

ChatGPT and Generative AI in Writing Instruction: Literature Review

Mochammad Indra Yumanto^{1*}, Ibnu Mas'ud Luthfi²

¹UIN Sunan Ampel Surabaya, Indonesia

²STAI di Tahdzib Jombang, Indonesia

^{1*}Correspondence Author's E-mail: mindrayumanto@gmail.com

Article History	Received [09/04/2026]	Revised [10/05/2026]	Accepted [30/06/2026]
-----------------	--------------------------	-------------------------	--------------------------

Article Info

Abstract

This study is motivated by the rapid adoption of ChatGPT and other generative artificial intelligence tools in writing instruction across educational levels. The growing use of these tools raises questions about their actual pedagogical value, the reliability of the feedback they generate, and the risks they pose to academic integrity. This study aims to map the role, feedback quality, and challenges of ChatGPT use in writing instruction, and to identify implications for Arabic writing instruction. The method used is library research with a descriptive-analytical qualitative design. Data were collected from journal articles, books, and scholarly reports indexed in Scopus, Scite.ai, and Google Scholar, then analyzed through thematic content analysis. The results show that ChatGPT supports idea development, drafting, and surface-level language correction, but its contribution to higher-order feedback depends on prompt design and revision structure. AI-generated citations remain unreliable and require independent verification. User acceptance is generally positive but conditional on transparency and human oversight. Based on these results, it can be concluded that ChatGPT functions best as a complementary tool guided by teachers, not as a replacement for human judgment, and that Arabic writing instruction needs its own empirical evidence base for AI-assisted feedback design.

Keywords

ChatGPT;
generative artificial intelligence;
writing instruction;
AI feedback;
Arabic writing

Abstrak

Penelitian ini dilatarbelakangi oleh pesatnya adopsi ChatGPT dan alat kecerdasan buatan generatif lain dalam pembelajaran menulis di berbagai jenjang pendidikan. Meluasnya penggunaan alat ini memunculkan pertanyaan tentang nilai pedagogisnya, keandalan umpan balik yang dihasilkan, dan risikonya terhadap integritas akademik. Penelitian ini bertujuan memetakan peran, kualitas umpan balik, dan tantangan penggunaan ChatGPT dalam pembelajaran menulis, serta mengidentifikasi implikasinya bagi pembelajaran menulis bahasa Arab. Metode yang digunakan adalah studi kepustakaan dengan desain kualitatif deskriptif-analitis. Data dikumpulkan dari artikel jurnal, buku, dan laporan ilmiah yang terindeks di Scopus, Scite.ai, dan Google Scholar, kemudian dianalisis melalui analisis isi tematik. Hasil penelitian menunjukkan bahwa ChatGPT mendukung pengembangan ide,

Kata Kunci

ChatGPT;
kecerdasan buatan generatif;
pembelajaran menulis;
umpan balik AI;
menulis bahasa Arab

<i>penyusunan draf, dan koreksi bahasa tingkat permukaan, tetapi kontribusinya pada umpan balik tingkat tinggi bergantung pada desain prompt dan struktur revisi. Sitasi hasil ChatGPT masih belum andal dan memerlukan verifikasi independen. Penerimaan pengguna secara umum positif namun bersyarat pada transparansi dan pengawasan manusia. Berdasarkan hasil tersebut, dapat disimpulkan bahwa ChatGPT berfungsi paling baik sebagai alat pelengkap yang dipandu guru, bukan pengganti penilaian manusia, dan bahwa pembelajaran menulis bahasa Arab memerlukan basis bukti empiris tersendiri untuk desain umpan balik berbantuan AI.</i>	
To cite this article: <i>Muhammad & Ibnu</i> . (2026). "ChatGPT and Generative AI in Writing Instruction: A Systematic Literature Review" JETLA: Journal of Educational Technology and Learning Arabic, Volume 1, Issue 1, 2026; Page: 40-54	
This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/deed.en) © 2026 by author/s	

INTRODUCTION

Background of the Problem

Generative artificial intelligence, especially ChatGPT, has developed rapidly. It now appears widely in writing instruction across many fields of education ([Mahapatra, 2024](#)). Its use covers idea development, drafting, language editing, information retrieval, and feedback on writing ([Chan & Hu, 2023](#)). [Wang \(2024\)](#) shows that ChatGPT use is relevant for foreign language learning. The model helps learners overcome linguistic difficulties and develop the content of their writing. Large language models no longer function only as grammar checkers. They now act as companions across the stages of the writing process ([Imran & Almusharraf, 2023](#)). Generative AI can generate ideas, develop arguments, structure texts, and support revision ([Wang, 2024](#)). ChatGPT can also provide fast feedback tailored to individual needs. Its effectiveness, however, still depends on the quality of interaction and the user's ability to evaluate AI output ([Escalante et al., 2023](#)). These changes push educators to revisit writing instruction strategies, set clear usage guidelines, and preserve human oversight in learning and assessment ([Barrett & Pack, 2023](#)).

