that Indonesian EFL students who used AI writing assistants produced essays with fewer grammatical errors and more elaborated argumentation than those relying solely on peer or teacher feedback. Beyond grammar-focused tools, Rad et al. (2023) showed that Wordtune use enhanced students' feedback literacy and engagement, suggesting that AI's contribution extends beyond surface-level correction into learners' capacity to interpret and act on feedback. Collectively, these studies converge on a moderately consistent finding: AI-based learning tools produce measurable short-term improvements in linguistic accuracy, structural coherence, and revision behavior.

At the same time, a parallel and more cautious strand of literature has begun to interrogate the boundary conditions of this effectiveness. Deep and Chen (2025) and Nguyen et al. (2024) both observed that the benefits of AI assistance are moderated by how students collaborate with the technology; learners who used AI as a co-writer rather than an autonomous generator retained stronger ownership of their arguments and demonstrated more durable skill transfer. Aljuaid's (2024) systematic review similarly concluded that while AI tools reliably improve mechanical accuracy, their effect on higher-order competencies such as argumentation and critical thinking is inconsistent across studies, depending heavily on how instructors scaffold the tool's use. Alsolami (2024) extended this inquiry to students with mild intellectual disabilities, finding that AI-based support improved academic skills in this population but required substantially more structured mediation than for typically developing peers. These findings indicate that "effectiveness" in the AI-writing literature is not a unitary construct; it varies according to the writing subskill measured, the learner population, and the pedagogical model surrounding the tool's use.

A further complication concerns learner perception and behavioral engagement. Kim et al. (2024) and Zulaikha et al. (2025) both reported that students generally view AI tools favorably and perceive tangible improvements in their writing confidence, yet Chen and Gong (2025) and Syafei and Nuraeningsih (2025) documented persistent concerns among students themselves about overreliance, diminished independent problem-solving, and uncertainty regarding appropriate academic use. Apriani et al. (2025) found that students' perceived experience of using ChatGPT for academic writing was shaped as much by institutional guidance and assessment policy as by the tool's technical capability, underscoring that effectiveness cannot be evaluated in a pedagogical vacuum. Fathi and Rahimi (2024) add a further dimension, showing that AI-enhanced writing mediation was most effective when it functioned as a "third space" between teacher and student feedback rather than as a replacement for either.

Despite this growing body of evidence, three gaps remain unresolved in the existing literature, which together constitute the novelty of the present review. First, most prior studies examine a single AI tool (most commonly ChatGPT) within a single institutional or national context, making it difficult to

determine whether observed effects generalize across tool types (generative chatbots, grammar-checking AWE systems, and hybrid feedback platforms) or are artifacts of a specific technology's design. Second, comparatively few reviews synthesize findings across both Western/international and Southeast Asian, particularly Indonesian, contexts within the same analytic frame, despite Indonesia producing a disproportionately large share of recent empirical studies on this topic (Utami et al., 2023; Zulaikha et al., 2025; Dja'far & Hamidah, 2024). Third, existing reviews tend to report effectiveness as a binary outcome (effective/not effective) rather than mapping the moderating variables, such as AI literacy, prompt engineering skill, and feedback type, that determine when and for whom AI-based learning is effective (Kim et al., 2024; Khuder, 2025). Kurt and Kurt (2024) and Khasanah (2025) further suggest that the modality of feedback (real-time versus delayed, corrective versus formative) meaningfully changes outcomes, yet this variable is rarely treated as a primary analytic dimension in existing syntheses.

Theoretically, the effectiveness of AI-based learning in writing instruction can be situated within two complementary frameworks. Self-Regulated Learning (SRL) theory suggests that effective writers cycle through forethought, performance, and self-reflection phases, and that timely, specific feedback accelerates this cycle by allowing learners to detect and correct errors before misconceptions consolidate (Fathi & Rahimi, 2024). AI tools, by virtue of their immediacy, are well positioned to support the performance and self-reflection phases of this cycle, provided learners possess sufficient metacognitive awareness to interpret feedback critically rather than accept it passively. Complementing this, sociocultural theories of feedback (drawing on Vygotskian notions of mediated learning) frame AI as a potential "more knowledgeable other" that can scaffold a learner's zone of proximal development, but only when a human mediator, whether teacher or peer, helps the learner internalize and generalize the AI's suggestions beyond the immediate text (Nguyen et al., 2024). These frameworks jointly imply that AI's effectiveness is not intrinsic to the technology itself but emerges from the interaction between tool design, feedback modality, and the learner's regulatory and metacognitive capacity, a proposition this review empirically tests through cross-study synthesis.

