
RESEARCH ARTICLE

The Impact of AI-based Feedback on English Writing Performance: A Systematic Review and a Meta-analysis Study

Mona Alzahrani

Assistant Professor, Foreign Languages Department, Taif University, Saudi Arabia

Corresponding Author: Mona Alzahrani, **E-mail:** m.alzahrane@tu.edu.sa

ABSTRACT

Artificial intelligence has increasingly transformed educational practices, particularly in writing instruction. Among these innovations, AI-based feedback tools have appeared as favourable solutions for giving personalized, immediate, and accessible support to learners. This study aims to assess the effect of AI-based feedback on English writing performance, filling research gaps regarding the impact of English language proficiency level, duration of intervention, and implementation setting on the effectiveness of AI-based feedback tools in improving students' English language writing performance. A systematic review and meta-analysis were utilized to examine 24 empirical studies published from January 2022 to March 2026 across five databases (including EBSCO, ProQuest, ERIC, Web of Science, and Wiley Online Library). The results indicated that AI-based feedback tools had a large overall effect size (Hedges' $g = 1.298$, $p < 0.001$) on students' writing performance. Moreover, moderator analyses revealed that learners with intermediate proficiency benefited more substantially compared to advanced learners, suggesting that AI feedback is particularly effective for developing writers. In terms of implementation settings, blended learning environments yielded stronger effects than traditional classroom settings, highlighting the importance of flexible and technology-supported contexts. Although the duration of intervention did not significantly moderate the results, consistently strong effects were observed across the three periods of intervention: short, medium, and long-term. These findings suggest that AI-based feedback tools represent a powerful and scalable approach to enhancing English writing skills.

KEYWORDS

Artificial intelligence, AI-based tools, English language, meta-analysis, writing feedback

ARTICLE INFORMATION

ACCEPTED: 30 June 2026

PUBLISHED: 31 July 2026

DOI: 10.32996/jeltal.2026.8.8.12

1. Introduction

The artificial intelligence area has progressed very fast in recent years. This technology has been widely utilized in education, including teaching the English language and particularly writing skills. With the appearance of technologies, such as natural language processing (NLP), AI feedback tools have been used to offer instant and personalized feedback to learners, which changes the traditional ways of teaching writing (Zhai et al., 2021; Zhang & Hyland, 2022). Such technologies have received considerable attention within the context of EFL and ESL due to the importance of instant feedback for students' writing proficiency, especially with larger groups (Thi et al., 2025).

The process of writing in English is rather complex and requires both linguistic and cognitive efforts associated with developing the content of written texts and organizing the material. Although conventional teacher feedback is essential, it is limited by the number of lessons per week and may be insufficient because of the constraints related to class size and time restrictions (Lee, 2020). Therefore, AI-based feedback can serve as a helpful addition, providing students with instant recommendations regarding their grammatical, lexical, and structural problems (Wang et al., 2022).

However, even though the use of artificial intelligence in teaching writing has increased, there is still significant irregularity in studies regarding the effectiveness of this technology. While some studies show that the use of feedback from AI improves the quality, accuracy, and fluency of writing by students (Barrot, 2023; Thi et al., 2025), other studies indicate that the use of AI might

not be beneficial in solving complex issues such as argumentative writing and critical thinking skills (Zhang & Hyland, 2022). While it seems that the duration of interventions does not affect their outcome, it appears that the effectiveness of AI-based feedback is highly dependent on various contextual factors, including language skills among students and conditions in which teaching occurs (Zhai & Ma, 2022; Fleckenstein et al., 2023; Alnemrat et al., 2025).

With all this conflicting evidence, it is clear that a comprehensive and systematic synthesis of all previous research about AI-based feedback tools should be conducted in order to discover the overall efficiency of the discussed tools and to reveal under what circumstances they work the best. The methodology of meta-analysis allows performing an integrative analysis of results obtained from several different studies and identifying potential moderator variables (Borenstein et al., 2021). More specifically, the analysis of the effects of factors such as English language proficiency, the environment where AI-based feedback is implemented (classroom, online, or blended), and the duration of the intervention can shed light on how the technology works. Thus, the current study will try to analyze the impact of AI-based feedback on English language writing performance. Besides, the possible mediating role of factors such as English language proficiency, the type of implementation environment, and the duration of the intervention period will be evaluated.

2. Literature Review

The fast development of AI technology has made great changes to teaching practices, especially L2 writing pedagogy. The use of AI-based feedback tools, such as automated writing evaluation (AWE) and automated written feedback (AWF) systems, makes it possible to provide instant, personalized feedback to students on their writing. Through analyzing linguistic, syntactic, and rhetorical features of written texts using NLP techniques, these tools give feedback, which helps students modify their writing (Xue, 2024; Zhai & Ma, 2022). This means that AI-based feedback has gained increasing prominence in EFL/ESL classrooms because timely feedback plays a critical role in developing students' writing skills.

Thus, there have been many applications of AI-based feedback tools in different educational settings, including automated, human-assisted, and hybrid applications involving both human beings and machines. In the beginning, AI-based feedback tools were used to score students' essays automatically or provide surface-level feedback on grammar, punctuation, and vocabulary usage in their essays (Shi & Aryadoust, 2024; Zhang & Hyland, 2022). With the development of generative AI and LLMs, however, feedback can be given regarding higher-order writing skills of students, such as organization, coherence, and reasoning. In other instances, the use of AI feedback complements traditional forms of feedback provided by instructors. According to Tran (2025), there were increased revision rates and improved writing outcomes when both AI and instructor feedback were combined compared to situations where each was applied individually. Similar results have been reported by Bai and Nordin (2025), who have found the combination of human-AI feedback beneficial for increasing writing accuracy, complexity, and fluency. Therefore, it can be argued that a combination of the two types of feedback provides the best opportunity for enhancing students' writing skills.