[Mahapatra \(2024\)](#) finds that ChatGPT can raise second-language learners' writing motivation while it automates routine instructional tasks. The model gives instant, personalized feedback. Structured prompt engineering makes this potential clearer. [Knoth et al. \(2024\)](#) describe how learners are trained to formulate questions and request explanations from AI. [Ma et al. \(2023\)](#) add that iterating on AI output can improve argumentation and content quality. [Marzuki et al. \(2023\)](#) find that explicit training in prompt composition raises students' self-efficacy and AI-related knowledge, which then supports writing quality.

Field-specific studies confirm this pattern. In biochemistry education, [Oates and Johnson \(2025\)](#) find that iterative essay tasks with ChatGPT feedback help students understand how generative AI works and how to write prompts. The same study, however, flags

serious problems with the accuracy of AI-generated citations. [Tlili et al. \(2023\)](#) report a similar issue in chemistry-related information literacy instruction.

Several studies provide direct evidence on ChatGPT's effectiveness in writing instruction and assessment. [Song and Song \(2023\)](#) find that ChatGPT use in English-as-a-foreign-language academic writing instruction improves writing skills and task completion. Learners still worry about long-term effects on their own writing ability and call for explicit instruction on appropriate AI use. [Kim et al. \(2024\)](#) show that students see ChatGPT-assisted writing systems as a writing assistant, a virtual tutor, and a learning peer, though they also face challenges from the technology, the learners, and the task itself. [Geçkin et al. \(2023\)](#) compare ChatGPT and human raters in second-language academic writing assessment. Agreement between the two falls in the low-to-fair range, which suggests combining AI and human assessment produces more reliable scores. [Yavuz et al. \(2024\)](#) test ChatGPT and Google Bard on essay grading using an analytic rubric. ChatGPT shows high reliability and some alignment with instructor grading, but still needs refinement for interpretive, non-objective aspects. [Franjić and Božinovski \(2025\)](#) find that ChatGPT use helps lower-proficiency ESL learners improve grammar, vocabulary, organization, and coherence, though it may increase writing time.

Together, these studies show that ChatGPT can support writing quality, assist multiple stages of the writing process, and inform rubric-based assessment. AI reliability, however, stays inconsistent across writing aspects. ChatGPT use therefore needs proper instruction, critical evaluation training, and human verification. Its effectiveness in writing instruction depends on instructional design and user engagement, not on technological capability alone.

AI feedback in writing instruction shows a mixed pattern. AI gives fast feedback and improves surface-level language, but it still struggles with higher-order discourse features such as argument organization, content coherence, and reasoning accuracy. When paired with a structured revision cycle, AI feedback can help learners review, improve, and progressively develop their writing. ChatGPT should therefore stay a companion in the writing process, not a substitute for critical thinking and human judgment.

Identification of Problems

Three problems emerge from this pattern. First, the actual pedagogical value of ChatGPT in writing instruction still varies across studies, research designs, and educational levels, which makes generalization difficult. Second, AI-generated feedback and citations show reliability gaps, particularly at the higher-order discourse level and in reference accuracy, which raises academic integrity risks. Third, most existing evidence comes from English-language and general higher-education contexts. Arabic writing instruction, including mahārah kitābah in pesantren and Islamic higher education settings, still lacks a dedicated empirical base for AI-assisted feedback. These problems show a gap between the growing use of ChatGPT in practice (das sein) and the need for verified, contextually grounded guidance on its use (das sollen), particularly for Arabic writing instruction.

Formulation of the Problem

Based on this background, the research problems are formulated as follows: first, what role does ChatGPT play in the writing instruction process according to existing literature; second, how effective and reliable is ChatGPT-generated feedback in supporting surface-level and higher-order writing skills; and third, what challenges and literature gaps remain regarding ChatGPT use in writing instruction, particularly for Arabic writing instruction.

Objectives and Benefits of Research

Research Objectives

This study aims to map the role of ChatGPT in the writing instruction process, examine the effectiveness and reliability of ChatGPT-generated feedback, and identify challenges and literature gaps relevant to Arabic writing instruction.

