

BEYOND ERROR CORRECTION: CHATGPT AS FORMATIVE FEEDBACK FOR ARABIC WRITING

Abdul Hafidz bin Zaid^{1*}, Muassomah², Moh. Ismail¹, Alif Cahya Setiyadi¹, Agus Yasin¹

¹Universitas Darussalam Gontor, Indonesia

²Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

*Corresponding Email: abuafadh@unida.gontor.ac.id

Received: July 2026. Accepted: August 2026. Available Online: October 2026.

ABSTRACT

The increasing adoption of large language models (LLMs) is transforming language education; however, evidence regarding their pedagogical value for Arabic writing remains limited. This study investigates the educational potential of ChatGPT as an AI-assisted formative feedback tool for Arabic academic writing in a pesantren-based university in Indonesia. To achieve this, employing a qualitative descriptive design, the study integrates Error Analysis with the Student Feedback Literacy framework to examine students' writing errors and evaluate the educational quality of AI-generated feedback. A corpus of 30 Arabic writing texts produced by 30 first-year students enrolled in an intensive Arabic matriculation program was analyzed. ChatGPT (GPT-4, May 2024 version) generated corrective feedback using a standardized prompt, and two Arabic linguistics experts evaluated the linguistic accuracy and pedagogical usefulness of each response. The findings identified 132 errors comprising 39.4% lexical, 32.6% syntactic, and 28.0% morphological errors. Notably, ChatGPT produced fully acceptable feedback for 50.76% of the errors, demonstrating its capacity in lexical and syntactic correction. However, its performance was less consistent in addressing morphologically complex and context-dependent constructions. Overall, the findings indicate that ChatGPT supports learners' ability to recognize and revise writing errors, although it still requires teacher validation. Consequently, this study contributes valuable insights into AI-assisted language education through meaningful engagement with formative feedback in Arabic writing.

Keywords: Arabic Writing, ChatGPT, Error Analysis, Formative Feedback, Student Feedback Literacy

INTRODUCTION

The rapid advancement of generative artificial intelligence (AI) is reshaping educational practice worldwide, particularly through its capacity to deliver immediate, personalized, and scalable feedback (Zhang et al., 2026). Among these innovations, large language models (LLMs) such as ChatGPT have attracted considerable attention for their ability to support second- and foreign-language learning by assisting learners in drafting, revising, and refining written texts (Al-Abri, 2025). As educational systems increasingly integrate AI into teaching and assessment, scholarly discussion has shifted from evaluating the technical performance of AI toward examining its pedagogical value in fostering meaningful learning. This shift is especially relevant in Arabic language education, where writing proficiency requires mastery of complex lexical, syntactic, and morphological systems, and where accurate written communication is essential for both academic engagement and access to Islamic intellectual traditions. Despite these developments, AI-supported Arabic language instruction remains considerably underexplored compared with English (Susiwati et al., 2026) and other widely taught languages, particularly within Islamic higher education contexts.

Recent studies demonstrate that ChatGPT can provide rapid corrective feedback, facilitate learner autonomy, and reduce teachers' feedback workload in language classrooms (Guo, 2025). Nevertheless, existing research has predominantly examined AI-assisted language

learning from perspectives such as writing improvement, pronunciation practice, translation support, or learner perceptions of AI technologies. Comparatively little attention has been devoted to understanding whether AI-generated feedback constitutes *educationally effective feedback* that learners can interpret and utilize to improve subsequent writing (Boudouaia et al., 2026). This limitation is particularly evident in Arabic language education, where linguistic accuracy depends not only on lexical choice but also on intricate morphosyntactic relationships that challenge both learners and AI systems. Moreover, empirical evidence remains scarce in *pesantren*-based universities, where Arabic serves simultaneously as the language of Islamic scholarship and academic instruction, creating distinctive pedagogical demands that differ from those of conventional foreign language classrooms.

To address this gap, this study adopts the Student Feedback Literacy framework proposed by Carless and Boud (2018). Feedback literacy conceptualizes effective feedback as a multifaceted process that extends beyond simply providing corrections. It underscores the importance of learners' ability to recognize and value feedback, interpret its underlying meaning, critically assess their work, and actively implement suggestions to enhance their future performance. From this comprehensive perspective, the educational significance of ChatGPT does not rest solely on its capacity to identify linguistic errors. Instead, its true potential lies in its ability to generate feedback that learners can thoroughly understand, reflect upon, and meaningfully apply within the writing process. As a result, by integrating feedback literacy with Error Analysis, this study advances beyond merely evaluating the accuracy of AI-generated corrections. It aims to assess their pedagogical effectiveness and their role in supporting Arabic writing development.