Moreover, the use of AI feedback in the context of blended learning is becoming more widespread due to its convenience. Since blended learning involves the use of technology to facilitate continuous learning outside the class, the use of AI can contribute to providing constant and instant feedback on writing assignments. Previous studies have found that, in a blended learning setting, the role of AI applications was mostly to support asynchronous learning (Park & Doo, 2024). Furthermore, current research has investigated the effectiveness of using generative AI software such as ChatGPT and Grammarly in writing instruction. As mentioned earlier, AI tools can offer more interactive feedback compared to traditional Automated Writing Evaluation (AWE). According to Chang et al. (2021), students who benefited from Grammarly outperformed their counterparts who were taught traditionally, with a medium effect size. In a similar path, Alanazi et al. (2025) reported that feedback generated from ChatGPT led to marked improvements in the writing ability of the participants.

In summary, AI software for generating feedback for writing tasks has become much more advanced than merely automated checking programs for basic grammatical errors. Such tools help in developing both lower-order and higher-order writing skills. Regarding the effectiveness of AI-based feedback as a means for improving writing, several studies have shown the effectiveness of the use of AI feedback in relation to certain dimensions of writing. For instance, Du and Nordin (2026) indicated that students who used AWE systems made marked gains in terms of writing accuracy, fluency, and complexity, in addition to experiencing lower levels of writing anxiety. Likewise, Chan et al. (2024) indicated that AI feedback resulted in considerable gains in the quality of essays written by students, in addition to raising their levels of motivation.

Nonetheless, there are some findings that reveal slightly different results. Alnemrat et al. (2025), for example, argue that although both feedback types result in improved writing performance, there are no significant differences between AI-generated and teacher-generated feedback, thus implying that the two types of feedback are equally effective, but neither can be considered superior to the other. Another important observation is related to what kind of writing problems AI tools can be used for successfully. While low-level writing issues, such as grammar and vocabulary, tend to be adequately handled by AI tools, higher-order writing skills, including argumentation and reasoning, remain underdeveloped despite the use of AI feedback (Tran, 2025; Shi & Aryadoust, 2024). Besides, students' involvement in the process is another factor that should be considered. Indeed, according to Zhang (2025), there is evidence that the effectiveness of feedback produced by AI tools depends greatly on how useful students find the process and how engaged they are in using suggested corrections and recommendations.

Despite the increasing number of empirical studies investigating the application and effectiveness of AI-based feedback in writing instruction, various gaps still exist in current literature. The first one is related to the high heterogeneity among existing studies in terms of their designs, contexts, and assessment of writing ability, which contributes to inconsistent findings regarding the size of effects of AI-based feedback in English writing instruction (Shi & Aryadoust, 2024). The second limitation of current literature is related to the under-representation of systematic studies investigating the influence of various factors (e.g., English proficiency level, learning context, etc.) on the effectiveness of AI-based feedback. Finally, there is a need for meta-analytic studies investigating how AI-based feedback is used in different contexts and its overall effectiveness.

Therefore, the present study intends to bridge these gaps by means of conducting a systematic review and meta-analysis of the literature on AI-based feedback in English writing instruction. To do so, this paper attempts to answer the following research questions:

- (1) How are AI-based tools implemented for students' academic writing feedback?
- (2) How effective is AI-based feedback in enhancing students' academic writing performance?
- (3) To what extent is the effect of AI-based feedback moderated by English language proficiency level, duration of intervention, and learning setting?

3. Methodology

3.1 Searching Process

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021) was used to carry out this review. The researcher conducted an extensive search across various online academic databases, including EBSCO, ProQuest, Web of Science, ERIC, and Wiley Online Library. The articles published from January 2022 to March 2026 was selected.

Furthermore, Boolean search strings were employed to identify relevant studies, specifically targeting the title, abstract, and keywords of the literature: ("Artificial intelligence" OR "Generative AI" OR "Generative AI tools" OR "AI-based feedback" OR "AI-based Tools") AND ("Writing skill" OR "Writing performance" OR "Academic writing" OR "EFL w" OR "ESL writing").

3.2 Inclusion and exclusion criteria

To minimize potential bias and ensure the relevance of the search results, specific selection criteria were applied to screen full-text studies. Table 1 displays both the inclusion and exclusion criteria.

Table 1. Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
Empirical Studies	Non-empirical studies: (e.g., review, meta-analysis, conceptual, opinion, and editorial articles).
Published between January 2022 and March 2026	Published before 2022
English-language articles	Non-English articles.
Available in Full text.	Not available in full text.
AI in academic writing.	AI in fields other than academic writing
Quantitative studies contain enough data to calculate effect sizes	Qualitative studies

3.3 Screening Process

The PRISMA checklist (Page et al., 2021) was employed to identify, select, and report the research findings (See Figure 1). During the preliminary literature search, duplicates were removed, and 1520 documents were identified. After reviewing titles and abstracts, 1418 articles were excluded. Subsequently, 102 full-text articles were read to assess if they met the inclusion and exclusion criteria. At the selection stage, full texts were screened based on these eligibility criteria: (1) include experimental and control groups, (2) numbers of participants, English language proficiency levels (beginners, intermediate, and advanced) are clearly shown; and (3) evaluation results of the intervention were reported (means and standard deviation). Finally, 24 articles were included in the current systematic review and meta-analysis.