Research Benefits

Theoretically, this study enriches the literature on generative AI in writing pedagogy, particularly regarding feedback effectiveness and citation reliability. Practically, the findings can guide teachers and institutions, including Arabic language education programs, in designing AI use policies that protect academic integrity while supporting learning autonomy. For Arabic language education programs and pesantren-based education, the findings offer a basis for developing an AI-assisted feedback model suited to the linguistic and rhetorical demands of Arabic writing.

THEORETICAL AND CONCEPTUAL FRAMEWORK

Theoretical Framework

This study builds on scaffolding theory and writing feedback theory. Scaffolding theory frames AI as temporary support that a learner needs less over time as competence grows. Applied to ChatGPT, this means AI support should concentrate on early writing stages, such as idea generation and surface-level correction, then gradually give way to independent student judgment ([Teng, 2024](#); [Teng & Huang, 2025](#)).

Writing feedback theory separates surface-level feedback from higher-order feedback. Surface-level feedback covers grammar, spelling, and diction. Higher-order feedback covers argument organization, content coherence, and reasoning quality ([Shi & Aryadoust, 2024](#)). Evidence shows that AI handles surface-level feedback well but needs structured revision cycles and careful prompt design to reach higher-order feedback quality ([Walter, 2024](#)).

Citation reliability theory also applies here. Academic writing depends on traceable, verifiable sources. [Walters and Wilder \(2023\)](#) show that AI-generated citations can look complete and convincing while remaining fictitious. This risk means AI output needs treatment as unverified information, not as validated academic material, until an independent check confirms it.

Together, these three theoretical strands, scaffolding, feedback level, and citation reliability, frame ChatGPT as a conditional support tool. Its pedagogical value depends on how instructors design tasks, how much students engage critically, and how far human oversight remains active in the process.

Conceptual Framework

Based on this theoretical foundation, the conceptual framework of this study positions ChatGPT as an input variable that interacts with three mediating factors: instructional design, prompt quality, and user AI literacy. These three factors shape two outcome dimensions, namely writing quality (idea development, drafting, revision) and feedback reliability (surface-level accuracy and higher-order validity). User perception and acceptance act as a moderating factor that shapes how far students and teachers trust and adopt ChatGPT output. Human oversight remains a constant background condition across all these relationships, since none of the reviewed evidence supports removing teacher judgment from the writing assessment process. Applied to Arabic writing instruction, this framework adds a fourth mediating factor, namely linguistic and rhetorical conformity, which covers *nahwu* (syntax), *sharaf* (morphology), and Arabic discourse conventions that AI models trained mainly on English corpora may not capture accurately.

METHODOLOGY

This study uses library research with a descriptive-analytical method (Creswell, 2003). The research is qualitative and draws on written sources, namely journal articles, books, policy reports, and scholarly reviews relevant to ChatGPT and generative AI in writing instruction. Primary sources cover articles that discuss ChatGPT, large language models, or generative AI in writing instruction. Secondary sources strengthen the theoretical foundation on digital literacy, writing pedagogy, and AI ethics.

Data collection used documentation techniques through Scopus, Scite.ai, and Google Scholar. Search keywords included ChatGPT, writing instruction, academic writing, and AI feedback.

Data analysis used descriptive-qualitative content analysis in three stages: data reduction, thematic data presentation, and conclusion drawing. The themes analyzed cover the role of AI in the writing process, types of feedback provided, citation accuracy, and user perceptions and challenges.

Data validity relied on source triangulation, which compared findings across multiple studies, particularly literature from reputable journals published between 2021 and 2026. This study did not collect primary data. Its results synthesize prior research and cannot be fully generalized, particularly to Arabic language learning contexts, which still need further field research.

RESULT AND DISCUSSION ANALYSIS

Research Data Results

The reviewed literature shows that ChatGPT plays a strong role in the early stages of writing, particularly idea development, argument planning, and initial drafting. [Levine et al. \(2024\)](#) find that students mainly use ChatGPT to generate ideas during planning and argument development. Students do not simply copy AI output. They develop, compare, and adapt these ideas to fit their own rhetorical goals and writing voice. This finding

confirms that ChatGPT functions as a resource within a social, distributed writing process, not as a machine that produces finished text on its own.

