International Journal of English Language and Literature Studies
ISSN(e): 2306-0646
ISSN(p): 2306-9910
DOI: 10.55493/5019.v14i1.5328
Vol. 14, No. 1, 17-30.
© 2025 AESS Publications. All Rights Reserved.
URL: www.aessweb.com

Effectiveness and moderating factors of computer-mediated feedback in L2 writing: A meta-analysis

Shinjae Park

College of General Education, Kookmin University, Seoul 02707, Korea.
Email: muhando@kookmin.ac.kr

Article History
Received: 2 January 2025
Revised: 14 March 2025
Accepted: 27 March 2025
Published: 4 April 2025

Keywords
ChatGPT
Computer-mediated feedback
Feedback
L2 writing performance
Meta-analysis
Moderator analysis.

ABSTRACT

Computer-mediated feedback (CMF) for writing has garnered growing attention among practitioners and researchers. However, comprehensive meta-analyses in English as a foreign language (EFL) and English as second language (ESL) contexts remain scarce. This study synthesized 19 valid effect sizes from 26 experimental studies conducted between 2021 and 2024 to evaluate the effectiveness of CMF in L2 learners' writing performance and determine the moderating factors influencing its variability. The results revealed CMF large overall effect ($g = 1.602$) with significant variability across studies. A moderator analysis identified meaningful differences based on feedback source, feedback timing, learner proficiency, and task characteristics. Automated and immediate feedback demonstrated the strongest impact ($g = 0.937$ and $g = 0.875$) highlighting their importance in facilitating real-time corrections. Intermediate learners benefited the most ($g = 0.789$) while advanced and beginner learners showed comparatively smaller effects. Greater improvements were observed in academic writing tasks other than argumentative writing tasks. Additionally, integrating educational software alongside CMF implementation at the university level appeared to enhance writing proficiency. These findings suggest that CMF significantly contributes to L2 writing development, particularly in structured academic environments.

Contribution/ Originality: This study examines the effectiveness of CMF in L2 writing through a meta-analysis of 19 effect sizes from recent studies conducted between 2021 and 2024. Exploring key moderating factors—proficiency, task type, and feedback timing provides a clearer picture of how CMF works. Additionally, it compares the impact of different AI-driven feedback tools.

1. INTRODUCTION

The integration of innovative feedback technologies has become increasingly prevalent in L2 writing instruction, reflecting a broader shift toward digital learning environments (Escalante, Pack, & Barrett, 2023; Guo & Wang, 2024; Mizumoto & Eguchi, 2023; Teng, 2024). Among these technologies, computer-mediated feedback (CMF) has been recognized in reducing L2 learners' writing apprehension and fostering metacognitive engagement. Previous studies indicate that CMF contributes to improvements in linguistic accuracy and overall composition quality (Barrot, 2023; Fitriana & Nurazni, 2022; Zhang & Zhang, 2024). Furthermore, CMF's structured and immediate feedback mechanisms facilitate a more iterative revision process, often impractical in traditional classroom-based feedback settings (Howell, Perez, & Abraham, 2021; Thi & Nikolov, 2022).

© 2025 AESS Publications. All Rights Reserved.

The existing findings exhibit notable inconsistencies regarding its pedagogical effectiveness while prior research has examined various aspects of technology-enhanced feedback in L2 writing instruction (Liu, Hou, Tu, Wang, & Hwang, 2023; Rahimi & Fathi, 2022; Yao, Wang, & Yang, 2021). A systematic meta-analysis can consolidate these findings and identify key factors influencing CMF outcomes across different instructional settings. To date, only a few meta-analyses (e.g., (Fleckenstein, Liebenow, & Meyer, 2023; Lv, Ren, & Xie, 2021; Ngo, Chen, & Lai, 2024; Seyyedrezaei, Amiryousefi, Gimeno-Sanz, & Tavakoli, 2024)) have systematically examined the impact of online feedback on L2 writing. Lv et al. (2021) synthesized findings from 17 studies published between 2000 and 2021, concluding that computer-mediated writing instruction significantly improves writing quality compared to non-technological methods. Furthermore, they identified moderating factors such as feedback source (e.g., online teacher feedback and online automated feedback), task genre (e.g., argumentative essay and report), and teaching status (e.g., collaborative task and individual task) to be key variables affecting the effectiveness of these interventions. Seyyedrezaei et al. (2024) conducted a meta-analysis of 64 studies primarily published from 1990 to 2020. They found that the use of technology had a large positive effect on English as a second/foreign language (ESL/EFL) learners' writing performance ($g = 1.00$). Notably, the genre of writing and the type of technology were identified as significant moderator variables with a meta-regression analysis revealing their strong relationship with effect size.

Several gaps exist in the literature despite these contributions. For example, little is known about the relative effects of different types of educational technologies (e.g., Pigai, Grammarly, and ChatGPT) nor has there been sufficient investigation into how learner proficiency or education level (e.g., secondary school and university) may moderate the impact of CMF. Furthermore, research on CMF has grown considerably in recent years, particularly after 2020, a comprehensive and updated meta-analysis is needed to evaluate these findings and provide actionable insights for future research.

To address these gaps, the current study conducted a meta-analysis of 26 experimental studies published between 2021 and 2024, incorporating 19 valid effect sizes to examine the overall effectiveness of CMF in L2 writing performance. In addition, this meta-analysis investigated the moderating effects of learner proficiency, feedback timing, and task characteristics on the impact of CMF providing an updated synthesis to advance the understanding of how CMF can be optimized for L2 learners.

2. LITERATURE REVIEW

2.1. The Role of Technology in Writing Instruction

L2 writing is widely recognized as a complex skill that requires the integration of multiple linguistic and cognitive components. Research indicates that many learners struggle with acquiring proficiency in academic writing due to difficulties in vocabulary selection, grammatical accuracy and text organization (Alkhalaf, 2020; Ferris & Eckstein, 2020; Seyyedrezaei et al., 2024). These challenges are further compounded by classroom environments that may not effectively accommodate diverse learning preferences, potentially hindering student engagement and skill development (Ankawi, 2023).

