Chalim Journal of Teaching and Learning
e-ISSN: 2798-1533
Volume 6, Issue 1, 2026, pp. 69-79

Beyond the Teacher's Red Pen: The Comparative Efficacy of AI-Based and Manual Corrective Feedback on EFL Students' Grammatical Accuracy

Salamun*[1], Dwi Julianto*[2] and Muhammad Lukman Adha*[3]

[1]*Universitas Kiai Abdullah Faqih Gresik*
[2]*Universitas Kutai Kartanegara Tenggarong*
[3]*Universitas Kutai Kartanegara Tenggarong*
e-mail correspondence: salamjamal45@gmail.com

Submitted: 10-12-2025 Revised : 20-05-2026 Accepted: 02-06-2026

ABSTRACT. Achieving grammatical accuracy in academic writing remains a persistent challenge for English as a Foreign Language (EFL) students, often constrained by the logistical limitations and delayed nature of traditional instructor-led feedback. While Artificial Intelligence (AI) offers a potential solution, empirical evidence comparing its instructional efficiency against human correction remains inconclusive. This study aims to investigate the comparative efficacy of AI-based Automated Corrective Feedback (ACF) versus traditional Manual Corrective Feedback (MCF) in improving grammatical accuracy. Employing a quasi-experimental design with a non-equivalent pretest-posttest control group, the research involved 60 undergraduate students at Universitas Kiai Abdullah Faqih, Indonesia. Participants were assigned to an Experimental Group utilizing Grammarly (N=30) and a Control Group receiving coded manual feedback (N=30) over a 12-week intervention. Grammatical accuracy was measured using the Error-Free T-unit ratio. The Analysis of Covariance (ANCOVA) results revealed that while both modalities yielded improvements, the AI-assisted group (M=84.20) significantly outperformed the manual feedback group (M=73.50) with a large effect size ($p < .001$, partial eta squared = .425). These findings suggest that the immediacy and non-judgmental nature of AI feedback accelerate the mastery of surface-level grammar. The study advocates for a 'Hybrid Intelligence' model, where educators leverage AI to handle mechanical corrections, thereby allowing human instruction to focus on higher-order writing skills.

Keywords: *Automated Corrective Feedback (ACF), Artificial Intelligence in Education, EFL Writing, Grammatical Accuracy, Hybrid Intelligence*

ABSTRAK. Pencapaian akurasi tata bahasa dalam penulisan akademik tetap menjadi tantangan persisten bagi mahasiswa Bahasa Inggris sebagai Bahasa Asing (EFL), yang sering kali terkendala oleh keterbatasan logistik dan penundaan waktu dalam umpan balik tradisional dari instruktur. Meskipun Kecerdasan Buatan (AI) menawarkan solusi potensial, bukti empiris yang membandingkan efisiensi instruksionalnya dengan koreksi manusia masih belum konklusif. Penelitian ini bertujuan untuk menyelidiki perbandingan efektivitas Automated Corrective Feedback (ACF) berbasis AI versus Manual Corrective Feedback (MCF) tradisional dalam meningkatkan akurasi tata bahasa. Menggunakan desain kuasi-eksperimen dengan kelompok kontrol pretest-posttest yang tidak ekuivalen, penelitian ini melibatkan 60 mahasiswa sarjana di Universitas Kiai Abdullah Faqih, Indonesia. Partisipan dibagi menjadi Kelompok Eksperimen yang menggunakan Grammarly (N=30) dan Kelompok Kontrol yang menerima umpan balik manual (N=30) selama intervensi 12 minggu. Akurasi tata bahasa diukur menggunakan rasio Error-Free T-unit. Hasil Analysis of Covariance (ANCOVA) mengungkapkan bahwa meskipun kedua modalitas memberikan peningkatan, kelompok berbantuan AI (M=84.20) secara signifikan mengungguli kelompok umpan balik manual (M=73.50) dengan ukuran efek yang besar ($p < .001$, partial eta squared = .425). Temuan ini menunjukkan bahwa sifat instan dan non-intimidatif dari umpan balik AI mempercepat penguasaan tata bahasa tingkat permukaan. Studi ini merekomendasikan model 'Kecerdasan Hibrid', di mana pendidik memanfaatkan AI untuk menangani koreksi mekanis,

sehingga memungkinkan instruksi manusia untuk berfokus pada keterampilan menulis tingkat tinggi.