The reviewed studies use varied research designs, including quasi-experiments, large-scale surveys, classroom case studies, and systematic reviews. [Ganjavi et al. \(2024\)](#) survey around 415 medical students to map their awareness, use, and concern regarding ChatGPT. [Reddy et al. \(2024\)](#) use a case study on a biochemistry writing task that involves iterative essay evaluation, AI feedback, and independent literature research.

This design diversity shows that ChatGPT's effectiveness cannot rest on final writing products or usage frequency alone. Its pedagogical impact depends on task integration, teacher scaffolding quality, students' ability to evaluate AI output, and the room available for gradual revision. Current evidence therefore supports directed, reflective ChatGPT use. It does not justify handing the entire writing process to AI. The topic remains young, so the evidence does not yet allow large-scale quantitative meta-analysis. Cross-study synthesis instead identifies general patterns and the conditions that shape successful ChatGPT use in writing instruction.

The literature search covers multiple educational levels and disciplines: second-language education, medicine, chemistry, biochemistry, and information literacy. This spread shows that ChatGPT use in writing instruction is a cross-disciplinary phenomenon, not a practice limited to one field.

Analysis and Discussion of Research Results

The Role of ChatGPT in the Writing Process

ChatGPT works best as a facilitator in the early stages of writing, especially idea development, language exploration, and initial drafting. At this stage, AI lowers students' affective barriers and encourages bolder experimentation with language before refinement ([Özçelik & Ekşi, 2024](#)). [Mahapatra \(2024\)](#) adds that in foreign-language writing, students often use ChatGPT to answer grammar questions, request correction suggestions, and detect surface errors. Sample gaps across age groups and education levels, however, still limit how far these findings generalize.

[Reddy et al. \(2024\)](#) report that students in biochemistry writing tasks do more than produce essays with ChatGPT's help. They evaluate the weaknesses of AI output, make iterative refinements, and combine this process with independent literature research. This process helps students understand the technology's functions, prompt-composition strategies, and AI's limits in generating scientific content and references. ChatGPT's main benefit, then, lies less in fast draft production and more in its capacity to trigger dialogue, critical evaluation, and decision-making during writing.

Directed ChatGPT use can also strengthen student engagement. [Teng and Huang \(2025\)](#) find that carefully designed prompts raise affective and behavioral engagement in English-as-a-foreign-language writing instruction, though cognitive engagement does not show a significant increase. This finding shows that the technology can raise interest, participation, and willingness to work on writing tasks, but higher engagement does not automatically raise higher-order thinking. ChatGPT use therefore still needs prompting

questions, evaluation instructions, and revision tasks that demand reasoning and writer accountability.

[Teng \(2024\)](#) finds that ChatGPT use associates with higher motivation, self-efficacy, engagement, and a tendency to collaborate in writing, although some students still raise concerns about its use. This finding supports the argument that ChatGPT can lower initial barriers to writing, especially when students struggle to start a text or form ideas. This support, however, must help students build their own writing capacity. It should not replace the work of planning, developing, and revising texts.

If ChatGPT use spreads evenly across every stage of writing, students risk forming wrong expectations about its capacity and growing dependent on its output. That dependence can reduce practice in forming ideas, reasoning about claims, and developing a personal writing voice. ChatGPT works better as a scaffold that gradually recedes as student competence grows. In early stages, AI can open up ideas and flag linguistic issues. In later stages, responsibility for evaluation, revision, and accountability should shift back to students. This approach aligns AI use with the core goal of writing instruction, namely forming independent, critical, decision-capable writers, not just producing fluent-looking texts.

Feedback Effectiveness: Surface-Level versus Higher-Order

[Shi and Aryadoust's \(2024\)](#) systematic review shows that AI-based feedback research produces positive, negative, neutral, and mixed results depending on task design and integration into the learning cycle. AI performs consistently well at surface-level correction: grammar, spelling, and diction. Its contribution to higher-order thinking, such as argument organization and reasoning quality, depends heavily on how instructors integrate it into a structured revision cycle, not on the model's sophistication alone.

[Walter \(2024\)](#) confirms that feedback quality depends greatly on careful prompt engineering and alignment with a discipline's writing conventions. Without well-developed prompt design, AI use tends to stay at the level of mechanical correction. [Teng and Huang \(2025\)](#) support this point: carefully designed prompts raise affective and behavioral engagement in EFL writing, though the effect on cognitive engagement stays limited. When integrated into an explicit revision cycle, AI's potential to support argument analysis and evaluation becomes visible, as shown in [Reddy et al.'s \(2024\)](#) biochemistry writing model, where students create, evaluate, and refine ChatGPT output through feedback and literature searches. AI, then, should work as a dialogic partner and revision aid, not a replacement for a writer's own judgment and decisions ([Teng, 2024](#)).