This theoretical perspective is particularly relevant to *pesantren*-based higher education in Indonesia. Although students are immersed in Islamic learning environments and frequently encounter Arabic through religious texts, opportunities for authentic written communication and individualized feedback remain limited. Instruction often prioritizes grammatical knowledge and textual interpretation, yet shortages of native Arabic-speaking instructors constrain the availability of timely and consistent formative feedback. Consequently, students often face persistent lexical, syntactic, and morphological difficulties, which can hinder their development of strong academic writing skills. In such contexts, generative AI has significant potential to enhance learning by complementing teacher feedback, offering immediate, personalized, and scalable support. Nonetheless, its effectiveness in education fundamentally depends on the quality and usability of the feedback it provides, rather than on automation alone, highlighting the importance of careful design and integration.

Against this background, this study investigates the pedagogical potential of ChatGPT as an AI-assisted feedback tool for Arabic writing in a *pesantren*-based university in Indonesia. Drawing upon the Student Feedback Literacy framework alongside Error Analysis, the study aims to analyze the lexical, syntactic, and morphological errors present in students' Arabic writing and to evaluate the extent to which ChatGPT generates feedback that is accurate, pedagogically meaningful, and capable of supporting learners' understanding and improvement of their written performance. Rather than viewing AI solely as an automated error-correction technology, this study conceptualizes ChatGPT as a formative feedback tool whose educational value depends on the quality, interpretability, and usability of the feedback it provides. By integrating linguistic error analysis with contemporary feedback theory, the study seeks to advance current understanding of how generative AI can support Arabic writing development while contributing to broader international discussions on AI-assisted formative assessment and feedback practices in Islamic higher education.

METHOD

This study was conducted at Universitas Darussalam Gontor, Ponorogo, East Java, Indonesia, a *pesantren*-based university where Arabic functions not only as the language of religious scholarship but also as a medium of daily communication and academic instruction. This setting provides a distinctive context for investigating AI-assisted Arabic writing because students experience intensive exposure to Arabic while continuing to encounter challenges typical of non-native learners in Islamic higher education. The study adopted a qualitative descriptive approach to explore both the linguistic characteristics of students' Arabic writing errors and the pedagogical value of ChatGPT-generated feedback. Two complementary frameworks guided the investigation. Error Analysis (EA) (Corder, 1981) served as the analytical framework for systematically identifying and classifying students' writing errors into lexical, syntactic, and morphological categories, whereas the Student Feedback Literacy framework provided the conceptual lens for evaluating whether ChatGPT's feedback was sufficiently accurate, understandable, and educationally meaningful to support learners in recognizing and improving their writing. This integration enabled the study to examine ChatGPT not merely as an automated error-correction system but as a potential formative feedback tool within Arabic language education.

The primary data comprised Arabic writing assignments produced by thirty first-year students enrolled in the university's Arabic matriculation program after completing an academic writing course. To gather relevant data, purposive sampling was employed because these students represented novice Arabic writers who were developing foundational linguistic competence while studying in an immersive *pesantren* environment. The writing tasks were selected because they reflected students' independent production of Arabic texts without external assistance. This approach thereby provided authentic evidence of their lexical, syntactic, and morphological difficulties. After anonymization, each writing sample was transcribed and subsequently submitted to ChatGPT (GPT-4, accessed in May 2024) using a standardized prompt. This prompt requested the model to identify linguistic errors, provide corrected versions, and explain the rationale for each correction. Employing a uniform prompting procedure ensured analytical consistency across all texts. Furthermore, this method allowed for a systematic comparison between students' original writing and the AI-generated feedback, thereby strengthening the validity of the analysis.

Moreover, the data analysis followed the iterative procedures of Error Analysis. Student writings were first examined to identify linguistic deviations, which were subsequently categorized into lexical, syntactic, and morphological errors. Each error was analyzed to determine its linguistic characteristics and possible sources, including first-language interference, incomplete understanding of Arabic grammatical structures, or limited mastery of morphological rules. Following this, ChatGPT's responses were then evaluated not only for linguistic accuracy but also for their pedagogical quality through the lens of Student Feedback Literacy. Specifically, the analysis considered whether the feedback clearly identified the error, provided an understandable explanation, and offered guidance that could facilitate learners' revision and future writing development. To strengthen the credibility of the findings, two senior lecturers in Arabic linguistics independently reviewed all identified errors and AI-generated corrections. The reviewers evaluated responses on linguistic accuracy, appropriateness, clarity, and usefulness, resolving disagreements through discussion to reach consensus. Ethical principles were strictly followed: participants gave informed consent, personal identifiers were removed for confidentiality, and procedures aligned with university standards to ensure integrity.