3.4 Document Coding

To ensure the reliability of results, high-quality data extraction and coding are critical (Cooper et al., 2019). The coding framework (See Table2) includes two parts: (1) Basic literature information, such as the article author, publication date, and

country or region. (2) Feature value coding, including English language proficiency level, intervention duration, implementation setting, sample sizes for the experimental and control groups, and effect size calculation.

Inter-rater reliability was calculated using Cohen's Kappa to assess the reliability of the document coding, yielding a value of 0.93 (> 0.7), which indicates a very high level of consistency between the researcher and an independent reviewer.

4. Data Analysis

To analyze the overall effect size and moderator variables' effect sizes, the Comprehensive Meta-Analysis program (CMA, version 4.0) was used. The number of samples (N), mean (M), and standard deviation (SD) from the several independent studies included in this article were manually entered into the CMA software for analysis. The effect sizes were calculated using Hedges's g to assess the impact of each study. The variability between research is taken into account in order to obtain a more accurate estimation of the true effect size and enable more precise comparisons between studies with different sample sizes (Bai et al., 2020).

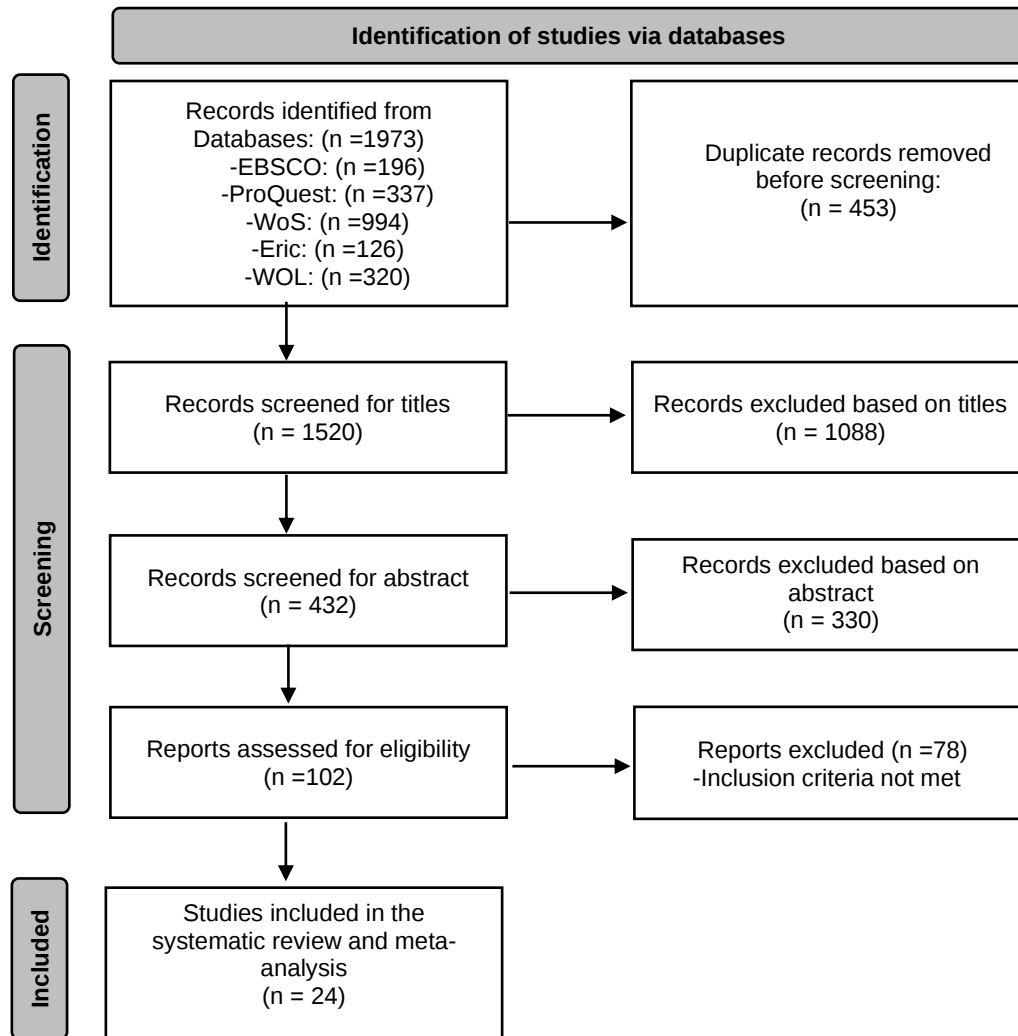


Figure 1. PRISMA Flow Diagram of the Article Search and Selection Process

Table 2. Document Coding

Author (Year)	Country/ Region	English Language Proficiency Level	Duration of Intervention	Implementation Setting	Number of Samples	
					EG	CG
Al Ghaithi & Behforouz (2025)	Oman	Intermediate	Short	Blended	20	20
Ali et al. (2025)	Saudi Arabia	Intermediate	Short	Blended	50	50
Alnemrat et al. (2025)	Jordan	Intermediate	Short	Classroom	60	60
Bašić et al. (2023)	Croatia	Advanced	Short	Classroom	9	9
Biju et al. (2024)	Bangladesh	Intermediate	Medium	Classroom	35	35
Chan et al. (2024)	China	Advanced	Short	Classroom	342	576
Chen & Gong (2025)	China	Advanced	Long	Blended	25	25
EKİZOĞLU & DEMİR (2025)	Turkey	Intermediate	Medium	Classroom	30	30
Escalante et al. (2023)	Asia–Pacific	Intermediate	Medium	Classroom	23	25
Gayed et al. (2022)	Japan	Intermediate	Short	Classroom	5	5
Jamshed et al. (2025)	India	Intermediate	Long	Blended	56	56
Mahapatra (2024)	India	Advanced	Short	Classroom	37	35
Mahdi et al. (2025)	Kuwait	Intermediate	Short	Classroom	23	22
Mekheimer (2025)	Egypt	Intermediate	Medium	Classroom	30	30
Sari & Han (2024)	Turkey	Intermediate	Long	Classroom	45	45
Sari & Lestari (2025)	Indonesia	Intermediate	Medium	Classroom	30	30
Sawangwan (2024)	Thailand	Intermediate	Short	Blended	25	25
Tedjo (2022)	Indonesia	Intermediate	Medium	Classroom	10	10
Vo & Ho (2026)	Vietnam	Intermediate	Medium	Classroom	45	45
Wale & Kassahun (2024)	Ethiopia	Intermediate	Medium	Blended	45	47
Wang & Han (2022)	China	Intermediate	Short	Classroom	35	35
Wiboolyasarin et al. (2024)	Thailand	Intermediate	Long	Classroom	18	21
Xiao (2024)	China	Intermediate	Long	Blended	25	26
Zhang et al. (2025)	Malaysia	Advanced	Long	Classroom	30	30