Feedback effectiveness still varies by genre, task purpose, and student competence level ([Özçelik & Ekşi, 2024](#)). [Mahapatra \(2024\)](#) shows ChatGPT can work as a formative feedback tool with adequate training, and [Teng \(2024\)](#) links ChatGPT use to higher motivation, self-efficacy, and collaborative writing tendencies. AI's benefits, then, do not always show up only in writing scores. They also appear in affective and metacognitive dimensions. Generalizing AI's benefits across genres and education levels therefore still needs caution.

Citation Accuracy and Reliability

One of the most consistent findings across the reviewed sources is the low reliability of AI-generated citations and references. [Reddy et al. \(2024\)](#) find that about 26 percent of the supporting references used in a chemistry writing task do not exist. AI output, then, cannot count as validated academic material, even when it looks convincing and carries seemingly complete bibliographic details.

This risk can damage writing credibility, block argument traceability, and cause academic integrity violations when students cite unverified references. [Walters and Wilder \(2023\)](#) confirm that AI-generated citations can appear complete while remaining fabricated. Instruction in citation verification, primary source retrieval, and AI literacy therefore needs a place as a core writing curriculum component, not just as technical tool training. Within this framework, ChatGPT can help formulate keywords or map possible sources, but the decision to accept and cite a reference must rest on the author's own verification and educator guidance.

For academic writing instruction, including Arabic-language scholarly writing central to Arabic language education programs, this risk calls for explicit strengthening of source verification skills as part of the curriculum. Every reference an AI model generates or suggests needs treatment as preliminary information, subject to independent verification before use in academic assignments or publication.

User Perception, Trust, and Acceptance

Student perceptions of ChatGPT in writing instruction lean positive, but this acceptance does not amount to unconditional trust. [Ganjavi et al. \(2024\)](#) find high awareness and use among medical students, yet respondents still raise concerns about inaccuracy, privacy, and plagiarism risk. [Stöhr et al. \(2024\)](#) find a similar pattern among general higher education students: exposure raises familiarity and trust, but almost half of respondents still express concern about future AI use, and most still expect human involvement in final assessment. Increased use, then, does not show that students accept AI as a teacher replacement. It reflects conditional acceptance that depends on transparency, accuracy, and active oversight.

[Khowaja et al. \(2024\)](#) find similar ethical concerns, namely privacy, bias, transparency, intellectual property, and misinformation, alongside the access divide as a pressing issue in AI use. Students' positive perceptions, then, need a critical reading. ChatGPT looks useful because it works fast, stays easily accessible, and gives personal support during writing, but these benefits only hold when balanced against the ability to evaluate AI output. The most constructive stance avoids absolute acceptance or rejection. It builds calibrated trust: students can use AI to generate ideas and get initial feedback, but they still need to check information accuracy, judge suggestion appropriateness, protect personal data, and take responsibility for their writing decisions. Human involvement in learning and assessment, in this framework, works as a pedagogical requirement for autonomy, fairness, and academic integrity, not just an administrative safeguard.

[Leslie et al. \(2021\)](#) add that responsible AI governance needs transparency and accountability frameworks, which reinforces the case for keeping human oversight active in assessment. [Sullivan et al. \(2023\)](#) raise a related concern, namely that over-reliance on AI can reduce students' opportunities to build independent judgment about their own

writing. AI literacy, then, needs a focus not only on technology use skill, but also on output evaluation, model-limitation recognition, and responsibility for decisions based on AI suggestions ([Walter, 2024](#)).

AI Literacy and the Layered Feedback Model

[Walter \(2024\)](#) shows that explicit prompt-composition training raises students' self-efficacy and AI-related knowledge. This finding matters because it confirms that AI output quality depends not only on model sophistication, but also on users' ability to formulate instructions, evaluate responses, and use them for conscious writing revision. Prompt engineering, then, should count as part of academic literacy and critical thinking strategy, not just an extra technical skill.

Teacher AI literacy matters as much as student AI literacy. [Chan and Hu \(2023\)](#) find that teachers need the ability to assess AI output quality, recognize potential bias, explain technology limitations, and design tasks that use AI without shifting the thinking responsibility onto machines. Without these competencies, AI risks producing convincing-looking texts without building students' understanding, revision skill, or reasoning development ([Chan, 2023](#)). AI integration success in writing instruction, then, depends more on pedagogical design and educators' capacity to mediate AI use than on access to AI tools alone.