RESULTS AND DISCUSSION

Characteristics of Students' Arabic Writing Errors

Analysis of the thirty Arabic writing assignments produced by first-year students revealed 132 linguistic errors, demonstrating that despite studying in a *pesantren*-based university where Arabic is widely used in academic and religious activities, learners continued to experience considerable difficulties in producing grammatically accurate written texts. Following the Error Analysis framework (Corder, 1981), all identified errors were independently examined by two senior lecturers in Arabic linguistics and subsequently classified into three major categories: lexical, syntactic, and morphological. As presented in Table 1, lexical errors constituted the largest proportion of the dataset, followed by syntactic and morphological errors.

Table 1. Distribution of Arabic Writing Errors

Error Type	Frequency	Percentage
Lexical	52	39.4%
Syntactic	43	32.6%
Morphological	37	28.0%
Total	132	100%

Source: Author's analysis of students' Arabic writing samples

The predominance of lexical errors, as shown in Table 1, indicates that students encountered greater difficulty selecting contextually appropriate vocabulary than applying grammatical rules. Lexical errors were primarily characterized by inaccurate verb selection, inappropriate lexical collocations, and direct translation from Indonesian into Arabic. For example, several students used *نُطِقُ* ("to utter or pronounce") instead of *نَتَكَلَّمُ* ("to speak"), *يُوزَعُونَ* ("to distribute") instead of *يُشَارِكُونَ* ("to share"), and *نَتَّبِعُ* ("to follow") rather than *نُشَارِكُ* ("to participate"). Although these lexical choices conveyed partially related meanings, they failed to represent the communicative intentions required by the context. Expert reviewers agreed that these deviations reflected semantic transfer from Indonesian expressions and insufficient sensitivity to contextual word usage in Arabic.

Furthermore, syntactic errors represented the second largest category and primarily involved incorrect sentence structure, agreement relations, and the misuse of grammatical particles. Common problems included disagreement between relative pronouns and their antecedents, inaccurate construction of *idāfah* phrases, omission of obligatory particles such as *أَنْ*, and inappropriate agreement between verbs and subjects. For instance, students frequently employed the masculine relative pronoun *الَّذِي* to modify the feminine plural noun *الْعُلُومُ*, whereas the grammatically appropriate form is *الَّتِي*. Other examples included incorrect constructions following *لَكِي* and structurally incomplete complement clauses after *أَتَمَنَّى*.

Subsequently, morphological errors accounted for slightly more than one-quarter of all identified deviations and were associated with inaccurate word formation, derivational patterns, and inflectional agreement. Students commonly misused *maṣḍar* forms, produced incorrect adjective patterns, and demonstrated confusion regarding gender, number, and derivational morphology. Representative examples included the use of *سَمِعَ* instead of *لَوْ سَمِعَ*, instead of *لَا تَسْمَعُ*, and the non-standard adjective *مَمْلُوءَةٌ* where a verbal construction such as *يَمْلَأُونَ* was required. Expert reviewers noted that these errors often arose because learners attempted to construct Arabic words by analogy with familiar morphological patterns without fully understanding the underlying principles of *ṣarf*. Given the morphological richness of Arabic,

such inaccuracies not only affect grammatical correctness but may also alter semantic interpretation and reduce textual coherence (Boudelaa et al., 2026).

ChatGPT's Performance in Generating Corrective Feedback

Following the identification and classification of students' writing errors, each erroneous sentence was submitted to ChatGPT (GPT-4) using a standardized prompt requesting the model to identify the linguistic error, provide a corrected version, and explain the rationale underlying the correction. Subsequently, the AI-generated responses were independently evaluated by two senior lecturers in Arabic linguistics to determine their linguistic accuracy, contextual appropriateness, and pedagogical clarity. Based on the consensus of both reviewers, each response was then categorized as fully acceptable, partially acceptable, or inaccurate. Finally, Table 2 portrays the expert evaluation of ChatGPT corrective feedback, offering a comprehensive overview of the assessment process.