4.1 Testing heterogeneity and calculating the overall effect size

Both the Q-test and the I^2 -test were used in the heterogeneity testing to evaluate the consistency of the included research. To ascertain the total effect magnitude, either a fixed-effect model or a random-effect model must be chosen. The fixed effects model assumes that the true effect size is constant throughout all investigations, with deviations resulting from random error. On the other hand, the random effects model permits genuine changes and true effect size discrepancies between studies because of random error (Schmidt et al., 2009). The fixed effect model is suitable when heterogeneity is small, while the random effects model should be applied when it is large (Borenstein et al., 2021). A higher I^2 score indicates increased heterogeneity. Low heterogeneity is indicated by values between 0% and 25%, moderate heterogeneity is indicated by values between 25% and 75%, and high heterogeneity is indicated by values between 75% and 100% (Higgins, 2003). The current study showed significant heterogeneity ($Q = 505.527$, $I^2 = 95.450$, $p < 0.000$), indicating the use of the random effects model (see Table 3).

Table 3. Overall Effect Size and the Heterogeneity Test.

Effect Model	Hedges' g	n	Standard error	95% (CI)		Heterogeneity test result			
				Lower	Upper	Q	df	I^2	P
Random	1.298	24	0.239	0.830	1.765	505.527	23	95.450	0.000

Notes: n, number of effect sizes; CI, confidence interval; Q, Cochran's Q statistic; df, degrees of freedom; I^2 , heterogeneity percentage; P, p value.

4.1 Publication bias

A funnel plot, the Begg and Mazumdar test, and the classic fail-safe N test were used to assess publication bias before the meta-analysis. The validity of the meta-analysis was visualized using a funnel plot, with a symmetric inverted funnel indicating the absence of publication bias (Egger et al., 1997). The asymmetrical funnel plot in Figure 2 suggests publication bias, as smaller studies with negative or non-significant results are probably absent. Begg and Mazumdar test and the fail-safe N test were then used to determine the statistical significance of this asymmetry. The Begg and Mazumdar test shows non-significant results (Kendall's tau ≈ 0.23 , $p \approx 0.101 > 0.05$), indicating no statistically significant evidence of publication bias according to this method. The classic fail-safe N estimates the number of missing studies needed to render the meta-analysis's p-value non-significant, demonstrating the robustness of the observed effect size (Borenstein et al., 2021). The anticipated effect size of unpublished research is unlikely to influence the effect size of the meta-analysis if the fail-safe N is more than $5N + 10$ (Rosenthal, 1979). The analysis of the current study found a classic fail-safe N of 2159, which exceeded the acceptance level of 130 ($5 \times 24 + 10$), demonstrating that 2,159 more "null" studies would be required to render the total effect non-significant, suggesting that the results of the meta-analysis are extremely reliable and difficult to reverse due to publication bias. These outcomes support the validity and reliability of the study's results.