[Tlili et al. \(2023\)](#) highlight ethical and pedagogical issues in TESOL instruction, namely authorship, accuracy, bias, privacy, security, and accountability. These issues call for AI use guided by transparency, responsibility, and process traceability. Students still need to explain their own contribution, separate personal ideas from AI suggestions, and verify information before use, given that ChatGPT output can carry inaccurate or unverifiable citations ([Reddy et al., 2024](#)). AI literacy development, then, needs to cover output evaluation, model-limitation recognition, and ethical impact consideration ([Chan & Hu, 2023](#); [Walter, 2024](#)). AI evaluation in education cannot rest on text-generation effectiveness alone. It must also weigh whether the use stays fair, safe, accountable, and aligned with learning goals ([Chan, 2023](#)).

Overly permissive policy risks leaving problems such as information inaccuracy, citation fabrication, plagiarism, privacy violation, and AI dependence unaddressed. Citation fabrication carries particular weight because AI-generated references can look convincing while staying unverifiable ([Walters & Wilder, 2023](#)). A total ban, however, works no better: it blocks the AI literacy students increasingly need and limits their chance to learn technology use critically, ethically, and responsibly ([Chan & Hu, 2023](#)). The real policy question, then, is not whether to accept or reject AI. It is which pedagogical and ethical conditions make its use accountable.

[Sullivan et al. \(2023\)](#) show that higher education institutions have begun to set ChatGPT use regulations, though these differ significantly across institutions and countries. This diversity shows that AI policy cannot follow one uniform approach, since effectiveness depends on learning goals, task characteristics, users' AI literacy level, and institutional capacity for training and oversight. A comprehensive policy framework needs to cover pedagogical, governance, and operational dimensions together: it should regulate AI use

purposes in learning, protect privacy and accountability, and ensure adequate infrastructure, training, and evaluation mechanisms ([Chan, 2023](#)).

Institutional policy, then, should avoid both blanket prohibition and blanket permission. Policies need a graded approach based on task stage: institutions can allow limited AI use for brainstorming, outlining, and initial feedback, but must restrict its use for generating full answers or replacing student thinking ([Barrett & Pack, 2023](#)). Institutions also need to require AI use disclosure, set information and source verification procedures, provide prompt-composition guidance, and assess the writing process trace, including drafts, revisions, reflections, and students' ability to explain their own linguistic decisions, rather than assess the final product alone. This approach keeps AI as a learning aid, not a substitute for writing competence.

For Arabic language education programs, this policy framework needs alignment with the demands of mahārah kitābah (writing skill) and Arabic academic conventions. Regulation should go beyond stating whether ChatGPT use is permitted. It should specify which linguistic aspects require verification by students and educators, such as nahwu (syntax) accuracy, sharaf (morphology), diction choices, cohesion, argument structure, and conformity with Arabic rhetorical conventions and academic genres. Teachers need to keep authority over meaning accuracy, context, and discourse quality validation, because grammatically correct AI output does not always align with communicative purpose or accepted academic norms. AI policy in Arabic language education, then, can aim at raising efficiency and providing revision support while it protects academic integrity, independent thinking, and linguistic-pedagogical relevance.

On this basis, AI-based writing instruction should follow a layered feedback model. In the first layer, AI handles surface corrections, such as grammar, spelling, word choice, and possible sentence improvements, while it raises prompting questions that encourage initial revision. In the second layer, students select, verify, and explain their reasons for accepting or rejecting AI suggestions. In the third layer, teachers keep control over aspects that need pedagogical and scholarly judgment, especially argument organization, content depth, discourse coherence, idea originality, and conformity with disciplinary conventions. This division of function keeps AI from replacing the thinking process while it ensures every AI use produces reflective activity, not passive acceptance of machine output.

This model also requires a shift in assessment practice. Teachers cannot assess the final product alone. They need to assess the process, namely drafts, revision history, responses to AI feedback, student reflection, and students' ability to take responsibility for linguistic decisions. AI use, in this way, can strengthen learning autonomy and revision skill rather than help students avoid the writing process. This layered feedback model aligns with [Chan's \(2023\)](#) recommendation to develop assessment policies and syllabi that explicitly regulate AI use boundaries in writing tasks while leaving room for its controlled integration.