Table 2. Expert Evaluation of ChatGPT's Corrective Feedback

Feedback Quality	Frequency	Percentage
Fully acceptable	67	50.76%
Partially acceptable	31	23.48%
Inaccurate	34	25.76%
Total	132	100%

Source: Author's analysis of expert evaluations

Overall, as observed in Table 2, ChatGPT was able to produce linguistically appropriate corrections for the majority of the errors identified. The expert reviewers concurred that the model reliably recognized straightforward lexical and syntactic deviations, showcasing its strength in handling simpler error types. However, its performance was relatively weaker when dealing with morphologically complex constructions that demand a deeper understanding of contextual nuances and intricate morphological features, which posed a greater challenge for accurate correction. This indicates that while ChatGPT demonstrates a solid ability to address basic language errors, there remains room for improvement in managing more complex, context-dependent linguistic structures.

An important finding of this study is that ChatGPT did not perform uniformly across lexical, syntactic, and morphological errors. The model consistently generated highly accurate feedback for lexical substitutions and sentence-level grammatical structures but demonstrated lower consistency when addressing morphologically complex constructions. This pattern reflects the linguistic characteristics of Arabic itself. Arabic morphology is based on intricate root-and-pattern systems, inflectional variation, and extensive agreement mechanisms involving gender, number, person, and case (Belal, 2019; Soudain et al., 2018).

These findings share a similar vein with recent studies reporting that large language models perform well when grammatical corrections depend on explicit linguistic patterns but encounter greater difficulty when successful correction requires deeper contextual and morphological reasoning. Aljanabi (2024), for example, found that ChatGPT demonstrates promising Arabic parsing capabilities but still struggles with complex grammatical structures requiring expert linguistic interpretation. Likewise, Juričić and Sarić (2024) reported that AI-based grammar correction systems achieve high levels of accuracy for surface-level grammatical errors while exhibiting reduced reliability in linguistically complex contexts.

ChatGPT demonstrated its strongest performance in correcting lexical errors. The model consistently detected inappropriate vocabulary choices and proposed contextually appropriate alternatives that aligned with standard Arabic usage. For example, a student wrote, “نُطَقُ أَنْ نَسْتَطِيعَ أَنْ نَنْطِقَ فِي الْيَوْمِيَّةِ” (“We can pronounce in daily life”), where the verb نَطَقَ (“to pronounce”) was used instead of the more appropriate verb تَحَدَّثَ (“to speak”). ChatGPT correctly revised the sentence to “نَسْتَطِيعُ أَنْ نَتَحَدَّثَ يَوْمِيًّا” (“We can speak every day”) and explained that the original verb did not correspond to the intended communicative meaning. Both expert reviewers considered this correction fully acceptable because it not only restored semantic accuracy but also provided a concise explanation that could help learners distinguish between two verbs with related but different meanings. Similar performance was observed in corrections involving verbs such as يُوزَعُونَ (“to distribute”) and يُشَارِكُونَ (“to share”), indicating that ChatGPT effectively resolved vocabulary errors influenced by literal translation from Indonesian.

Performance remained consistently high when correcting syntactic errors, particularly those involving agreement relationships, sentence structure, and the use of grammatical particles. For instance, the student sentence “أَتَلَقَّى الْعُلُومَ الْكَثِيرَةَ الَّتِي تُسَاعِدُنِي” incorrectly employed the masculine relative pronoun الَّذِي to modify the feminine plural noun الْعُلُومَ. ChatGPT successfully reconstructed the sentence using the grammatically appropriate form الَّتِي, thereby restoring agreement between the antecedent and the relative clause. In another example, the model accurately inserted the particle أَنْ following the verb أَتَمَمْتُ to produce the correct complement structure “أَتَمَمْتُ أَنْ يَسْتَمِرَّ”. Nevertheless, the accompanying explanations occasionally focused on the corrected form without fully clarifying the underlying grammatical principles, resulting in several responses being categorized as only partially acceptable.

By contrast, ChatGPT demonstrated lower consistency in correcting morphological errors, which constituted the most challenging category for both learners and the AI system. Although the model generally recognized incorrect word forms, it occasionally proposed alternatives that were grammatically acceptable but less appropriate within the broader textual context. For example, when correcting the phrase “يَدْعُو الْأُسْتَاذُ الطَّلَّابَ إِلَى سَمْعِ الْأَسْلُوبِ الْجَمِيلِ”, ChatGPT appropriately replaced سَمْعَ with سَمَاعَ to produce the conventional maṣḍar form. However, in more complex cases involving derivational morphology and contextual word formation, expert reviewers noted that the AI’s explanations lacked sufficient detail regarding the underlying ṣarf principles.