Technological advancements have introduced new approaches to writing instruction, offering potential solutions to persistent challenges in L2 writing development. Writing-assistance tools such as Grammarly provide immediate, structured feedback on grammar, style, and clarity, facilitating independent learning beyond the constraints of traditional classroom instruction (Armanda, Nugraheni, Wulansari, & Imron, 2022). Studies suggest that integrating technology into feedback mechanisms may help reduce learner anxiety and foster a more supportive writing environment (Cui, Schunn, Gai, Jiang, & Wang, 2021; Taskıran & Goksel, 2022). Moreover, technology supports higher-order cognitive processes such as meta-cognitive evaluation and critical thinking, which are essential for effective writing (Waer, 2023; Zhai & Ma, 2022).

© 2025 AESS Publications. All Rights Reserved.

Empirical studies have demonstrated the efficacy of specific technological tools in enhancing L2 writing proficiency. For instance, Zhang and Zhang (2024) observed that Pigai-based instruction facilitated improvements in critical thinking and academic essay composition due to its extensive learning resources and emphasis on higher-order cognitive skills. Similarly, Waer (2023) reported that technology-supported feedback mechanisms not only enhanced grammatical accuracy and content quality but also contributed to increase learner motivation in writing tasks. Conflicting findings have emerged despite these promising results. Lv et al. (2021) noted that some learners struggle with online environments due to technical issues and distractions. Zhao and Yang (2023) observed that public formats, such as blogs may discourage participation due to privacy concerns. These discrepancies underscore the need for systematic reviews to clarify the overall effectiveness of technology in writing instruction.

Grammarly and Pigai are among the most frequently adopted feedback tools in L2 writing instruction, each offering distinct advantages for language learners. Grammarly is widely recognized for its ability to provide immediate corrections on grammar, style, and clarity, benefiting non-native English speakers by enhancing writing fluency and accuracy (Fitriana & Nurazni, 2022). In contrast, Pigai, which is primarily utilized in Chinese educational settings, employs an automated evaluation system that aligns with linguistic patterns characteristic of Chinese EFL learners. Its localized approach to error detection differentiates it from other automated writing feedback systems as it tailors suggestions to the specific needs of its user base (Zhang & Zhang, 2024).

In addition to these tools, platforms such as Criterion and Write and Improve provide structured feedback aligned with educational standards. Criterion is commonly used for test preparation, whereas Write & Improve, developed by Cambridge English emphasizes iterative feedback to help learners revise their drafts. Moodle, which is a learning management system integrates feedback capabilities that support both automated and teacher-mediated feedback. Finally, generative artificial intelligence (AI) tools, such as ChatGPT, have introduced conversational feedback mechanisms, offering learners a unique way to interact with the program, clarify doubts, and explore alternative writing strategies. However, concerns regarding accuracy and ethical implications remain (Guo & Wang, 2024).

2.2. Previous Meta-Analyses

Recent meta-analyses have played a crucial role in consolidating research on CMF, a technology-driven approach to writing instruction that emphasizes iterative feedback processes. Unlike broader concepts such as technology-enhanced language learning (TELL), CMF is specifically designed to enhance feedback mechanisms through digital platforms, including automated assessment tools, teacher-mediated feedback systems, and peer-review environments (Li, 2023). Its targeted nature allows for more timely and actionable feedback addressing key challenges in L2 writing instruction.

Li (2023) synthesized data from 28 studies on CMF in L2 vocabulary learning, covering research published between 2000 and 2022. This analysis provided valuable insights into how CMF facilitates vocabulary acquisition by integrating automated feedback mechanisms with learner engagement strategies. The study found CMF to have a large positive effect ($g = 0.853$) on adult learners' vocabulary learning, highlighting language distance (e.g., the degree of similarity or difference between the writing systems of two languages) as a significant moderating factor. Seyyedrezaei et al. (2024) extended this research by analyzing 64 studies primarily from 1990 to 2020 and reported large effect size ($g = 1.00$) for CMF in ESL/EFL writing contexts. They identified the genre of writing and the type of feedback as statistically significant moderators with automated feedback showing particular promise for improving grammatical accuracy and coherence.

However, these studies have notable limitations. Li (2023) focused on learners' vocabulary learning while Seyyedrezaei et al. (2024) focused on pre-2020 studies, excluding the impact of newer technologies developed during the COVID-19 pandemic—a period of rapid digital transformation in education (Lim & Richardson, 2021).

© 2025 AESS Publications. All Rights Reserved.

Previous analyses have not fully examined the interaction between specific feedback mechanisms, such as immediate versus delayed feedback and variables like learner proficiency and task complexity.

This meta-analysis synthesizes findings from studies published between 2021 and 2024 to assess the impact of CMF on L2 writing performance based on previous research. It further investigates how moderating variables—including feedback timing, learner proficiency, task characteristics and the type of feedback software (e.g., Grammarly, Criterion, Pigai and ChatGPT) influence writing outcomes. This study contributes to refining instructional practices in L2 writing by identifying the conditions under which CMF tools yield the most substantial benefits. Additionally, it examines global and temporal patterns by exploring how preferences for feedback software vary based on researchers' national affiliations and the publication year, thereby providing insights into emerging trends in the field.

2.3. Research Questions

This study investigates the impact of CMF on L2 writing proficiency by addressing the following research questions:

1. How effective is CMF in improving L2 writing performance?
2. To what extent do moderating factors, such as feedback timing, learner proficiency, task characteristics, and f eedback software preferences (e.g., Grammarly, Criterion, Pigai, and ChatGPT) influence the effectiveness of CMF?

3. METHOD

3.1. Inclusion and Exclusion Criteria

The inclusion of studies in this meta-analysis was based on the following criteria which ensured the relevance and methodological rigor of the selected research:

1. The study employed systematic quantitative data suitable for meta-analysis, published between 2021 and 2024.
2. The research investigated the effects of online, automated, or electronic feedback on ESL/EFL writing performance.
3. The study clearly defined independent variables related to online feedback and its impact on ESL/EFL writing.
4. Participants were instructed in either a second or foreign language.

Studies were excluded from the analysis based on the following criteria:

1. The study presented quantitative data but lacked descriptive statistics.
2. The research did not focus on writing quality but rather explored students' attitudes or perspectives.
3. The study was published in a language other than English.