Implications for Arabic Writing Instruction

This review points to a clear gap in the existing literature. Automated written feedback research has grown and now covers a wide range of systems, but its focus still centers on

English language learning at the tertiary level ([Shi & Aryadoust, 2024](#)). [Almelhes \(2023\)](#) shows that AI use for Arabic language learning can support self-directed learning environments, help learners identify pronunciation errors, and support non-native learners, but this study does not examine writing instruction or mahārah kitābah specifically. Research on AI in Arabic writing instruction, then, stays limited in both quantity and focus.

This limitation carries weight because AI needs assessment not only for its ability to correct language errors, but also for its conformity with Arabic rhetorical conventions, discourse structures, and communicative purposes. The layered feedback model applies well to Arabic academic writing instruction: AI can give initial support on surface aspects, while students and teachers verify meaning accuracy, argument organization, coherence, and conformity with academic genres. This caution matters because AI output can carry errors, and AI-generated citations risk inaccuracy or unverifiability ([Walters & Wilder, 2023](#)). The teacher's role in validating AI output, then, stays essential, particularly in the Arabic context, which carries linguistic and rhetorical features distinct from English.

This gap opens room for original research, particularly on how surface-level and higher-order feedback patterns work in Arabic writing instruction specifically. Future research also needs to test students' ability to verify AI suggestions, assess citation and reference accuracy, and compare AI use effectiveness across pesantren and Islamic higher education settings. Such research would extend AI application from the English learning context to Arabic, and could also produce pedagogical models suited to the needs of mahārah kitābah and Arabic academic demands.

Limitations

Most evidence in this review comes from general higher education and medical education contexts, not from Arabic language learning specifically. The systematic evidence, even without addressing the Arabic context directly, shows that differences in writing systems and discourse structures across languages can shape how AI gives relevant feedback. AI reliability in assessing writing also depends heavily on training data, and most popular AI models still rely on corpora dominated by English. Generalizing these findings to the Arabic language education or pesantren context, then, needs care, given the differences in learner characteristics, instructional goals, and cross-linguistic rhetorical conventions that the existing literature has not tested extensively.

Contribution

This review consolidates fragmented findings on ChatGPT's role, feedback quality, and reliability risk in writing instruction into a single layered feedback framework. It links scaffolding theory, feedback-level theory, and citation reliability concerns into one coherent model that instructors can apply directly to task design and assessment practice. By pointing to the underexamined intersection between generative AI and Arabic writing instruction, particularly mahārah kitābah in pesantren and Islamic higher education settings, this review also opens a concrete direction for future empirical work in a context that current literature has not yet addressed in depth.

CONCLUSION AND SUGGESTIONS

Conclusion

ChatGPT and other large language models can support writing instruction, particularly in idea development, initial drafting, and grammar, spelling, and word choice improvement. These benefits depend on instructional design, guidance quality, and active engagement from both students and teachers. AI cannot yet fully replace human judgment in assessing argument organization, content depth, coherence, originality, and conformity with academic conventions. AI output also needs checking, since it can carry errors, bias, or unverifiable information. AI literacy, then, needs a place in academic writing instruction so students can check facts, compare information against credible sources, and take responsibility for how they use AI suggestions.

AI use in Arabic writing instruction, particularly mahārah kitābah, works best through collaboration among AI, students, and teachers. AI can give initial feedback on linguistic aspects. Students select and verify these suggestions. Teachers assess more complex aspects, namely meaning, argument organization, coherence, and cultural-academic appropriateness. With directed, transparent, and responsible use, AI can expand practice opportunities and support learning autonomy without reducing students' intellectual responsibility or the teacher's pedagogical role.

Suggestions

For teachers, this review suggests adopting the layered feedback model: let AI handle surface-level correction, require students to verify and justify their acceptance or rejection of AI suggestions, and keep teacher authority over higher-order judgment.

For Arabic language education programs and pesantren-based institutions, this review suggests developing explicit AI use policies for mahārah kitābah, covering source verification training and criteria for evaluating nahwu, sharaf, and rhetorical conformity in AI-assisted writing tasks.

For future researchers, this review suggests conducting empirical studies on AI-assisted Arabic writing instruction, testing students' ability to verify AI-generated citations, and comparing AI use effectiveness across pesantren and Islamic higher education settings to build the field-specific evidence base that current literature still lacks.

AUTHOR CONTRIBUTIONS STATEMENT

Mochammad Indra Yumanto conceived the research idea, formulated the research problem, conducted the literature search and data analysis, and drafted and revised the manuscript. Additional author names, if any, should be listed here along with their specific contributions. All authors approved the final manuscript for publication.