The present findings, therefore, suggest that AI should not be viewed as a replacement for expert linguistic instruction. Instead, ChatGPT appears most effective when supporting aspects of writing that involve explicit grammatical regularities. Meanwhile, teachers remain indispensable for explaining morphologically rich constructions, semantic subtleties, and context-dependent language use. This complementary relationship, in turn, reinforces emerging perspectives that advocate for the responsible integration of generative AI into education (Meniado, 2023; Sullivan et al., 2023).

Educational Quality of ChatGPT Feedback through the Student Feedback Literacy Framework

While the preceding section evaluated the linguistic accuracy of ChatGPT’s corrections, the Student Feedback Literacy framework offers a broader perspective by examining whether the feedback meaningfully supports students’ learning. Guided by the framework proposed by Carless and Boud (2018), the AI-generated responses were reviewed to assess if they helped learners identify errors, grasp the underlying linguistic principles, stay confident during revisions,

and independently enhance their future writing. Table 3 presents an assessment of how well ChatGPT's responses align with the criteria of the student feedback literacy framework.

Table 3. Educational Quality of ChatGPT Feedback Based on the Student Feedback Literacy Framework

Feedback Literacy Dimension	Evidence from ChatGPT Feedback	Expert Evaluation
Appreciating feedback	Errors were explicitly identified and generally distinguished as lexical, syntactic, or morphological.	High
Making judgments	Explanations frequently clarified why corrections were required, although grammatical reasoning varied in depth.	Moderate–High
Managing affect	Feedback employed neutral, supportive language without evaluative or punitive expressions.	High
Taking action	Most responses enabled independent revision, but several morphological corrections lacked sufficient explanation to support future transfer of learning.	Moderate

Source: Author's analysis of ChatGPT Feedback Based on the SFL Framework

The first dimension, in Table 3, appreciating feedback, concerns learners' ability to recognize what is incorrect and why feedback deserves attention. Across the analyzed responses, ChatGPT consistently identified problematic words, grammatical constructions, and morphological forms before proposing revisions. In lexical examples, the model explicitly differentiated between semantically related verbs, explaining that *نُطِيقُ* refers to “pronouncing” rather than “speaking,” thereby justifying its recommendation to replace it with *نَتَحَدَّثُ*. Similarly, syntactic feedback clearly identified agreement problems between relative pronouns and their antecedents instead of merely presenting corrected sentences.

The second dimension, making judgments, evaluates whether learners are provided with sufficient information to understand why a correction is appropriate. The findings indicate that ChatGPT frequently explained the grammatical or semantic rationale underlying its revisions, particularly for lexical and syntactic errors. For example, when correcting the sentence “*أَتَمَنَّى هَذَا الْبِرْتَامَج سَيَسْتَمِرُّ*”, ChatGPT not only inserted the particle *أَنْ* but also explained that the verb *أَتَمَنَّى* requires a subordinate verbal clause introduced by *أَنْ*. The quality of explanations varied. In many morphological cases, ChatGPT replaced incorrect forms with correct ones without explaining *ṣarf* or derivational morphology. While learners identified correct answers, they often didn't understand the linguistic reasoning. Expert reviewers saw these responses as only partially effective for learning.

The third dimension, managing affect, concerns the emotional and motivational qualities of feedback. Throughout the dataset, ChatGPT consistently adopted supportive and non-judgmental language. Rather than labeling students' writing as “wrong” or “incorrect,” the model typically used constructive expressions such as “a more appropriate form is...” or “this verb is commonly used in this context.” This finding suggests that generative AI can create psychologically supportive feedback environments that complement teacher-led instruction. Consequently, this indicates the potential for AI to foster a more positive learning atmosphere.

The fourth dimension, taking action, represents the ultimate goal of feedback literacy by examining whether learners can use feedback to improve subsequent performance. For lexical and syntactic errors, the combination of explicit error identification, corrected forms, and concise explanations provided learners with sufficient information to revise their texts

independently. Expert reviewers highlighted that responses which explained the underlying linguistic principles were more valuable than those merely rewriting incorrect sentences. When ChatGPT provided only the corrected form without grammatical justification, students could imitate the revision but might not apply the knowledge elsewhere. Explanations based on semantics or grammar foster deeper understanding and help students learn independently.