A comprehensive search was conducted using Google Scholar to identify relevant studies. The search incorporated terms related to automated and computer-based writing feedback, including "automated writing evaluation," "computer-based essay feedback," and related variations. From the initial search, 39 articles were retrieved. Each article was screened based on its title and abstract leading to the exclusion of 26 articles that did not align with the research focus. A more detailed examination of the methodology and results sections resulted in the further exclusion of nine studies due to insufficient statistical data for calculating effect sizes. Ultimately, 19 quantitative studies met the inclusion criteria for this meta-analysis (see Appendix 1).

3.2. Coding Study Features

A coding scheme was established to extract relevant study characteristics based on variables commonly employed in previous meta-analyses of applied linguistics (Norris & Ortega, 2000). This scheme was further refined

© 2025 AESS Publications. All Rights Reserved.

in accordance with the recommendations of Li (2023). The characteristics coded included potential moderating variables that could influence the effects of online feedback on writing, such as sample characteristics, research methodology, and effect size data.

Sample characteristics encompassed factors such as education levels, majors, and research settings. Research method variables included aspects such as feedback source, study design, task type, and task setting. Effect size information comprised data points such as total sample size, treatment and control group details, and pre- or post-test differences. In this study, these variables were classified into population data and treatment data as shown in the table to investigate the effects of CMF on L2 writing. The coding scheme is described in detail in Appendix 2.

3.3 Research Instrument and Data Selection

A meta-analytic approach was chosen as the research instrument to systematically synthesize findings from multiple empirical studies, offering a quantitative and generalizable understanding of CMF effects on L2 writing. This method provides several advantages over individual experimental studies. First, meta-analysis mitigates biases associated with single-study variability and enhances statistical power, producing a more reliable estimate of treatment effects by aggregating multiple effect sizes. Second, it enables the identification of moderating factors such as learner proficiency, feedback source, and feedback timing in which individual studies often lack the statistical scope to analyze comprehensively. Third, prior meta-analyses in applied linguistics (e.g., (Fleckenstein et al., 2023; Li, 2023; Ngo et al., 2024; Norris & Ortega, 2000)) have demonstrated the robustness of this approach in summarizing intervention effects in L2 acquisition research.

The statistical methods used in this study were employed to assess the validity of selecting meta-analysis as the research approach. The weighted mean effect size, calculated using Hedges' g was used to evaluate the impact of CMF on L2 writing. Additionally, a funnel plot and Rosenthal's fail-safe N were utilized to assess publication bias, helping to examine the reliability of the study's findings and consider the potential for selective reporting.

4. RESULTS

4.1. Overall Effect Size

The meta-analysis included 19 studies ($k = 19$) with 13,589 participants. The weighted mean effect size, calculated using Hedges' g was 1.602, reflecting a moderate to large overall effect. The 95% confidence interval (CI) for this effect size ranged narrowly from 1.585 to 1.619, indicating considerable precision and consistency in the estimates. Detailed results for each study, including effect sizes, CIs, and significance levels are presented in Appendix 3.

The heterogeneity analysis showed a Q statistic of 2849.86 ($df = 18$, $p < 0.001$) and an I-squared (I^2) value of 99.37%, indicating very high heterogeneity among the included studies. The tau-squared (τ^2) value of 0.207 confirmed substantial variability in true effect sizes beyond sampling error. τ^2 measures the absolute variance in true effect sizes while I^2 reflects the proportion of observed variance due to heterogeneity rather than sampling error (Borenstein, Hedges, Higgins, & Rothstein, 2021). An I^2 value above 75% signals significant heterogeneity, which requires further investigation (Borenstein et al., 2021; Chandler, Cumpston, Li, Page, & Welch, 2019). In the current study, despite this heterogeneity, the recalculated weighted mean effect size and its narrow CI indicate a stable and significant treatment effect. To better understand why results vary across studies, researchers should examine various factors, such as participant characteristics or task types that may influence the outcomes. Table 1 summarizes the overall effect size results including the number of studies and participants, the weighted mean effect size (Hedges' g), the 95% confidence interval and heterogeneity statistics.

© 2025 AESS Publications. All Rights Reserved.

Table 1. Overall effect size results.

	95% CI				Heterogeneity			
k	N	g	LL	UL	Q	df	I^2	τ^2
19	13,589	1.602	1.585	1.619	2849.86	18	99.37	0.207

Note: k = Number of effect sizes; N = The total number of participants; g = Hedges' g (Weighted mean effect size); LL = Lower limit of the 95% CI; UL = Upper limit of the 95% CI; Q = Cochran's Q statistic for heterogeneity; I^2 = Percentage of observed variance due to heterogeneity; τ^2 = Estimate of between-study variance ($p < 0.001$).

Publication bias was assessed using a funnel plot that displayed a slightly asymmetrical distribution of effect sizes around the mean (g = 1.602). The asymmetry was more pronounced in studies with higher standard errors, suggesting that smaller studies with larger effect sizes may be overrepresented. This pattern indicates a potential for publication bias; namely, studies with significant results are more likely to be published. Figure 1 visualizes this pattern illustrating the distribution of effect sizes (Hedges' g) relative to their standard errors. The red dashed line represents the mean effect size (g = 1.602) serving as a reference for evaluating the spread of effect sizes.

Rosenthal's fail-safe N was calculated to further evaluate this bias. The value obtained was 65 indicating that 65 null-effect studies would be required to negate the observed effect size (Rosenthal, 1979). Although this result suggests that the current findings are robust, the funnel plot's visible asymmetry and the high mean effect size warrant caution when interpreting the overall results. The findings suggest that further statistical tests, such as Egger's regression are necessary to assess the potential presence of publication bias and its impact on effect size estimates. The results offer initial insights into the variability of effect sizes and the factors that may influence these variations while there are limitations to this study.