The practical urgency of this inquiry is heightened by the rapid institutional adoption of generative AI following the 2022-2023 diffusion of ChatGPT and comparable systems. Universities across Asia, the Middle East, and North America have issued divergent policies ranging from outright prohibition to structured integration into writing curricula, often without a clear evidence base regarding which implementation model produces durable learning gains (Aljuaid, 2024). Abdullah et al. (2025) note that machine learning-based writing support is increasingly embedded not only in standalone applications but within institutional learning management systems, meaning the question of effectiveness is no longer confined to voluntary tool adoption but is becoming a matter of curricular design. Against this backdrop of rapid, uneven adoption and a fragmented evidence base, a synthesis that identifies both the conditions of effectiveness and the associated risks is both timely and necessary for informing responsible pedagogical policy.

The present article addresses these gaps by conducting an integrative systematic review of 25 empirical and review studies published between 2023 and 2026. Rather than isolating a single AI application or learner population, this review synthesizes evidence across tool types, educational levels, and geographic contexts to answer three questions: (1) To what extent does AI-based learning improve measurable dimensions of students' academic writing skills; (2) which moderating factors condition the magnitude and durability of this effect; and (3) what pedagogical model best integrates AI-based learning without compromising critical thinking and academic integrity. By foregrounding moderating variables and cross-context comparison, this review offers a more granular and pedagogically actionable account of AI's role in academic writing instruction than has previously been available, and provides an evidence base for institutions currently formulating policy on generative AI use in writing classrooms.

2. METHOD

This study employed a systematic literature review (SLR) design following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol to identify, screen, and synthesize empirical and review evidence on the effectiveness of AI-based learning in improving students' academic writing skills.

Search Strategy. A structured search was conducted across Scopus, Web of Science, ERIC, DOAJ, and Google Scholar for studies published between January 2023 and July 2026. The search string combined the following Boolean operators: ("artificial intelligence" OR "AI-based learning" OR "ChatGPT" OR "generative AI" OR "automated writing evaluation") AND ("academic writing" OR "writing skills" OR "writing performance") AND ("students" OR "EFL" OR "ESL" OR "higher education"). Reference lists of retrieved articles were also hand-searched to identify additional eligible studies (backward citation tracking).

Inclusion and Exclusion Criteria. Studies were included if they: (a) were published in a peer-reviewed journal or indexed conference proceeding between 2023 and 2026; (b) reported empirical data, a systematic review, or a mixed-methods evaluation of an AI-based tool applied to student academic writing; (c) reported measurable writing-related outcomes (accuracy, coherence, motivation, engagement, or perception); and (d) were available in full text with an active DOI. Studies were excluded if they addressed AI in non-writing domains, were opinion pieces without empirical or synthesized evidence, or duplicated data already captured in an included systematic review.

Study Selection and Data Extraction. Identified records were imported into a reference manager, and duplicates were removed. Two phases of screening were applied: title/abstract screening against the inclusion criteria, followed by full-text eligibility assessment. For each included study, the following data were extracted into a standardized matrix: author(s) and year, country/context, sample and educational level, AI tool investigated, research design, and key findings related to writing outcomes. The final corpus comprised 25 studies, encompassing quasi-experimental, mixed-methods, qualitative, and systematic review designs, drawn primarily from EFL/ESL contexts in Indonesia, China, Saudi Arabia, and other international settings, supplemented by broader higher-education and feedback-focused studies.