Viewed through the Student Feedback Literacy framework, the model effectively supported learners in recognizing writing problems, understanding many grammatical and lexical principles, and revising their texts through constructive and non-threatening feedback. Moreover, this approach provided learners with a supportive environment that encouraged ongoing improvement and self-reflection. Its pedagogical effectiveness varied with linguistic complexity, especially in Arabic morphology, where explanations were less comprehensive. Evaluating AI feedback requires assessing correctness, depth, and clarity, focusing on developing learners' feedback literacy and self-regulated writing in Arabic. Building these skills enhances learners' autonomy and confidence, leading to better learning outcomes.

Beyond linguistic accuracy, the findings further illustrate how immediate AI-generated feedback can serve as a significant aid in students' development as self-regulated writers. Students were able to revise their texts immediately after receiving ChatGPT's feedback, instead of postponing revision until instructor comments became available. This real-time feedback loop offers several pedagogical advantages, such as encouraging students to critically monitor their own writing process, compare original and revised versions, and actively engage with linguistic improvements. Consequently, this iterative process not only enhances students' writing skills but also promotes a mindset of continuous learning and self-improvement, which are essential components of effective language education.

Nevertheless, the study also demonstrates that immediacy alone does not guarantee effective learning. The expert evaluation revealed clear differences between feedback that merely supplied corrected sentences and feedback that explained the linguistic rationale behind those corrections. Explanatory feedback provided greater opportunities for students to internalize grammatical principles and transfer their knowledge to subsequent writing tasks, whereas simple rewriting primarily supported short-term textual revision. These findings align with studies by Athanassopoulos et al. (2023), Xiao and Zhi (2023), and Barrot (2024), which emphasize that learners benefit most from generative AI when they critically engage with its explanations rather than accepting its outputs uncritically.

Pedagogical Implications of AI Feedback for Arabic Writing in *Pesantren*-Based Higher Education

The findings have important implications for Arabic language pedagogy within *pesantren*-based universities, where Arabic serves simultaneously as the language of Islamic scholarship, academic communication, and daily interaction. Despite the immersive linguistic environment, students continued to experience substantial lexical, syntactic, and morphological difficulties, indicating that exposure alone does not ensure writing proficiency. This observation supports previous research highlighting that Arabic instruction in many *pesantren* contexts continues to emphasize grammatical knowledge and textual analysis while providing fewer opportunities for sustained writing practice accompanied by timely formative feedback (Zarkasyi et al., 2024).

When interpreted collectively through the Student Feedback Literacy framework, the results suggest that AI-generated feedback can function as an important pedagogical resource that complements conventional Arabic writing instruction in *pesantren*-based higher education. This highlights the potential for AI to enhance educational practices, especially in environments where resources may be limited. In such settings, where one instructor is often responsible for providing feedback to large numbers of students, AI-assisted feedback offers valuable

opportunities to expand formative assessment beyond the traditional limitations of classroom time, thus supporting more personalized and continuous learning processes.

However, the findings equally demonstrate that teacher supervision remains indispensable, particularly when students encounter complex morphological constructions, culturally nuanced expressions, or linguistically ambiguous contexts. This emphasizes that despite the advantages of AI, human oversight is essential to ensure accuracy and contextual relevance. Rather than replacing teachers, ChatGPT should therefore be integrated into a layered feedback model, in which AI provides the initial stage of formative feedback. Subsequently, instructors can validate, elaborate on, and contextualize the corrections, ensuring that feedback is both accurate and pedagogically meaningful. This collaborative approach aims to optimize the strengths of both AI technology and human expertise, ultimately fostering a more effective and nuanced learning experience.

The findings further indicate that immediate AI-generated feedback may significantly reduce students' dependence on teacher-centered correction by encouraging greater learner autonomy during the writing process. This is because, rather than relying exclusively on instructors to identify every linguistic error, students can engage in an iterative cycle of drafting, receiving feedback, revising, and reflecting before ultimately submitting their work for formal assessment. Such a process aligns closely with the principles of self-regulated learning, whereby learners actively monitor and improve their own performance through continuous feedback and revision, thereby fostering a more independent and proactive approach to learning. (Lu et al., 2026; Selvi et al., 2026; Shen & He, 2026).

The results also highlight the importance of positioning ChatGPT as a scaffold for Arabic writing rather than as an autonomous instructional authority. Within pesantren-based universities, where Arabic instruction emphasizes both grammatical precision and the accurate interpretation of Islamic texts, teacher expertise remains indispensable for validating AI-generated corrections, explaining nuanced morphological structures, and connecting language use with broader communicative and religious contexts. Moreover, this underscores the crucial role of human instructors in ensuring the reliability and depth of language learning, especially in specialized educational settings. Based on these findings, a pedagogical model of AI-assisted Arabic writing can be proposed (Figure 1).