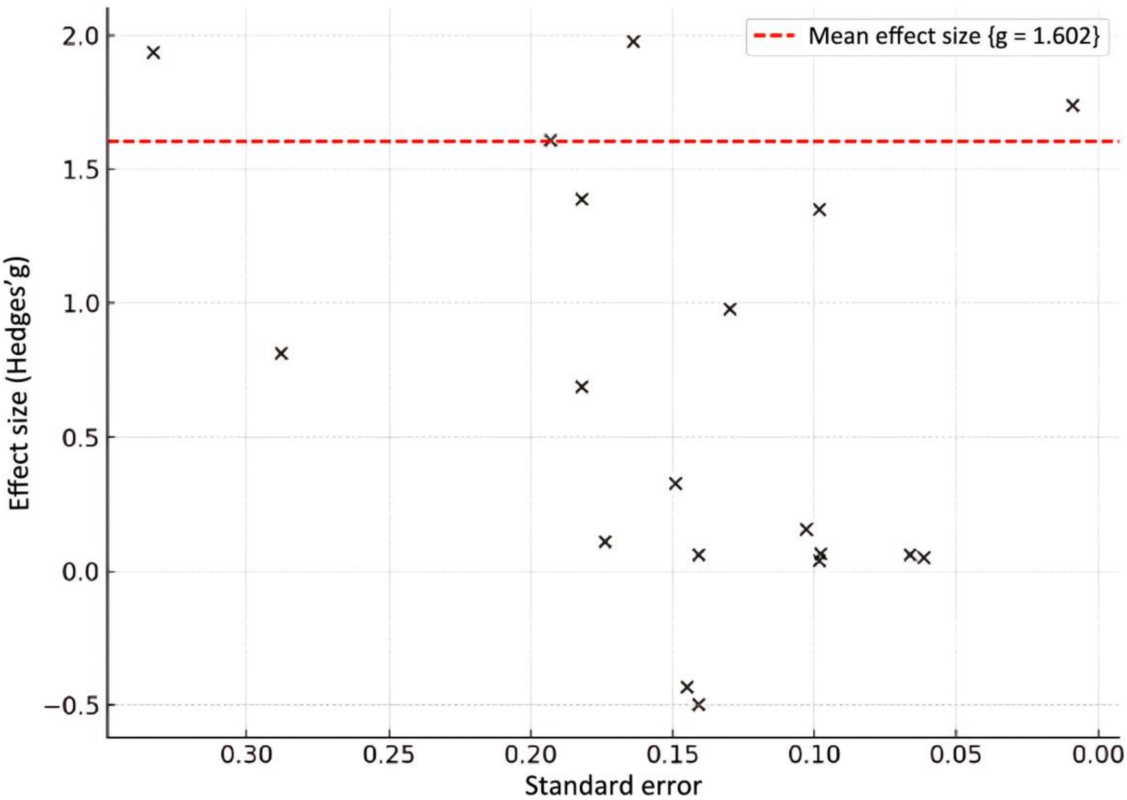

Figure 1. Funnel plot of effect sizes (Hedges's g) against the standard error.

4.2. Moderator Analysis

Table 2 summarizes the moderator effects on the overall effect sizes (Hedges' g) observed in the meta-analysis. The analysis examined how various study characteristics—namely, education level, feedback source, feedback

© 2025 AESS Publications. All Rights Reserved.

timing, learner proficiency, software use, and task genre influence the effectiveness of feedback in ESL/EFL writing.

Table 2. Summary of moderator variables.

Moderators	Category	k	g	LL	UL	Z_value	p
Education level	P	2	0.254	0.095	0.413	3.136	0.0017
	U	17	0.74	0.326	1.155	3.5	0.0005
Feedback source	AF	8	0.937	0.263	1.611	2.723	0.0065
	M	5	0.704	0.234	1.174	2.934	0.0033
	PF	6	0.347	-0.335	1.028	0.997	0.3188
Feedback timing	D	9	0.482	-0.021	0.986	1.876	0.0607
	I	10	0.875	0.324	1.427	3.113	0.0019
Proficiency	A	4	0.421	-0.659	1.501	0.764	0.4449
	B	3	0.646	-0.128	1.42	1.636	0.1018
	I	12	0.789	0.319	1.26	3.287	0.001
Software name	ChatGPT	3	0.433	-0.887	1.753	0.643	0.5202
	Pigai	5	0.568	-0.341	1.477	1.224	0.221
	None	4	0.196	0.101	0.29	4.057	0.00005
	Other	7	1.168	0.67	1.665	4.601	0.000004
Software type	E	15	0.697	0.267	1.127	3.176	0.0015
	G	4	0.661	-0.233	1.554	1.45	0.1471
Task genre	A	5	0.485	-0.278	1.247	1.246	0.2128
	AC	7	1.057	0.426	1.689	3.282	0.001
	M	7	0.467	-0.116	1.05	1.57	0.1164

Note: Education level: U (University) and P (Pre-university); proficiency: B (Basic), I (Intermediate), and A (Advanced); feedback source: PF (Peer), AF (Automated), TF (Teacher), and M (Mixed); feedback timing: I (Immediate) and D (Delayed); task genre: A (Argumentative writing), AC (Academic writing), and M (Mixed genre); software type: G (General purposes) and E (Educational purposes).

Regarding the education level, it was revealed that university-level studies (g = 0.74 and p = 0.0005) showed a stronger and more statistically significant effect compared to primary-level studies (g = 0.254 and p = 0.0017). This suggests that feedback may have a greater impact on university learners due to their advanced cognitive abilities and deeper engagement with academic tasks.

In terms of feedback source, automated feedback (g = 0.937 and p = 0.0065) and mixed feedback (g = 0.704 and p = 0.0033) demonstrated significant positive effects. In contrast, peer feedback (g = 0.347 and p = 0.3188) was not statistically significant indicating that technology-mediated feedback may offer advantages over peer-provided feedback in improving writing quality.

Furthermore, feedback timing played a notable role in improving learners' writing skills. Immediate feedback (g = 0.875 and p = 0.0019) had a stronger and more statistically significant impact compared to delayed feedback (g = 0.482 and p = 0.0607) underscoring the importance of timely responses in enhancing learners' writing performance.

In terms of proficiency levels, intermediate learners (g = 0.789 and p = 0.001) benefited the most from feedback. Advanced (g = 0.421 and p = 0.4449) and basic learners (g = 0.646 and p = 0.1018) showed weaker and non-significant effects suggesting that learners at an intermediate stage are more receptive to feedback as they transition towards higher proficiency levels.