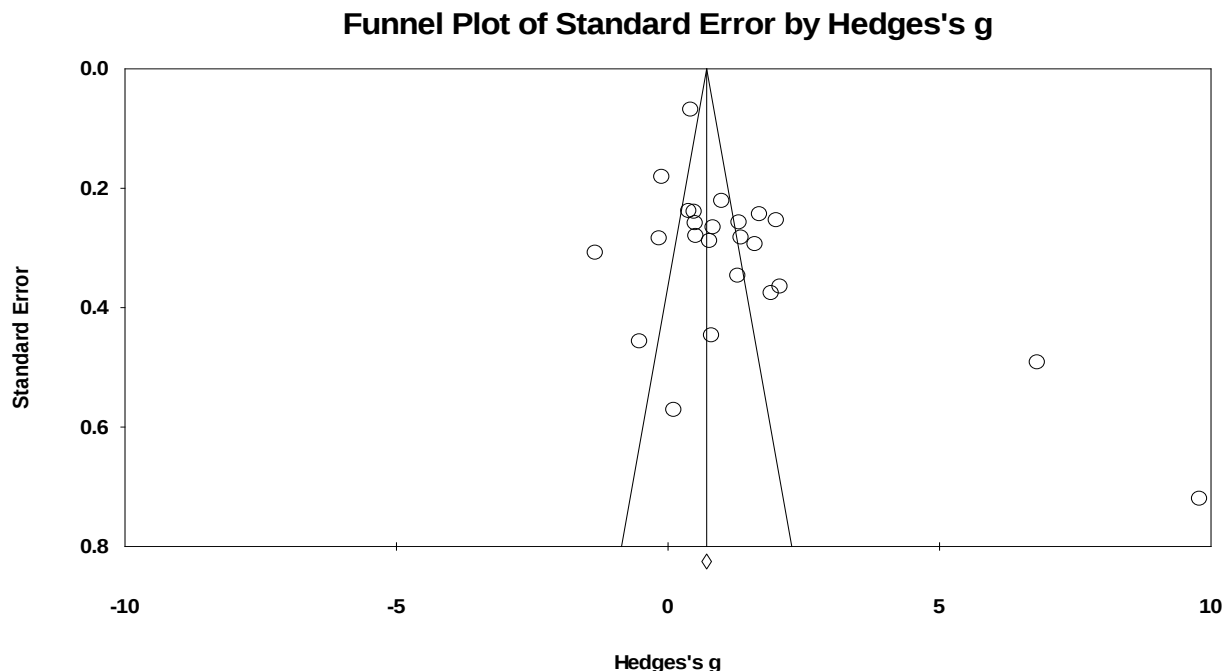


Figure 2. Funnel Plot for AI-based Feedback on English Academic Writing

5. Findings

5.1 The implementation of AI-based tools in students' academic writing feedback

Variables, including English language proficiency level, implementation setting, and intervention duration, were selected to address RQ1 because they are widely discussed in prior studies and facilitate data analysis. Furthermore, these aspects have the potential to significantly influence the implementation and effectiveness of AI-based feedback tools in enhancing students' academic writing performance (Ranalli, 2018).

5.1.1 English Language Proficiency Level

The students' levels of English language proficiency were classified based on the CERF framework (Council of Europe, 2001) into beginners (A1 and A2), intermediate (B1 and B2), and advanced (C1 and C2) (see Table 4). These levels shape learning by determining the complexity of input, the degree of learner autonomy, and the type of instructional support required at each stage. Nineteen of the included studies involved participants with an intermediate level of English proficiency, while five studies focused on advanced learners. This distribution indicates that students at the intermediate level were the main users of AI-based feedback tools, with comparatively fewer studies examining outcomes at higher proficiency levels.

Table 4. English Language Proficiency Level of Participants in Review Papers

Dimension	Category	N (%)
English language proficiency level	Intermediate	19 (79.17%)
	Advanced	5 (20.83%)

5.1.2 Implementation Setting

As shown in Table 5, the implementation settings of AI-based feedback tools in writing were categorized as Classroom, Online, and Blended. Most studies (N = 17) implemented the AI-based feedback in formal classroom settings (e.g., Mahapatra, 2024; Mahdi et al., 2025; Mekheimer, 2025; Sari & Han, 2024; Sari & Lestari, 2025). Seven studies implemented AI-based feedback tools in writing in a blended environment (e.g., Al Ghaithi, 2025; Jamshed et al., 2025; Wale & Kassahun, 2024), and no study implemented the assessment tool online.

Table 5. Implementation Setting of Review Papers

Dimension	Category	N (%)
Implementation setting	Classroom	17 (70.83%)
	Blended	7 (29.17%)
	Online	0

5.1.3 Duration of intervention

As indicated in Table 6, the duration of interventions ranged from less than 5 weeks to more than 8 weeks, with less than 5 weeks classified as short, 5-8 weeks classified as medium, and more than 8 weeks classified as long. Most of the studies lasted for less than 5 weeks (N = 10). Other studies applied the AI-based feedback tools in both medium-intervention (5-8 weeks) designs (N = 8) and long-intervention (>8 weeks) designs (N = 6).

Table 6. Duration of Intervention

Dimension	Category	N (%)
Duration of intervention	Short (< 5 weeks)	10 (41.67%)
	Medium (5-8 weeks)	8 (33.33%)
	Long (> 8 weeks)	6 (25%)

5.2 The Impact of AI-based feedback on enhancing students' academic writing performance

Table 3 displays the overall effect size and its heterogeneity. A small effect is indicated when g is less than 0.2, a medium effect is suggested when g is between 0.2 and 0.8, and a major effect is indicated when g is greater than 0.8 (Hedges, 1981). Therefore, an effect size of 1.298 suggests that it is substantial enough to be considered educationally significant, as it points to a high positive effect on improvement in English language writing performance that can have a practical impact on educational practices and enhance the overall learning experiences.

5.3 The effect of moderator variables on AI-based feedback

5.3.1 Effect sizes of moderator variables

To investigate the variables that might influence how well AI-based feedback tools enhance students' writing performance, the moderator variable results were presented using a mixed-effects model. The moderator variables' distribution and effect sizes are displayed in Table 7.