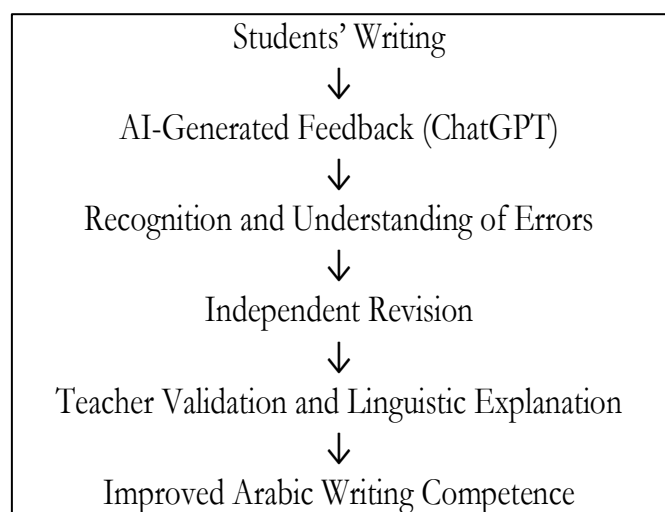


Figure 1. Pedagogical Model of AI-Assisted Feedback for Arabic Writing

Figure 1 depicts the model and illustrates that students initially produce independent writing, then receive formative feedback from ChatGPT, which helps them identify and better understand their errors. Next, students develop a preliminary understanding of their linguistic mistakes and revise their texts accordingly, which serves as an essential step towards improving their language skills. Subsequently, they obtain validation and further explanation from their instructors, who offer expert insights to deepen understanding. Through this iterative cycle, AI functions primarily as an accessible first layer of formative assessment, providing immediate feedback, while teachers, with their specialized knowledge, offer the higher-order linguistic interpretation, contextualization, and pedagogical guidance necessary for sustained language development and mastery.

Overall, these findings suggest that the integration of generative AI into Arabic writing instruction should not be understood as replacing teachers but as strengthening formative assessment practices within Islamic higher education. By providing immediate, individualized, and supportive feedback, ChatGPT enables students to engage more actively with their own writing while allowing instructors to concentrate on higher-order aspects of language learning, including grammatical reasoning, contextual interpretation, and academic writing development. Consequently, the pedagogical significance of AI lies in its capacity to enhance feedback literacy, promote self-regulated writing, and create more sustainable feedback practices within *pesantren*-based universities and other Arabic language learning contexts.

CONCLUSION

This study demonstrates that ChatGPT has considerable potential to function as an AI-assisted formative feedback tool for Arabic writing in *pesantren*-based higher education. Using Error Analysis, the findings revealed that students' writing was characterized by recurring lexical, syntactic, and morphological errors, while expert evaluation showed that ChatGPT generated generally accurate corrective feedback, particularly for lexical and syntactic problems, although its performance was less consistent in addressing morphologically complex structures. More importantly, interpreting these findings through the Student Feedback Literacy framework revealed that the educational value of ChatGPT extends beyond linguistic correction by supporting learners in recognizing errors, understanding corrective explanations, and engaging in independent revision. These findings suggest that the effectiveness of generative AI in language education should be evaluated not only by its technical accuracy but also by its capacity to foster meaningful learner engagement with feedback and self-regulated writing development. Accordingly, ChatGPT should be positioned as a complementary pedagogical resource rather than a substitute for teachers, with instructor expertise remaining essential for validating AI-generated feedback and facilitating deeper grammatical understanding. Future research should investigate how sustained interaction with AI-assisted feedback influences the development of students' feedback literacy and Arabic writing proficiency over time, while also comparing the pedagogical effectiveness of different large language models across diverse instructional contexts and learner proficiency levels.

AUTHORSHIP CONTRIBUTION STATEMENT

AHBZ: conceptualization, writing original draft, review, interpretation of results and discussion, final revision of the manuscript. M: methodology, writing review, editing. MI: data collection, data analysis. ACS and AY: investigation, data validation, editing.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest related to the research, authorship, or publication of this manuscript.

GENERATIVE AI STATEMENT

During the preparation of this manuscript, the authors used ChatGPT to assist with the translation and language editing of selected portions of the manuscript into English and Scopus AI to support the identification and exploration of relevant literature. In this study, ChatGPT (GPT-4 model, accessed May 2024) served as the object of investigation and was used to generate suggested corrections to students' Arabic written texts based on a standardized prompt. The authors reviewed and edited all AI-assisted content and take full responsibility for the accuracy, originality, and integrity of the final manuscript.