In terms of software use, studies utilizing "other" software tools (g = 1.168 and p = 0.000004) and studies that did not involve specific software (g = 0.196 and p = 0.00005) showed statistically significant effects. However, ChatGPT (g = 0.433 and p = 0.5202) and Pigai (g = 0.568 and p = 0.221) demonstrated non-significant effects. This indicates variability in the effectiveness of tools influenced by the context and implementation of feedback systems.

Finally, the task genre analysis showed that feedback yielded significant benefits with regard to academic writing tasks (g = 1.057 and p = 0.001) whereas argumentative tasks (g = 0.485 and p = 0.2128) and mixed tasks

© 2025 AESS Publications. All Rights Reserved.

($g = 0.467$ and $p = 0.1164$) did not show statistically significant effects. Therefore, feedback approaches must be tailored to specific task types for optimal effectiveness.

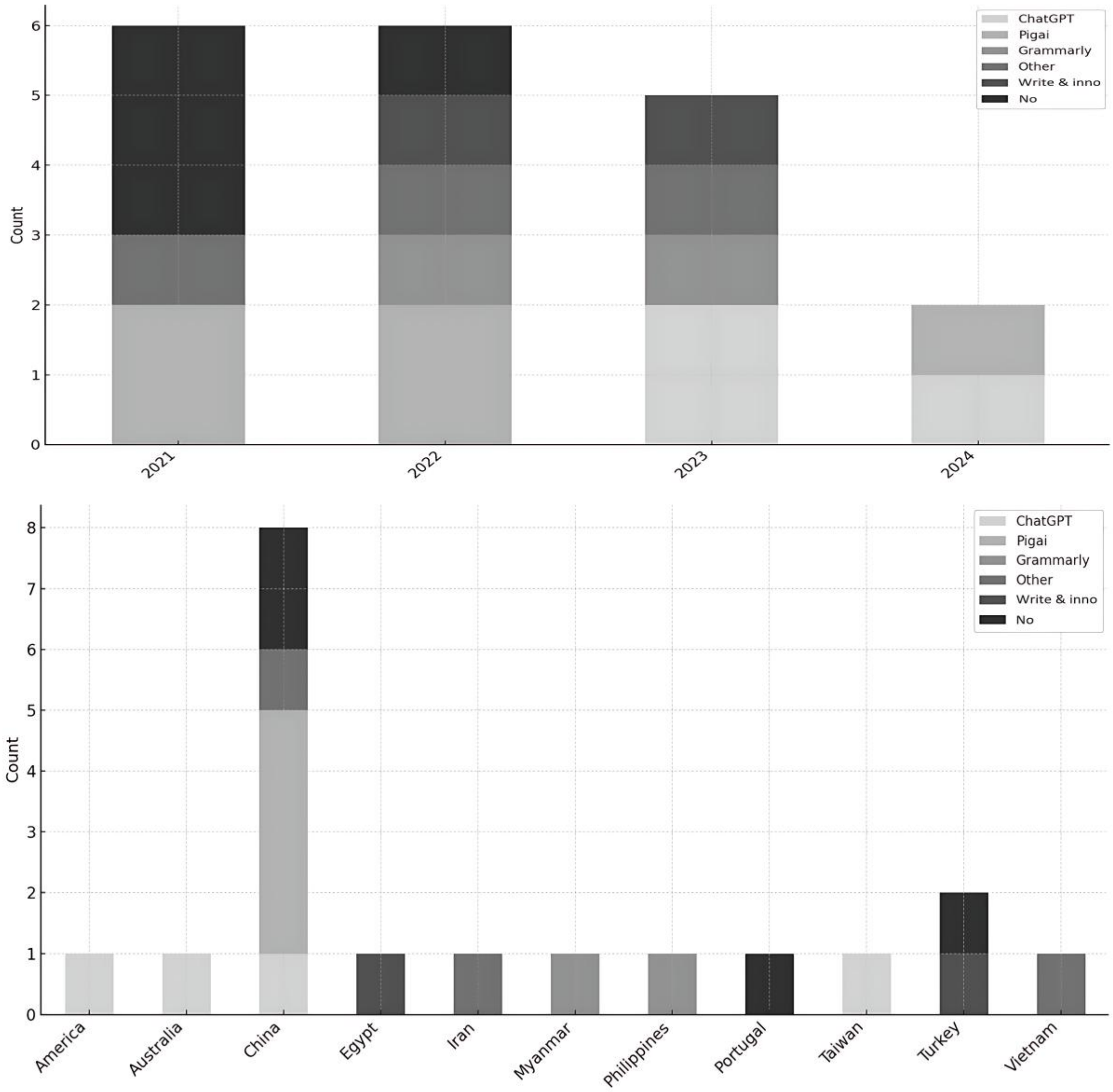

Figure 2. Software usage patterns by year and nationality.

Although the dataset is limited, several notable trends emerge. Since 2021, Pigai has remained consistently popular, particularly in China where 42.11% (eight out of 19) of all data points were generated. Furthermore, since 2023, ChatGPT has emerged as a frequently used tool, accounting for 42.86% (three out of seven) of software usage in the most recent data. Figure 2 shows these trends illustrating the distribution of software usage by year (top panel) and authors' nationalities (bottom panel). The data reveal a distinct regional pattern with Pigai being predominantly used in China while other tools, such as ChatGPT show more varied adoption across different countries.

Overall, the findings emphasize that feedback timing, feedback source and education level significantly influence the effectiveness of feedback in ESL/EFL writing. Immediate feedback, automated feedback, and feedback provided at the university level demonstrated statistically significant positive effects. These results highlight the importance of aligning feedback strategies with specific learner characteristics and task contexts to facilitate optimal writing outcomes.

© 2025 AESS Publications. All Rights Reserved.

5. DISCUSSION

This study contributes to a better understanding of the effectiveness of CMF in L2 writing performance and highlights the role of moderating factors in shaping its outcomes. The findings offer several important implications for research and pedagogy, as well as directions for future study. Nevertheless, notable limitations are also present.

The meta-analysis revealed a weighted mean effect size of 1.602 indicating that CMF has a profound effect on L2 writing performance. This underscores its overall effectiveness in enhancing writing skills. The moderation analysis offered insights into factors influencing the impact of CMF. For example, feedback timing emerged as a critical factor, with immediate feedback ($g = 0.875$) proving significantly more effective than delayed feedback ($g = 0.482$). This aligns with prior research suggesting that immediate feedback enables learners to internalize changes while their memory of the task is still fresh. Similarly, Li's (2023) CMF research on vocabulary learning highlights the benefits of immediate feedback, yet it did not find a significant moderating effect. This suggests that feedback timing may be more crucial in writing where learners can revise immediately. Vocabulary learning may depend on other factors such as task structure and retention processes.