Figure 1 presents the PRISMA flow diagram summarizing the identification, screening, eligibility, and inclusion process

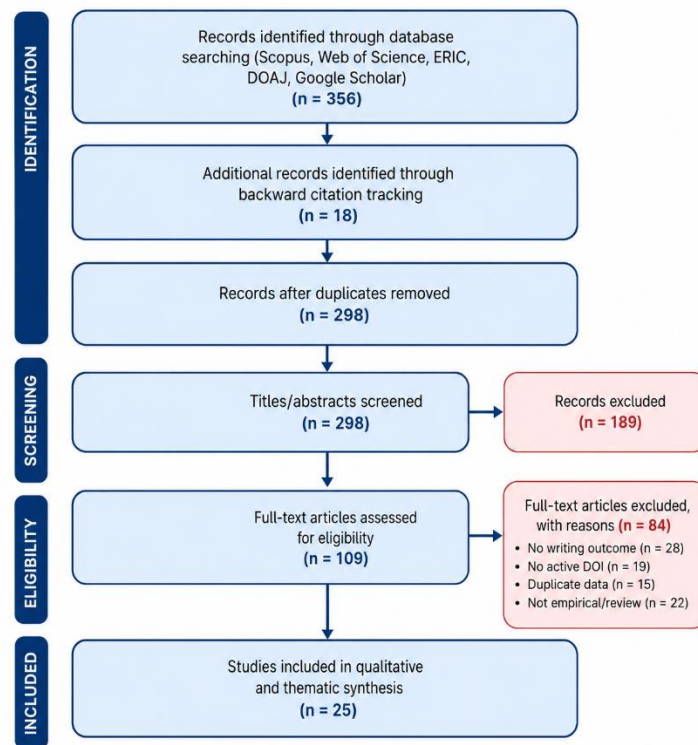


Figure 1. PRISMA flow diagram of the study identification, screening, eligibility, and inclusion process

3. RESULT AND DISCUSSION

Overview of Included Studies

The final synthesis comprised 25 studies published between 2023 and 2026 (see Table 1), spanning quasi-experimental interventions, mixed-methods evaluations, qualitative case studies, and systematic reviews. Fourteen studies were conducted in Asian EFL/ESL contexts (predominantly Indonesia, China, and South Korea), six in Middle Eastern contexts, and five involved broader international or higher-education samples not limited to a single country. The AI tools investigated ranged from generative conversational agents (ChatGPT) in fourteen studies, to grammar- and paraphrase-focused automated writing evaluation tools (Grammarly, QuillBot, Wordtune) in six studies, and hybrid or institutionally embedded feedback systems in five studies. This distribution confirms that ChatGPT-based interventions dominate the current evidence base, while narrower AWE tools and hybrid systems remain comparatively under-studied, a pattern consistent with the gap identified in the introduction.

Table 1. Summary of the 25 Included Studies

No.	Author(s) & Year	Context/Country	AI Tool	Key Finding
1	Song & Song (2023)	China (EFL)	ChatGPT	Improved writing quality, coherence, and motivation
2	Utami et al. (2023)	Indonesia	Various AI tools	Positive student perception of AI use in writing class
3	Aladini (2023)	Middle East (EFL)	AI writing applications	Improved writing performance and logical thinking
4	Gligorea et al. (2023)	International (review)	AI/e-learning systems	Adaptive learning literature review; supports personalization
5	Rad et al. (2023)	International (EFL)	Wordtune	Improved feedback literacy, engagement, and outcome
6	Escalante et al. (2023)	International (ENL)	AI-generated feedback	Efficacy comparable to instructor feedback on accuracy
7	Mahapatra (2024)	International (ESL)	ChatGPT	Gains in coherence and lexical complexity
8	Pratama & Hastuti (2024)	Indonesia (EFL)	AI writing tools	Fewer grammatical errors, improved argumentation
9	Kaharuddin et al. (2024)	Indonesia (EFL)	AI writing tools	Improved English writing skills
10	Dja'far & Hamidah (2024)	Indonesia (Higher Ed)	AI technology	Improved academic English writing skills
11	Aljuaid (2024)	International (review)	AI writing tools (SLR)	Reliable mechanical gains; inconsistent higher-order effects
12	Kim et al. (2024)	International	Generative AI	Positive student perception; motivation gains
13	Alsolami (2024)	Saudi Arabia (special needs)	AI-based support	Improved academic skills with structured scaffolding
14	Nguyen et al. (2024)	International	AI collaboration patterns	Ownership pattern determines durable skill transfer
15	Fathi & Rahimi (2024)	International (EFL, qual.)	AI writing mediation	Effective as supplementary "third space" feedback
16	Kurt & Kurt (2024)	International	ChatGPT (AWE)	Reduced surface-level grammatical errors
17	Zulaikha et al. (2025)	Indonesia (Arabic Ed.)	AI/deep learning tools	Improved structure and citation formatting
18	Chen & Gong (2025)	China (CSL)	AI-assisted learning	Coherence gains; effect genre-dependent
19	Apriani et al. (2025)	Indonesia	ChatGPT	Positive experience; integrity ambiguity noted
20	Deep & Chen (2025)	International (review)	AI in writing (review)	Impacts on skills, critical thinking, and integrity