Table 7. Results of Moderator Variable Analysis

Moderator variables	n	Hedges' g	Standard error	Lower level	Upper level	Z	Q_b	p
English language proficiency							5.938	0.015
Intermediate	19	1.552***	0.341	0.884	2.219	4.555		
Advanced	5	0.573**	0.213	0.156	0.990	2.694		
Implementation setting							4.275	0.039
Classroom	17	0.785***	0.150	0.490	1.079	5.219		
Blended	7	2.855**	0.990	0.915	4.795	2.884		
Duration of intervention							0.989	0.610
Short	10	1.256**	0.417	0.438	2.073	3.011		
Medium	8	1.076***	0.269	0.549	1.603	4.004		
Long	6	1.765**	0.649	0.493	3.037	2.721		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

5.3.2 English language proficiency level

The moderator analysis in Table 7 shows that English language proficiency significantly influenced the effectiveness of AI-based feedback tools in students' writing performance ($Q_b = 5.938$, $p = 0.015 < 0.05$). For students with intermediate proficiency, the effect size was very large (Hedges' $g = 1.552$), indicating a high effect on writing performance. In contrast, students with advanced proficiency showed a smaller but still meaningful improvement ($g = 0.573$), corresponding to a medium effect. This pattern suggests that AI-based feedback tools are particularly beneficial for students who are still developing their writing skills, while more proficient learners experience comparatively moderate gains.

5.3.3 Implementation setting

The findings for the implementation setting (See Table 7) also revealed a significant moderating effect ($Q_b = 4.275$, $p = 0.039 < 0.05$). In classroom settings, the effect size was $g = 0.785$, which can be interpreted as a medium-to-high effect on students' writing performance. However, in blended learning environments, the effect size was substantially larger ($g = 2.855$), indicating a very high effect. This suggests that combining AI-based feedback with flexible or technology-enhanced learning contexts maximizes its impact, possibly due to increased opportunities for interaction and iterative revision.

5.3.4 Duration of intervention

Regarding the duration of intervention, Table 7 shows that all duration categories had a high effect on writing performance. Short-term interventions ($g = 1.256$), medium-term interventions ($g = 1.076$), and long-term interventions ($g = 1.765$) all demonstrated strong effects. However, the value of Q_b ($Q_b = 0.989$, $p = 0.610 > 0.05$) indicates that there was no statistically significant difference in the moderating effect of duration on the relationship between AI-based feedback tools and enhancing writing performance. These findings show that AI-based feedback tools are effective regardless of how long they are implemented, suggesting that even shorter interventions can yield substantial improvements in students' writing.

6. Discussion

This current systematic review and meta-analysis have studied the effectiveness of using AI-based feedback in writing performance. Besides, this study also focused on the moderating role of three factors in relation to AI-based feedback application – English language proficiency, implementation setting, and intervention duration. The results obtained from this study suggest that there is substantial evidence that AI-based feedback tools had a very high impact on students' academic writing performance ($g=1.298$), which further verifies their increasing importance in language teaching settings. The results found in this meta-analysis are consistent with some existing empirical studies in which AI feedback tool applications were found to improve writing quality by enhancing accuracy and coherence in writing (Chan et al., 2024; Wiboolyasarin et al., 2024).

It should be noted that the results of the first research question show that the implementation of AI feedback tools is most frequent in classroom environments, and there is also some evidence of using blended learning models. The significance of the role of an instructor in the learning process is highlighted despite the presence of new technologies. Some papers presented in the review (for example, Mahdi et al., 2025; Mekheimer, 2025; Sari & Han, 2024) state that AI tools are considered auxiliary means in the learning process. This finding is aligned with the outcomes obtained previously and shows that cooperation between human beings and AI contributes to the provision of comprehensive feedback because it combines the benefits of both AI and teachers' experience (Chen & Gong, 2025; Al Ghaithi & Behforouz, 2025). It also appears that using blended learning models allows for applying AI tools more creatively, as they can contribute to providing continuous feedback beyond classroom borders (Sawangwan, 2024; Wale & Kassahun, 2024).

Regarding the second research question, the high effect size reported in this research clearly indicates that the use of AI-based feedback has a strong positive influence on the effectiveness of students' writing. As mentioned earlier, previous findings indicate that AI technologies allow for enhancing students' writing by means of giving immediate and specific feedback regarding grammar, vocabulary, and organization (Escalante et al., 2023; Gayed et al., 2022). In addition to previously cited works, more recent publications considered within the scope of this review (for example, Ali et al., 2025; Mahapatra, 2024) demonstrate that the use of AI-based feedback results in an iterative process of revising work and in a greater involvement in writing. Although there was a high effect size in general, some of the works found within the current review suggest that the effectiveness of feedback depends on the nature of such feedback. Namely, AI can be used effectively to solve low-level issues such as lexical errors or grammar problems, while high-level features of writing, including argumentation or critical thinking, require human involvement (Bašić et al., 2023; Jamshed et al., 2025).

As for the third research question, moderator analysis has shown that English language proficiency greatly affects the effectiveness of AI-based feedback. Intermediate level students were characterized by a high effect, while those on the advanced level had a medium effect. This result resonates with the findings of several authors (e.g., Wang & Han, 2022; Xiao, 2024), according to whom the use of AI tools is much more beneficial to those learners whose language skills need further improvement. The use of explicit correction and scaffolding offered by the tool contributes greatly to the noticing and error correction process among intermediate-level students. Advanced students, on the other hand, may not benefit as much from the AI tool as their needs are broader and include not only language accuracy but also rhetoric and analysis, which cannot always be offered by such software (Alnemrat et al., 2025; Mahdi et al., 2025).