Furthermore, the proficiency level played a significant role in the effectiveness of CMF. Intermediate learners ($g = 0.789$) benefited the most from CMF because they are at a developmental stage where detailed feedback complements their evolving skills. In contrast, beginner learners ($g = 0.646$) may find complex feedback to be overwhelming due to limited cognitive and linguistic resources while advanced learners ($g = 0.421$) may focus more on the global aspects of writing, such as style or argumentation, which require less corrective input. These findings suggest that, for maximum impact, feedback strategies should align with learners' proficiency levels. Lv et al. (2021) also found that educational level moderated the effectiveness of online feedback with upper secondary school students benefiting more than university students. While both studies highlight the importance of tailoring feedback to learners' developmental stages, they differ in their interpretation of why higher proficiency levels yield lower effects. The present study suggests that advanced learners benefit less from CMF because they focus more on global writing aspects rather than corrective feedback whereas Lv et al. (2021) attribute the lower effect size at the university level to a lack of sufficient studies in that subgroup. These differences demonstrate the need for further investigation into how proficiency levels and educational settings interact with feedback effectiveness in L2 writing.

The analysis of software preferences revealed meaningful patterns. Pigai demonstrated a moderate effect size ($g = 0.568$), particularly in structured educational contexts. ChatGPT showed a smaller effect ($g = 0.433$), possibly reflecting its conversational feedback style which may lack the structured error detection found in tools such as Pigai. Nonetheless, ChatGPT's growing popularity since 2023 highlights the increasing appeal of AI-driven tools (Teng, 2024) although their long-term effectiveness warrants further research. Moreover, the geographical concentration of Pigai use in China underscores the influence of institutional and cultural factors on CMF implementation.

Task characteristics also influenced CMF outcomes. CMF yielded significant benefits with regard to academic writing tasks ($g = 1.057$) whereas argumentative ($g = 0.485$) and mixed ($g = 0.467$) tasks showed weaker effects. This suggests that feedback strategies must be tailored to task demands; notably, tools offering detailed, adaptive feedback are particularly beneficial for complex writing tasks. Similarly, Seyyedrezaei et al. (2024) found that the effectiveness of technology-integrated writing instruction varied by genre with collaborative technologies benefiting argumentative writing and non-collaborative technologies being more effective for narrative writing. Both studies highlight the importance of aligning instructional approaches with task demands, yet a key difference lies in the effectiveness of academic writing. While the present study found that CMF had the greatest impact on academic writing, Seyyedrezaei et al. (2024) did not emphasize this genre, instead focusing on the differential impact of technology types on argumentative and narrative writing. This suggests that CMF's effectiveness in academic writing may be attributed to its ability to provide structured and detailed feedback, whereas the success of technology-enhanced writing instruction may be more dependent on the level of collaboration and interaction

© 2025 AESS Publications. All Rights Reserved.

involved. These findings underscore the importance of considering both feedback mechanisms and technology types when designing instructional approaches for different writing genres.

This study has several limitations despite offering significant contributions. First, the relatively small dataset (k = 19) limits the generalizability of the findings, particularly for emerging tools such as ChatGPT. Second, although the overall effect size is robust, the high heterogeneity ($I^2 = 99.37\%$) indicates substantial variability among studies which may affect the consistency of the results. Funnel plot asymmetry suggests potential publication bias as smaller studies with significant results may be overrepresented. Although Rosenthal's fail-safe N supports the robustness of the findings, additional tests such as Egger's regression are recommended. Finally, this study primarily focuses on short-term outcomes, thereby neglecting the long-term impacts of CMF on writing development and learner autonomy.

Future research should address these limitations by expanding the dataset to include more diverse linguistic and cultural contexts. A larger pool of studies would enable robust comparisons between traditional CMF tools such as Pigai and emerging AI-driven systems such as ChatGPT. In addition, longitudinal studies are necessary for evaluating the sustainability of CMF's effects on writing performance and the role of CMF in fostering learner autonomy. Furthermore, future research should explore how learner characteristics, such as motivation and cognitive load interact with feedback complexity to inform the development of more personalized feedback strategies. Future research can refine CMF approaches to better align with learners' proficiency levels, task demands, and instructional contexts, ultimately enhancing L2 writing outcomes by addressing these gaps.

6. CONCLUSION

This study has systematically analyzed the effectiveness of CMF in L2 writing performance, utilizing 19 quantitative studies selected through rigorous screening. The overall effect size (g = 1.602) confirms the positive impact of CMF although significant heterogeneity highlights the role of contextual and moderating factors.

Key findings indicate that immediate feedback, automated tools, and feedback tailored to intermediate learners are particularly effective in improving L2 writing outcomes. Among software tools, Pigai demonstrated higher effectiveness than ChatGPT. However, the difference was not statistically significant. These results underscore the importance of aligning CMF strategies with learner proficiency, task characteristics, and feedback timing.

Further research must address the observed heterogeneity and evaluate the long-term impacts of CMF although this study provides valuable insights. Future studies can help facilitate more effective applications of feedback in language learning by refining CMF strategies to better suit diverse learner needs and instructional contexts.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The author states that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

Alkhalaf, N. A. (2020). Saudi female EFL learners and collaborative writing: Attitudes and challenges. *Theory and Practice in Language Studies*, *10*(9), 1118-1127. http://dx.doi.org/10.17507/tpls.1009.16

Ankawi, A. (2023). Saudi Arabian university students' perspectives on issues and solutions in academic writing learning. *English Language Teaching*, *16*(8), 1-1. https://doi.org/10.5539/elt.v16n8p1

Armanda, M. L., Nugraheni, A. F., Wulansari, A., & Imron, A. (2022). "Grammarly" as English writing assistant from EFL students' perspective. *English Education: Journal of English Teaching and Research*, 7(2), 128-137. https://doi.org/10.29407/jetar.v7i2.17988

© 2025 AESS Publications. All Rights Reserved.