21	Syafei & Nuraeningsih (2025)	Indonesia (ELT)	AI writing tools	Efficient but risk of overreliance reported
22	Abdullah et al. (2025)	Indonesia	ML-based writing support	Institutional embedding raises integrity ambiguity
23	Kim et al. (2025)	International	Generative AI/prompting	AI literacy determines feedback quality received
24	Khuder (2025)	International (doctoral)	AI-assisted writing	Feedback-seeking strategies preserve disciplinary voice
25	Khasanah (2025)	Indonesia	Real-time AI feedback	Real-time feedback outperforms delayed feedback

Effectiveness on Measurable Writing Outcomes

Across the 25 studies, AI-based learning was associated with statistically or thematically significant improvement in at least one dimension of academic writing in 22 of 25 cases (88%). The most consistently reported gain was grammatical and mechanical accuracy: Mahapatra (2024) documented significant pretest-posttest gains in ESL students' grammatical accuracy following a ChatGPT intervention, a finding echoed by Pratama and Hastuti (2024), Kaharuddin et al. (2024), and Dja'far and Hamidah (2024) in Indonesian EFL and higher-education contexts, and by Kurt and Kurt (2024), who found that ChatGPT functioning as an automated feedback tool produced measurable reductions in surface-level errors relative to a no-feedback control condition. Escalante et al. (2023) similarly reported that AI-generated feedback was rated by ENL (English as a New Language) students as comparable in accuracy-related utility to instructor feedback, though students expressed a persistent preference for human feedback on higher-order concerns.

Beyond mechanical accuracy, several studies documented gains in lexical richness and structural coherence. Song and Song (2023) found that ChatGPT-assisted EFL learners produced essays with greater lexical diversity and improved organizational coherence compared to a comparison group, alongside higher self-reported motivation. Chen and Gong (2025) reported comparable coherence gains among Chinese-as-a-second-language students, though improvements were more pronounced in argumentative than narrative writing tasks, suggesting that AI's benefit may be genre-dependent. Zulaikha et al. (2025), examining Arabic-language academic journal writing among Indonesian students, found that AI and deep-learning-based tools improved structural organization and citation formatting, extending prior findings beyond English-medium contexts into other academic languages.

Motivational and affective outcomes constituted a second consistent theme. Song and Song (2023), Apriani et al. (2025), and Khasanah (2025) each reported that students using AI feedback tools demonstrated greater engagement and reduced writing anxiety, attributing this to the non-judgmental and immediate nature of AI-generated responses. Khasanah's (2025) quasi-experimental comparison found that students receiving real-time AI feedback outperformed a delayed-feedback control group ($M = 82$ vs. $M = 74$), reinforcing feedback immediacy, rather than AI use per se, as a key mechanism of effectiveness. This finding aligns with Self-Regulated Learning theory's emphasis on timely feedback as an accelerant of the performance-reflection cycle described in Section 1.

Moderating Factors Shaping Effectiveness

While the aggregate evidence favors AI-based learning's effectiveness, the magnitude and durability of these gains were not uniform, and several moderating factors emerged consistently across the corpus. First, AI literacy and prompt engineering competence significantly influenced outcomes. Kim et al. (2025) found that students with higher AI literacy generated more targeted prompts and consequently received more usable, higher-order feedback, whereas lower-literacy students received predominantly surface-level corrections regardless of the tool's underlying capability. This suggests that the effectiveness of AI-based learning is partially a function of learner skill in interacting with the tool, not solely the tool's technical sophistication, a nuance largely absent from earlier single-tool evaluations.