In addition, the implementation setting was identified as an important moderator that influenced the effectiveness of AI-based feedback. Specifically, the analysis revealed a very large effect of AI feedback in blended learning environments, while there was a medium to high effect on traditional classroom settings. This is consistent with the findings of a study conducted recently, which concluded that blended learning environments contribute to improved effects of AI-based feedback through increased flexibility, more practice opportunities, and continuous availability of feedback (Biju et al., 2024; Sari & Lestari, 2025). More specifically, learners can participate in the revision process more actively by using feedback that can be provided both synchronously and asynchronously.

On the other hand, the duration of implementation does not appear to moderate the effectiveness of AI-based feedback tools. While this result may seem counterintuitive, it should be noted that in all cases considered here, the duration had an extremely significant impact on the level of achievement, and any implementation time appeared to provide good results. Therefore, this result could be explained by the fact that, irrespective of the duration of intervention, the usage of AI feedback tools helps produce good academic results. It is interesting to note that there is some research that corroborates this idea, including Tedjo (2022) and Wale and Kassahun (2024). This result may have several important practical applications since it will help educators understand that the immediate application of AI tools is beneficial if learners receive timely and sufficient feedback.

In conclusion, it is important to emphasize that the results obtained during the current meta-analysis confirm the general efficacy of AI feedback tools in academic writing. By analyzing multiple studies in a single paper, one can form an overall picture of how the tool in question is used across settings and which factors may affect its effectiveness.

7. Implications

Findings from this study have major implications for teachers, curriculum designers, and policymakers in the domain of language education. Considering the overall effect of using AI-based feedback tools on writing performance, there is an opportunity for educational institutions to incorporate such tools in the process of writing instruction so as to promote positive academic results

for students. It is suggested that teachers use AI-based tools as supplementary tools that allow giving instant and personalized feedback, especially to intermediate-level learners who are positively influenced by this method. The high effectiveness of using these tools in the blended mode of instruction also suggests that incorporating AI tools in flexible teaching models is a way to maximize their influence due to more autonomy.

8. Limitations

Despite its strengths, this study also has some limitations that are worth mentioning. First, differences in research design, sample size, and measures used to evaluate participants' writing abilities might have led to the heterogeneity of the results obtained. Second, most of the analyzed articles deal with learners of an intermediate proficiency level. Thus, the results are difficult to generalize to those learners who are either at the beginner or advanced level. Moreover, the analysis of the selected moderators, including English proficiency level, implementation context, and the duration of intervention, does not provide insights into other important aspects that influence the use of artificial intelligence in writing instruction. These aspects include types of AI tools, the quality of feedback provided by the tool, and students' involvement. Therefore, more research needs to be done to cover all these limitations.

8. Conclusion

The current paper confirms that the introduction of AI-based feedback systems is highly beneficial for improving the students' academic writing skills. As the findings indicate, these solutions are particularly effective when used by intermediate learners or in a blended environment. At the same time, the performance of AI-based tools remains high across various intervention periods. Thus, such systems may serve as an efficient and scalable solution to enhancing writing skills in EFL and ESL settings. Yet, several points need to be addressed when applying AI systems successfully. Namely, more moderators should be identified, and the ways to develop higher-level writing skills through AI-based systems should be discussed.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

ORCID iD: 0000-0003-0942-4232

Publisher's Note: All claims expressed in this article are solely those of the author and do not necessarily represent those of her affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Al Ghaithi, A., & Behforouz, B. (2025). Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance. *Knowledge Management & E-Learning*, 17(2), 206–224. <https://doi.org/10.34105/j.kmel.2025.17.009>
- [2] Alanazi, S. M., Elmotri, B., Khamis, G. S., & Darem, A. A. (2025). Assessing the efficacy of ChatGPT's automated corrective feedback in enhancing students' writing proficiency. *International Journal of Advanced and Applied Sciences*, 12(2), 205–214. <https://doi.org/10.21833/ijaas.2025.02.022>
- [3] Ali, Y.A., Ali, M.M., Alharbi, W., Hamid, S., & Abbas, S. (2025). The Influence of ChatGPT on English Grammar Among Saudi ESL Learners. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence*:358-371. <https://dx.doi.org/10.24093/awej/AI.21>
- [4] Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1614673>
- [5] Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, 30, 100322. <https://doi.org/10.1016/j.edurev.2020.100322>
- [6] Bai, X., & Nordin, N. R. M. (2025). Human-AI collaborative feedback in improving EFL writing performance: An analysis based on natural language processing technology. *Eurasian Journal of Applied Linguistics*, 11(1), 1–19. <https://doi.org/10.32601/ejal.11101>
- [7] Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>
- [8] Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities and social sciences communications*, 10(1), 1–5. <https://doi.org/10.1057/s41599-023-02269-7>
- [9] Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: An exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14(1), 45. <https://doi.org/10.1186/s40468-024-00322-z>
- [10] Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- [11] Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes. *Contemporary Educational Technology*, 16(4), ep541. <https://doi.org/10.30935/cedtech/15607>