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

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. Hoboken: John Wiley & Sons.

Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. (2019). *Cochrane handbook for systematic reviews of interventions*. Hoboken: Wiley.

Cui, Y., Schunn, C. D., Gai, X., Jiang, Y., & Wang, Z. (2021). Effects of trained peer vs. teacher feedback on EFL students' writing performance, self-efficacy, and internalization of motivation. *Frontiers in Psychology*, *12*, 788474. https://doi.org/10.3389/fpsyg.2021.788474

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

Ferris, D., & Eckstein, G. (2020). Language matters: Examining the language-related needs and wants of writers in a first-year university writing course. *Journal of Writing Research*, *12*(2), 299–339. https://doi.org/10.17239/jowr-2020.12.02.02

Fitriana, K., & Nurazni, L. (2022). Exploring English department students' perceptions on using Grammarly to check the grammar in their writing. *Journal of English Teaching*, *8*(1), 15-25.

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

Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system pigai. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(11), 322-330.

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*(7), 8435-8463. https://doi.org/10.1007/s10639-023-12146-0

Högemann, J., Cunha, J., Núñez, J. C., Vallejo, G., Rodríguez, C., & Rosário, P. (2021). Writing intervention with elementary students struggling with writing: Examining approach profiles to the teacher feedback on writing quality and motivational variables. *Reading and Writing*, *34*, 1681-1710. https://doi.org/10.1007/s11145-021-10159-0

Howell, E., Perez, S., & Abraham, W. T. (2021). Toward a professional development model for writing as a digital, participatory process. *Reading Research Quarterly*, *56*(1), 95-117. https://doi.org/10.1002/rrq.294

Jiang, L., & Yu, S. (2022). Appropriating automated feedback in L2 writing: Experiences of Chinese EFL student writers. *Computer Assisted Language Learning*, *35*(7), 1329-1353. https://doi.org/10.1080/09588221.2020.1799824

Li, R. (2023). Investigating effects of computer-mediated feedback on L2 vocabulary learning. *Computers & Education*, *198*, 104763. https://doi.org/10.1016/j.compedu.2023.104763

Lim, J., & Richardson, J. C. (2021). Predictive effects of undergraduate students' perceptions of social, cognitive, and teaching presence on affective learning outcomes according to disciplines. *Computers & Education*, *161*, 104063. https://doi.org/10.1016/j.compedu.2020.104063

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, *31*(9), 5614-5632. https://doi.org/10.1080/10494820.2021.2012812

Lv, X., Ren, W., & Xie, Y. (2021). The effects of online feedback on ESL/EFL writing: A meta-analysis. *The Asia-Pacific Education Researcher*, *30*(6), 643-653. https://doi.org/10.1007/s40299-021-00594-6

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

© 2025 AESS Publications. All Rights Reserved.

Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive Learning Environments*, *32*(2), 727-744. https://doi.org/10.1080/10494820.2022.2096642

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Pham, V. H. (2021). The effects of lecturer's model e-comments on graduate students' peer e-comments and writing revision. *Computer Assisted Language Learning*, *34*(3), 324-357. https://doi.org/10.1080/09588221.2019.1609521

Rahimi, M., & Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: A mixed methods study. *Computer Assisted Language Learning*, *35*(9), 2627-2674. https://doi.org/10.1080/09588221.2021.1888753

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Seyyedrezaei, M. S., Amiryousefi, M., Gimeno-Sanz, A., & Tavakoli, M. (2024). A meta-analysis of the relative effectiveness of technology-enhanced language learning on ESL/EFL writing performance: Retrospect and prospect. *Computer Assisted Language Learning*, *37*(7), 1771-1805. https://doi.org/10.1080/09588221.2022.2118782

Shang, H.-F. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, *30*(1), 4-16. https://doi.org/10.1080/10494820.2019.1629601

Taskıran, A., & Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education*, *23*(2), 120-139. https://doi.org/10.17718/tojde.1096260

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, *7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Thi, N. K., & Nikolov, M. (2022). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher*, *31*(6), 767-779. https://doi.org/10.1007/s40299-021-00625-2

Tunagür, M. (2021). The effect of peer assessment application on writing anxiety and writing motivation of 6th grade students. *Shanlax International Journal of Education*, *10*(1), 96-105. https://doi.org/10.34293/education.v10i1.4352

Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, *17*(1), 47-71. https://doi.org/10.1080/17501229.2021.1914062

Weng, F., Ye, S. X., & Xue, W. (2023). The effects of peer feedback on L2 students' writing motivation: An experimental study in China. *The Asia-Pacific Education Researcher*, *32*(4), 473-483. https://doi.org/10.1007/s40299-022-00669-y

Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, *50*, 100568. https://doi.org/10.1016/j.asw.2021.100568

Zhai, N., & Ma, X. (2022). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, *35*(9), 2817-2842. https://doi.org/10.1080/09588221.2021.1897019

Zhang, J., & Zhang, L. J. (2024). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning*, *37*(5-6), 1198-1223. https://doi.org/10.1080/09588221.2022.2069822

Zhao, X., & Yang, Y. (2023). Impact of social media-supported flipped classroom on English as a foreign language learners' writing performance and anxiety. *Frontiers in Psychology*, *13*, 1052737. https://doi.org/10.3389/fpsyg.2022.1052737

© 2025 AESS Publications. All Rights Reserved.

APPENDIX

Appendix 1. Features of the included studies.