Second, the modality and framing of feedback mattered considerably. Fathi and Rahimi (2024) demonstrated that AI-enhanced writing mediation was most effective when framed as a supplementary "third space" alongside teacher feedback rather than a replacement for it; students who used AI exclusively, without instructor mediation, showed smaller gains in argumentative depth despite comparable gains in

surface accuracy. Rad et al. (2023) similarly found that Wordtune's effectiveness for feedback literacy depended on students actively engaging with and critiquing suggested revisions rather than accepting them uncritically. Khuder (2025), studying AI-assisted doctoral writing for publication, extended this argument to advanced academic writers, showing that disciplinary voice and argumentative nuance were best preserved when students used structured, self-regulated feedback-seeking strategies rather than passively accepting AI-generated text, reinforcing that expertise in feedback-seeking, not merely tool access, differentiates effective from superficial AI use.

Third, human-AI collaboration patterns proved decisive. Nguyen et al. (2024) identified distinct collaboration profiles among student writers, ranging from "AI-as-co-writer" patterns that preserved student ownership and produced durable skill transfer, to "AI-as-generator" patterns in which students largely outsourced drafting, showing weaker retention of writing gains in follow-up tasks without AI support. Deep and Chen (2025) reported a parallel pattern, noting that students who used AI primarily for idea generation and revision suggestions, rather than full-text generation, demonstrated stronger critical thinking outcomes alongside writing improvement. Institutional and disciplinary context also moderated outcomes: Alsolami (2024), studying students with mild intellectual disabilities, found that AI-based support was effective only when accompanied by substantially more structured scaffolding than typically developing peers required, indicating that population-specific implementation guidelines are necessary rather than a single generalized model.

Risks, Tensions, and Unresolved Concerns

Despite convergent evidence of effectiveness, 14 of the 25 studies explicitly flagged risks that complicate a straightforwardly positive interpretation. Overreliance emerged as the most frequently cited concern: Syaifei and Nuraeningsih (2025) and Chen and Gong (2025) both reported that students expressed ambivalence about AI use, valuing its efficiency while worrying about diminished independent writing capability. Aljuaid's (2024) systematic review concluded that AI tools reliably improve mechanical and organizational accuracy but show inconsistent effects on higher-order argumentation and critical thinking, precisely because inconsistent instructor scaffolding across studies confounds the technology's intrinsic effect. Deep and Chen (2025) explicitly linked heavy AI reliance to reduced opportunities for productive struggle, a cognitive process considered important for durable skill acquisition in writing pedagogy.

Academic integrity concerns were the second major tension. Apriani et al. (2025) found that students' willingness to disclose AI use was strongly shaped by perceived institutional tolerance, with many students reporting uncertainty about where legitimate assistance ended and academic misconduct began. Abdullah et al. (2025) noted that as machine learning-based writing support becomes embedded within institutional learning management systems, the boundary between sanctioned scaffolding and unsanctioned substitution becomes increasingly difficult for both students and instructors to adjudicate, a concern echoed in broader systematic reviews of AI in higher education (Gligorea et al., 2023).

Comparative Effectiveness by Tool Type

Disaggregating findings by AI tool category reveals a more nuanced picture than treating "AI-based learning" as a monolithic intervention. Generative conversational tools (ChatGPT and comparable large language models), examined in fourteen of the twenty-five studies, produced the broadest range of reported benefits, spanning grammatical accuracy, idea generation, structural revision, and motivational gains (Song & Song, 2023; Mahapatra, 2024; Apriani et al., 2025). However, this breadth came with a correspondingly wider range of reported risks, particularly regarding overreliance and integrity concerns, since generative tools are capable of producing complete draft text rather than isolated corrections. By contrast, narrower automated writing evaluation tools such as Grammarly, QuillBot, and Wordtune, examined in six studies, showed more circumscribed but more consistently positive effects concentrated on mechanical accuracy and feedback literacy (Rad et al., 2023; Escalante et al., 2023), with comparatively fewer integrity concerns reported, likely because their output is corrective rather than generative. Hybrid or institutionally embedded systems, examined in five studies, showed the most variable outcomes, contingent almost entirely on implementation quality: Abdullah et al. (2025) and Khasanah (2025) both found strong effects when institutional feedback systems were paired with structured instructional scaffolding, while effects were comparatively muted in studies where such systems operated with minimal teacher involvement. This pattern suggests that tool category alone does not determine effectiveness; rather, tool category interacts with implementation fidelity, corroborating the moderating-factors analysis in Section 3.3.