REFERENCES

Almelhes, S. (2023). A Review of Artificial Intelligence Adoption in Second-Language Learning. *Theory and Practice in Language Studies*, 13(5), 1259-1269. <https://doi.org/10.17507/tpls.1305.21>

- Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00427-0>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00408-3>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00411-8>
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed., pp. 34-35). Sage Publications.
- 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(1). <https://doi.org/10.1186/s41239-023-00425-2>
- Franjić, M., & Božinovski, B. (2025). Exploring ChatGPT's Impact On ESL Writing Proficiency: An Experimental Approach. *Teaching English With Technology*, 2025(2). <https://doi.org/10.56297/vaca6841/zddd4208/luzg7476>
- Ganjavi, C., Eppler, M., O'Brien, D., Ramacciotti, L. S., Ghauri, M. S., Anderson, I., Choi, J. H., Dwyer, D., Stephens, C., Shi, V., Ebert, M., Derby, M., Yazdi, B., & Cacciamani, G. (2024). ChatGPT and large language models (LLMs) awareness and use. A prospective cross-sectional survey of U.S. medical students. *PubMed Central*, 3(9). <https://doi.org/10.1371/journal.pdig.0000596>
- Geçkin, V., Kızıltaş, E., & Çınar, Ç. (2023). Assessing second-language academic writing: AI vs. Human raters. *Journal of Educational Technology and Online Learning*, 6(4), 1096-1108. <https://doi.org/10.31681/jetol.1336599>
- 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). <https://doi.org/10.30935/cedtech/13605>
- Khowaja, S. A., Khuwaja, P., Dev, K., Wang, W., & Nkenyereye, L. (2024). ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. *Cognitive Computation*, 16(5), 2528-2550. <https://doi.org/10.1007/s12559-024-10285-1>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). 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>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>

- Leslie, D., Burr, C., Aitken, M., Cows, J., Katell, M., & Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3817999>
- Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2024). How do students use ChatGPT as a writing support? Journal of Adolescent & Adult Literacy, 68(5), 445-457. <https://doi.org/10.1002/jaal.1373>
- Ma, Y., Liu, J., Fan, Y., Qikai, C., Yong, H., Wei, L., & Xiaozhong, L. (2023). AI vs. Human - Differentiation Analysis of Scientific Content Generation. arXiv. <https://doi.org/10.48550/arxiv.2301.10416>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. Smart Learning Environments, 11(1). <https://doi.org/10.1186/s40561-024-00295-9>
- Marzuki, M., Widiati, U., Rusdin, D., Darwin, D., & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. Cogent Education, 10(2). <https://doi.org/10.1080/2331186x.2023.2236469>
- Oates, A., & Johnson, D. M. (2025). ChatGPT in the Classroom: Evaluating its Role in Fostering Critical Evaluation Skills. International Journal of Artificial Intelligence in Education, 35(4), 1793-1824. <https://doi.org/10.1007/s40593-024-00452-8>
- Özçelik, N. P., & Ekşi, G. Y. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant, a case study. Smart Learning Environments, 11(1). <https://doi.org/10.1186/s40561-024-00296-8>
- Reddy, M. R., Walter, N. G., & Sevryugina, Y. V. (2024). Implementation and Evaluation of a ChatGPT-Assisted Special Topics Writing Assignment in Biochemistry. Journal of Chemical Education, 101(7), 2740-2748. <https://doi.org/10.1021/acs.jchemed.4c00226>
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187-209. <https://doi.org/10.1017/s0958344023000265>
- 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, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education Artificial Intelligence, 7, 100259. <https://doi.org/10.1016/j.caeai.2024.100259>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. Journal of Applied Learning & Teaching, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.17>
- Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education Artificial Intelligence, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>

Teng, M. F., & Huang, J. (2025). Incorporating ChatGPT for EFL Writing and Its Effects on Writing Engagement. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), 1-21. <https://doi.org/10.4018/ijcallt.367874>

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00237-x>

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00448-3>

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13(1), 14045. <https://doi.org/10.1038/s41598-023-41032-5>

Wang, C. (2024). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology Knowledge and Learning*, 30(3), 1825-1846. <https://doi.org/10.1007/s10758-024-09744-3>

Yavuz, F., Çelik, Ö., & Çelik, G. (2024). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, 56(1), 150-166. <https://doi.org/10.1111/bjet.13494>

Yılmaz, R., & Yılmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education Artificial Intelligence*, 4, 100147. <https://doi.org/10.1016/j.caeai.2023.100147>