Year	Author	Nationality (Author)	Effect size (g)	Ed. lv.	Major	Context	Prof.	Study design	Fb. source	Fb. timing	Task genre	Task type	Task setting	SW type	SW name
2021	Cui et al. (2021)	China	0.16	U	English	EFL	I	P	PF	I	A	I	CA	E	None
	Högemann et al. (2021)	Portugal	0.335	P	No	EFL	B	P	M	D	M	I	CA	G	None
	Tunagür (2021)	Turkey	0.173	P	No	EFL	B	EC	PF	D	M	I	CA	G	None
	Yao et al. (2021)	China	0.052	U	English	EFL	A	EC	PF	D	M	I	CA	E	Pigai
	Pham (2021)	Vietnam	2.021	U	Various	EFL	A	EC	PF	I	AC	I	CA	G	Other (Moodle)
	Gao (2021)	China	0.066	U	Various	EFL	I	P	AF	I	A	I	CA	E	Pigai
2022	Weng, Ye, and Xue (2023)	China	0.114	U	English	EFL	A	EC	PF	D	AC	I	CA	G	None
	Link, Mehrzad, and Rahimi (2022)	Iran	0.883	U	English	EFL	I	P	AF	D	AC	I	CA	E	Other (Criterion)
	Shang (2022)	Taiwan	-0.44	U	English	EFL	I	P	PF	D	M	I	CA	E	Pigai
	Taskıran and Goksel (2022)	Turkey	1.43	U	Various	EFL	B	P	M	I	AC	I	CA	E	Other (Write & improve)
	Jiang and Yu (2022)	China	2.18	U	Various	EFL	I	EC	AF	D	AC	I	CA	E	Pigai
	Thi and Nikolov (2022)	Myanmar	0.71	U	Various	EFL	I	P	M	I	AC	I	CA	E	Other (Grammarly)
2023	Barrot (2023)	Philippines	1.72	U	English	ESL	I	EC	AF	I	A	I	CA	E	Other (Grammarly)
	Waer (2023)	Egypt	1.37	U	English	EFL	I	P	AF	I	M	I	CA	E	Other (Write & improve)
	Liu et al. (2023)	China	0.04	U	Various	EFL	I	EC	AF	I	M	I	CA	E	Other (Mosoteach)
	Escalante et al. (2023)	Australia	0.063	U	Various	ESL	I	P	M	D	AC	I	CA	E	ChatGPT
	Mizumoto and Eguchi (2023)	USA	1.74	U	Various	EFL	I	P	AF	I	M	I	AT	E	ChatGPT
2024	Zhang and Zhang (2024)	China	0.98	U	Various	EFL	I	P	M	D	A	I	CA	E	Pigai
	Guo and Wang (2024)	China	-0.503	U	Various	EFL	A	EC	AF	I	A	I	CA	E	ChatGPT

Note: Education level: U (University) and P (Pre-university); proficiency: B (basic), I (Intermediate), and A (Advanced); study design: P (Pre- or post-test design) and EC (Treatment/Control group design); feedback source: PF (Peer), AF (Automated), TF (Teacher), and M (Mixed); feedback timing: I (Immediate) and D (Delayed); task genre: A (Argumentative writing), AC (Academic writing), and M (Mixed genre); task type: I (Individual task) and C (Collaborative task); task setting: CA (Classroom assignment) and AT (Assessment task); software type: G (General purposes) and E (Educational purposes).

© 2025 AESS Publications. All Rights Reserved.

Appendix 2. Descriptive information of the coding scheme.

Moderators	Subtypes	Definitions
Education level	Pre-tertiary	Education levels before university
	Tertiary	Education at the university
Context	EFL	English as a foreign language
	ESL	English as a second language
Proficiency	Basic	CEFR A1–A2 or pre-university students
	Intermediate	CEFR B1, first- or second-year English majors, or non-English major undergraduates
	Advanced	CEFR B2 or higher, third-year or above English majors, or graduate students
Study design	Pre- or post-test design	Compares performance before and after feedback
	Treatment/Control group design	Compares a feedback group with a no-feedback group
Feedback source	Peer feedback	Feedback received from peers via technologies
	Automated feedback	Feedback occurs between students and technologies
	Teacher feedback	Feedback received from teachers via technologies
Task genre	Argumentative writing	Writing to present and defend a position
	Academic writing	Formal, organized writing for academic purposes
	Mixed	Writing involving different genre types
Task type	Individual task	Tasks completed alone
	Collaborative task	Tasks completed in groups
Feedback timing	Immediate	Given right after the task
	Delayed	Given after some time
Task setting	Classroom assignment	Tasks done during instruction
	Assessment task	Tasks for evaluation purposes
Software type	General purposes	Technologies that were designed for non-educational purposes (e.g., Microsoft Word and SMS)
	Educational purposes	Technologies that were designed for educational purposes
Software name	ChatGPT	AI-based tool for dynamic, conversational feedback
	Pigai	Automated feedback tool widely used in educational settings
	Other	Software tools other than ChatGPT and Pigai

Appendix 3. Effect sizes of individual studies.

Study	Hedges's g	Standard error	Sample size	Lower CI	Upper CI	Z-value	p-value
Cui et al. (2021)	0.159	0.103	94	-0.042	0.362	1.552	0.120
Högemann et al. (2021)	0.329	0.149	45	0.043	0.627	2.274	0.029
Tunagür (2021)	0.052	0.048	26	-0.042	0.147	0.844	0.399
Yao et al. (2021)	1.978	0.164	37	1.698	2.343	12.293	0
Pham (2021)	0.094	0.126	104	-0.126	0.285	0.749	0.454
Gao (2021)	0.113	0.118	76	-0.11	0.339	0.993	0.320
Weng et al. (2023)	0.057	0.101	58	0.013	0.36	2.308	0.022
Link et al. (2022)	-0.432	0.146	47	-0.725	-0.154	-3.016	0.002
Shang (2022)	1.398	0.167	36	1.103	1.757	8.58	0
Taskıran and Goksel (2022)	1.978	0.334	9	1.526	2.833	6.54	0
Jiang and Yu (2022)	0.69	0.185	30	0.352	1.068	3.888	0
Thi and Nikolov (2022)	1.699	0.124	65	1.476	1.963	13.867	0
Barrot (2023)	1.359	0.098	103	1.176	1.563	13.903	0
Waer (2023)	0.039	0.105	103	-0.151	0.233	-0.405	0.685
Liu et al. (2023)	0.062	0.105	91	-0.142	0.268	0.6	0.547
Escalante et al. (2023)	0.063	0.066	226	-0.067	0.193	0.947	0.344
Mizumoto and Eguchi (2023)	1.74	0.009	12100	1.722	1.758	191.462	0
Zhang and Zhang (2024)	0.972	0.101	97	0.708	1.237	9.681	0
Guo and Wang (2024)	-0.495	0.141	50	-0.781	-0.225	-3.556	0

Views and opinions expressed in this article are the views and opinions of the author(s), International Journal of English Language and Literature Studies shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.

© 2025 AESS Publications. All Rights Reserved.