A related comparative observation concerns writing subskill specificity. Across the corpus, gains in grammatical and mechanical accuracy were reported with the greatest consistency and the largest effect sizes where quantitative data were available, while gains in argumentative depth, critical thinking, and disciplinary voice were reported less consistently and, where present, were more strongly conditioned by

human mediation (Aljuaid, 2024; Khuder, 2025). This asymmetry has direct implications for how institutions should calibrate expectations: AI-based learning appears highly reliable as a tool for reducing surface-level error and accelerating revision cycles, but should not, on the basis of current evidence, be treated as a reliable independent driver of higher-order academic reasoning without deliberate pedagogical design around its use.

Implications for Practice

The convergent findings across Sections 3.2 through 3.5 carry several practical implications for writing instructors and curriculum designers. First, because feedback immediacy and modality substantially influence outcomes, institutions integrating AI-based learning should prioritize tools and workflows that preserve rapid iterative feedback while explicitly requiring students to justify or critique AI suggestions before incorporating them, operationalizing the "feedback-seeking" competence highlighted by Khuder (2025) and Rad et al. (2023). Second, given the demonstrated importance of AI literacy and prompt engineering skill as moderators (Kim et al., 2025), writing curricula should incorporate explicit instruction in effective AI interaction, rather than assuming students will develop these skills incidentally through unstructured tool exposure. Third, the consistent finding that human-AI collaboration patterns determine durability of gains (Nguyen et al., 2024; Deep & Chen, 2025) supports designing assignments that require documented drafting stages, such as an AI-assisted draft followed by a student-authored critical revision, so that AI functions as a scaffold within a visible process rather than an opaque substitute for student authorship. Finally, because academic integrity ambiguity was a recurring concern (Apriani et al., 2025; Abdullah et al., 2025), institutions should pair any AI integration with explicit, course-level disclosure norms rather than relying on generic institutional policy, since the studies reviewed here indicate that ambiguity itself, more than AI use per se, generates the greatest anxiety and inconsistency in student behavior.

Synthesis Across the Three Research Questions

Synthesizing across the corpus, this review's first research question, the extent of AI's effectiveness, receives strong but conditional support: 88% of included studies reported measurable gains in at least one writing dimension, with grammatical accuracy showing the most consistent and lexical/structural coherence showing moderate but genre-dependent gains. The second research question, moderating factors, is answered by a convergent set of three variables identified independently across multiple studies: AI literacy/prompt competence, feedback modality and framing, and human-AI collaboration pattern, each of which determined whether gains were superficial (accuracy-only) or durable (transferable to unaided writing). The third research question, an appropriate pedagogical model, is addressed by the recurring finding, present in Fathi and Rahimi (2024), Rad et al. (2023), Khuder (2025), and Nguyen et al. (2024), that AI functions most effectively as a scaffolding mechanism embedded within, rather than substituting for, human-mediated feedback cycles. This cross-study convergence, drawn from heterogeneous tools and contexts rather than a single intervention, constitutes the integrative contribution of the present review and substantiates the pedagogical model proposed in the Conclusion

4. CONCLUSION

This systematic review synthesized evidence from 25 studies published between 2023 and 2026 to examine the effectiveness of AI-based learning in improving students' academic writing skills. The evidence indicates that AI-based learning tools, including generative conversational agents, automated writing evaluation systems, and hybrid institutional platforms, can improve grammatical accuracy, lexical richness, and revision efficiency, while also increasing motivation and reducing writing anxiety across EFL, ESL, and other academic writing contexts. However, these benefits are neither automatic nor uniform, as they are influenced by students' AI literacy and prompt engineering competence, the modality and framing of feedback, and the patterns of human-AI collaboration adopted during the writing process. The effects on higher-order competencies, including argumentation, disciplinary voice, and critical thinking, remain less consistent and depend heavily on deliberate instructor mediation. Therefore, AI-based learning is most appropriately viewed as an effective scaffolding mechanism rather than an autonomous substitute for writing instruction, with AI-generated feedback functioning as a supplementary "third space" supported by instructor guidance, structured disclosure norms, and documented drafting stages.

This review is limited by its reliance on previously published studies and the heterogeneity of AI tools, learning contexts, and implementation approaches examined across the literature. Consequently, the findings may not fully capture the long-term effects of AI-assisted writing or the practical challenges associated with overreliance and academic integrity. Future research should prioritize longitudinal designs

to assess whether AI-assisted writing gains remain after tool withdrawal, as well as comparative studies that isolate the effects of specific feedback modalities across different AI tool categories. Further empirical investigation involving instructors and students would also help clarify how AI literacy, human-AI collaboration, and pedagogical mediation influence the development of sustainable and transferable academic writing competencies.