- [12] Chang, T.-S., Li, Y., Huang, H.-W., & Whitfield, B. (2021). Exploring EFL students' writing performance and their acceptance of AI-based automated writing feedback. In *Proceedings of the 2nd International Conference on Education Development and Studies* (pp. 31–35). ACM. <https://doi.org/10.1145/3459043.3459065>
- [13] 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*, 15(2), 141. <https://doi.org/10.3390/educsci15020141>
- [14] Cooper, H., Hedges, L. V., Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. https://muse.jhu.edu/pub/207/edited_volume/book/65827
- [15] Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- [16] Du, J., & Nordin, N. R. M. (2026). The impact of the iWrite automated writing evaluation system on university EFL students' writing performance and writing anxiety. *Education Sciences*, 16(3), 411. <https://doi.org/10.3390/educsci16030411>
- [17] Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634.
- [18] EKİZOĞLU, M., & DEMİR, A. N. (2025). AI-Assisted Writing Feedback for Enhancing Secondary Students' Writing Skills: An Experimental Study. *Research Square*. <https://doi.org/10.21203/rs.3.rs-6430737/v1>
- [19] 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), 57. <https://doi.org/10.1186/s41239-023-00425-2>
- [20] Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. <https://doi.org/10.3389/frai.2023.1162454>
- [21] Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100055>
- [22] Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- [23] Higgins, J. P. T. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- [24] Jamshed, M., Albedah, F., Ansari, M.S., Banu, S. (2025). Assessing the Efficacy of AI-Driven Corrective Feedback via WhatsApp Application to Improve ESL Learners' Writing Skills: An Experimental Study. *International Journal of Interactive Mobile Technologies (IJIM)*, 19(7), 190–205. <https://doi.org/10.3991/ijim.v19i07.52709>
- [25] Lee, I. (2020). *Classroom writing assessment and feedback in L2 school contexts*. Springer.
- [26] Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- [27] Mahdi, H. S., Milad, M., & Saeed, A. T. (2025). Evaluating the Effectiveness of AI Tools Across the Essay Writing Process: A Comparative Study of Pre-Writing, Drafting, and Post-Writing Stages. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 15(1), 1–20. <http://doi.org/10.4018/IJCALLT.391354>
- [28] Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality. *Discover Education*, 4(1), 170. <https://doi.org/10.1007/s44217-025-00602-7>
- [29] Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906.
- [30] Park, Y. & Doo, M. (2024). Role of AI in Blended Learning: A Systematic Literature Review. *International Review of Research in Open and Distributed Learning*, 25(1), 164–196. <https://doi.org/10.19173/irrodl.v25i1.7566>
- [31] Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674. <https://doi.org/10.1080/09588221.2018.1428994>
- [32] Sari, E., & Han, T. (2024). The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *Journal of Computer Assisted Learning*, 40(5), 2065–2080. <https://doi.org/10.1111/jcal.13004>
- [33] Sari, N., & Lestari, Y. (2025). The effect of artificial intelligence-based feedback on improving the grammatical accuracy of EFL students. *NeuroLingua: Journal of Cognitive, Technological, and Cultural Language Learning*, 1(1), 40–48.
- [34] Sawangwan, S. (2024). ChatGPT vs teacher roles in developing EFL writing. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 14(1), 1–21. <https://doi.org/10.4018/IJCALLT.361235>
- [35] Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62(1), 97–128.
- [36] 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>

- [37] Tedjo, E. W. (2022). Providing online feedback using an AI proof-reading tool to enhance student surface-level writing skills. *EBONY: Journal of English Language Teaching Linguistics and Literature*, 2(1), 37-45.
- [38] Thi, X. H. N., Thien, H. V. H., Vuong, K. N., & Nguyen, T. T. (2025). Enhancing writing skills through AI-powered tools: perceived benefits and challenges among Vietnamese EFL students. *Discover Education*, 4(1), 472. <https://doi.org/10.1007/s44217-025-00905-9>
- [39] Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences*, 15(2), 232. <https://doi.org/10.3390/educsci15020232>
- [40] Vo, T. L., & Ho, T. M.P. (2026). A Quasi-Experimental Study on Enhancing EFL Learners' Grammar Precision in Writing Skills under Supported-ChatGPT Approach at a University in a Remote Region of Vietnam. *Arab World English Journal, Special Issue on Artificial Intelligence*, (3):361-375. <https://dx.doi.org/10.24093/awej/A3.23>
- [41] Wale, B. D., & Kassahun, Y. F. (2024). The transformative power of AI writing technologies: Enhancing EFL writing instruction through the integrative use of Writerly and Google Docs. *Human Behavior and Emerging Technologies*, 2024(1), 1-15. <https://doi.org/10.1155/2024/9221377>
- [42] Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. *Frontiers in psychology*, 13, 1-12. <https://doi.org/10.3389/fpsyg.2022.909802>
- [43] Wiboolyasarín, W., Wiboolyasarín, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, 6, 100228. <https://doi.org/10.1016/j.caeai.2024.100228>
- [44] Xiao, Q. (2024). ChatGPT as an Artificial Intelligence (AI) Writing Assistant for EFL Learners: An Exploratory Study of its Effects on English writing Proficiency. *Mulfimodal Teaching and Learning Methods*, 51. <https://doi.org/10.1145/3664934.3664946>
- [45] Xue, Y. (2024). Towards automated writing evaluation: A comprehensive review with bibliometric, scientometric, and meta-analytic approaches. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12596-0>
- [46] Zhai, N., & Ma, X. (2022). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, 61(4), 875–900. <https://doi.org/10.1177/07356331221127300>
- [47] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1), 8812542. <https://doi.org/10.1155/2021/8812542>
- [48] Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences*, 15(5), 600. <https://doi.org/10.3390/bs15050600>
- [49] Zhang, Y., Jalaluddin, I. B., Mamat, R., & Zhao, T. (2025). The effects of automated writing evaluation (AWE)+ human feedback on the quality of argumentative writing by Chinese EFL learners. *Journal of Language Teaching and Research*, 16(3), 899-910. <https://doi.org/10.17507/jltr.1603.20>
- [50] Zhang, Z. V., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, 100586.