BIBLIOGRAPHY

- Al-Abri, A. (2025). Exploring ChatGPT as a virtual tutor: A multi-dimensional analysis of large language models in academic support. *Education and Information Technologies*, 30(12), 17447–17482. <https://doi.org/10.1007/s10639-025-13484-x>
- Aljanabi, M. (2024). Assessing the Arabic parsing capabilities of ChatGPT and Claude: An expert-based comparative study. *Mjals*, (2024), 16–23. <https://doi.org/10.58496/mjals/2024/002>
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/amlr.2023.02.009>
- Barrot, J. S. (2024). ChatGPT as a language learning tool: An emerging technology report. *Technology, Knowledge and Learning*, 29(2), 1151–1156. <https://doi.org/10.1007/s10758-023-09711-4>
- Belal, A. S. (2019). Comprehensive processing for Arabic texts to extract their roots. *Iraqi Journal of Science*, 60(6), 1404–1411. <https://doi.org/10.24996/ij.s.2019.60.6.25>
- Boudelaa, S., Fathi, N., & Ansary, S. A. (2026). Novel dynamics in semantic ambiguity processing: Insights from a Semitic language. *Journal of Memory and Language*, 146. <https://doi.org/10.1016/j.jml.2025.104709>
- Boudouaia, A., Khattala, A., & Bennouioua, H. (2026). Assessing the integration of artificial intelligence-generated content feedback in English language writing learning. *Journal of Computing in Higher Education*. <https://doi.org/10.1007/s12528-026-09491-7>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Guo, X. (2025). Facilitator or thinking inhibitor: Understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, 33(5), 3526–3544. <https://doi.org/10.1080/10494820.2024.2445177>
- Juričić, V., & Sarić, D. (2024). Evaluation of AI-based grammar correction for portuguese. *Hrčak – Portal of Croatian Scientific and Professional Journals*, 25(2), 177–190. <https://doi.org/10.29162/jez.2024.6>
- Lu, Q., Yao, Y., Xiao, L., Zhu, X., & Yin, H. (2026). Can GenAI help undergraduate students become independent writers? An intervention study on its effects on writing motivation, feedback literacy, and self-regulated learning strategies. *Educational Psychology*, 1–27. <https://doi.org/10.1080/01443410.2026.2654526>
- Meniado, J. C. (2023). The impact of ChatGPT on English language teaching, learning, & assessment: A rapid review of literature. *Arab World English Journal*, 14(4), 3–18. <https://doi.org/10.24093/awej/vol14no4.1>
- Selvi, A. E., Irgatoğlu, A., Yurtadur, O., & Dağbaşı, G. (2026). Sustainable L2 writing pedagogy in Turkish higher education: Effects of AI-mediated feedback on self-regulated learning and writing performance. *PloS One*, 21(7), e0344618. <https://doi.org/10.1371/journal.pone.0344618>

- Shen, Y., & He, S. (2026). Exploring the impact of AI-generated feedback on EAP writing: Writing performance, self-regulation, and AI literacy. *2026 14th International Conference on Information and Education Technology, ICIET 2026*, 22–26. <https://doi.org/10.1109/ICIET69664.2026.11561596>
- Soudain, N., Bounhas, I., & Babbis, S. B. (2018). Ambiguity aware Arabic document indexing and query expansion: A morphological knowledge learning-based approach. *Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018*, 230–235. Retrieved from <https://www.scopus.com/pages/publications/85064814024?origin=resultslist>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6. <https://doi.org/10.37074/jalt.2023.6.1.17>
- Susiawati, I., Mardani, D., Matin, I., & Thamrin, M. H. (2026). Exploring AI-assisted Arabic learning for linguistic intelligence development among Indonesian university students. *Indonesian Journal of Applied Linguistics*, 16(1), 81–91. <https://doi.org/10.17509/8c995612>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>
- Zarkasyi, H. F., Mas' ud, F., Hidayatullah, R. A., & Khakim, U. (2024). Strategy of Indonesian Pesantren University in achieving competence of student: A grounded research at UNIDA Gontor. *KnE Social Sciences*, 9(6), 119-138. <https://doi.org/10.18502/kss.v9i6.15260>
- Zhang, L., Deng, X., & Shadiev, R. (2026). A meta-review of generative AI in education: Synthesizing findings from systematic reviews. *Journal of Educational Computing Research*, 64(4), 1068–1092. <https://doi.org/10.1177/07356331261419689>