5. REFERENCES

- Abdullah, S., Aqil, M., Sanjani, F., Wahyudin, A., Putri, T. H., Dinda, F., & Thohir, M. (2025). Role of artificial intelligence-based machine learning in improving the quality of students' academic writing. 2025 5th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), 24-29. <https://doi.org/10.1109/ice3is66769.2025.11281284>
- Aladini, A. (2023). AI applications impact on improving EFL university academic writing skills and their logical thinking. *Educational Sciences Journal*. <https://doi.org/10.21608/ssj.2023.320166>
- Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal*. <https://doi.org/10.24093/awej/chatgpt.2>
- Alsolami, A. (2024). The effectiveness of using artificial intelligence in improving academic skills of school-aged students with mild intellectual disabilities in Saudi Arabia. *Research in Developmental Disabilities*, 156, Article 104884. <https://doi.org/10.1016/j.ridd.2024.104884>
- Apriani, E., Daulay, S. H., Aprilia, F., Marzuki, A. G., Warsah, I., Supardan, D., & Muthmainnah. (2025). A mixed-method study on the effectiveness of using ChatGPT in academic writing and students' perceived experiences. *Journal of Language and Education*. <https://doi.org/10.17323/jle.2025.17913>
- Chen, C., & Gong, Y. (2025). The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students. *Education Sciences*. <https://doi.org/10.3390/educsci15020141>
- Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, 15(9), Article 247. <https://doi.org/10.3390/soc15090247>
- Dja'far, V. H., & Hamidah, F. N. (2024). The effectiveness of AI technology in improving academic English writing skills in higher education. *Journal of Language and Literature Studies*, 4(3). <https://doi.org/10.36312/jolls.v4i3.2031>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: A qualitative study. *Computer Assisted Language Learning*, 39, 263-302. <https://doi.org/10.1080/09588221.2024.2374772>
- Gligorea, I., Cioca, M., Oancea, R., Gorski, A., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*. <https://doi.org/10.3390/educsci13121216>
- Kaharuddin, Ahmad, D., Mardiana, Latif, I., Arafah, B., & Suryadi, R. (2024). Defining the role of artificial intelligence in improving English writing skills among Indonesian students. *Journal of Language Teaching and Research*, 15(2). <https://doi.org/10.17507/jltr.1502.25>
- Khasanah, S. (2025). Effectiveness of AI real-time feedback assistance in improving writing skills. *Belantika Pendidikan*, 8(1), 40-48. <https://doi.org/10.47213/bp.v8i1.422>
- Khuder, B. (2025). Enhancing disciplinary voice through feedback-seeking in AI-assisted doctoral writing for publication. *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amaf022>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265-1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 58, 638-655. <https://doi.org/10.1080/15391523.2025.2456043>
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, Article 24. <https://doi.org/10.28945/5370>

- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11. <https://doi.org/10.1186/s40561-024-00295-9>
- Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49, 847-864. <https://doi.org/10.1080/03075079.2024.2323593>
- Pratama, R. M. D., & Hastuti, D. (2024). The use of artificial intelligence to improve EFL students' writing skill. *English Learning Innovation*, 5(1). <https://doi.org/10.22219/englie.v5i1.30212>
- Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32, 5020-5040. <https://doi.org/10.1080/10494820.2023.2208170>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Syafei, M., & Nuraeningsih, N. (2025). Leveraging artificial intelligence in writing: ELT students' perspectives and experiences. *International Journal of Learning, Teaching and Educational Research*. <https://doi.org/10.26803/ijlter.24.5.22>
- Utami, S., Andayani, A., Winarni, R., & Sumarwati, S. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemporary Educational Technology*. <https://doi.org/10.30935/cedtech/13419>
- Zulaikha, Z., Setyawan, C. E., Maburi, M., & Rauhilah, S. (2025). The effectiveness of artificial intelligence and deep learning tools in enhancing academic journal writing: A mixed methods study of Arabic language education students in Indonesia. *TRANSFORMASI*, 21(1). <https://doi.org/10.56357/jt.v21i1.429>