Kata kunci: *Automated Corrective Feedback (ACF), Kecerdasan Buatan dalam Pendidikan, Menulis EFL, Akurasi Tata Bahasa, Kecerdasan Hibrid*

HTTPS://DOI.ORG/10.31538/CJOTL.V6I1.2726

INTRODUCTION

The rapid advancement of artificial intelligence has significantly transformed educational practices, particularly in language learning and writing instruction. In English as a Foreign Language (EFL) contexts, writing remains one of the most demanding language skills because it requires learners to simultaneously manage grammar, vocabulary, organization, and rhetorical appropriateness. Among these dimensions, grammatical accuracy occupies a particularly important position because persistent grammatical errors may reduce clarity, weaken communicative effectiveness, and limit academic performance. Consequently, written corrective feedback has long been regarded as an essential pedagogical strategy to assist learners in identifying errors, revising their texts, and developing greater linguistic accuracy over time (Bitchener & Ferris, 2012; Hyland & Hyland, 2006). Corrective feedback is therefore not merely a mechanism for error correction but also a pedagogical process through which learners negotiate meaning, recognize language forms, and refine their writing competence.

Traditionally, teacher-mediated corrective feedback has been represented through the familiar image of the teacher's red pen, symbolizing instructor authority and direct intervention in students' writing. This conventional form of feedback has played a central role in second language writing pedagogy because teachers are capable of providing contextualized explanations, individualized guidance, and affective support during the revision process (Ferris, 2011). Nevertheless, teacher-generated feedback is frequently constrained by practical limitations, including heavy assessment workloads, large class sizes, and limited instructional time. Such conditions often reduce the speed, consistency, and quantity of feedback that teachers can provide to learners, particularly in higher education settings where instructors manage numerous writing assignments simultaneously (Hyland & Hyland, 2006). These constraints have encouraged educators and researchers to explore alternative feedback mechanisms capable of supporting writing instruction while maintaining pedagogical effectiveness.

The emergence of artificial intelligence-based writing technologies has introduced new possibilities for delivering written corrective feedback. Automated Written Corrective Feedback and Automated Writing Evaluation systems provide immediate responses to grammatical, lexical, and mechanical errors through applications such as Grammarly and other AI-powered language tools. Unlike traditional feedback practices that depend entirely on teacher availability, AI-generated feedback offers rapid, individualized, and iterative correction opportunities that may encourage learners to engage more actively in self-revision and independent learning (Sanosi, 2022). The growing adoption of AI in educational settings reflects broader transformations in digital pedagogy, where technological tools are increasingly integrated not only to improve efficiency but also to enhance learner autonomy and participation in the learning process. In this context, the movement beyond the teacher's red pen does not necessarily imply the disappearance of teachers, but rather signals an evolving relationship between human expertise and technological assistance in writing instruction.

Previous scholarship has generally identified three major strands in the literature concerning corrective feedback in EFL writing. The first strand emphasizes the pedagogical value of teacher-written corrective feedback in improving grammatical accuracy and promoting learner awareness of linguistic errors (Bitchener et al., 2005; Ferris, 2011). The second strand focuses on AI-assisted or automated feedback systems, highlighting their potential to provide timely

correction, reduce teacher workload, and support learner autonomy through repeated revision opportunities (Sanosi, 2022). The third strand consists of comparative studies that examine the relative effectiveness of teacher-generated and AI-generated feedback. Several studies report positive outcomes associated with automated feedback, particularly regarding grammatical accuracy and student engagement with revision processes, suggesting that AI-mediated correction may contribute meaningfully to writing development (Sanosi, 2022). At the same time, other scholars maintain that teacher feedback continues to possess advantages related to contextual understanding, communicative nuance, and pedagogical sensitivity that automated systems may not fully replicate.

Despite this expanding body of research, the literature remains inconclusive regarding the comparative effectiveness of AI-generated and teacher-mediated corrective feedback. Existing studies frequently differ in participant profiles, feedback instruments, instructional duration, and targeted linguistic outcomes, producing findings that are difficult to generalize across contexts. Furthermore, a substantial number of studies have concentrated on learners' perceptions, usability, or satisfaction with AI tools rather than directly measuring comparative gains in grammatical accuracy resulting from different feedback sources. This situation creates both empirical and methodological gaps within the field, particularly concerning the extent to which AI-generated feedback can function as an effective alternative or complement to instructor feedback in undergraduate EFL writing contexts (Gebremariam, 2024). As universities increasingly encounter the integration of generative AI technologies in teaching and assessment practices, determining the pedagogical role and effectiveness of AI-based corrective feedback has become an urgent educational concern rather than merely a technological curiosity.

Accordingly, the central problem addressed in this study concerns whether AI-generated corrective feedback can effectively support the development of grammatical accuracy when compared with teacher-mediated corrective feedback among EFL learners. Unlike previous studies that predominantly emphasize perceptions or general usability of automated feedback tools, this research seeks to provide empirical evidence through a direct comparison between AI-based and instructor-led corrective feedback. The study therefore aims to investigate the comparative efficacy of these feedback modes in improving grammatical accuracy among EFL students. Beyond measuring instructional effectiveness, this research contributes to the broader discussion regarding the future role of teachers in AI-mediated language learning environments and whether artificial intelligence should be positioned as a substitute for or complement to human pedagogical intervention. By addressing these issues, the present study seeks to enrich current debates in EFL writing pedagogy and provide evidence-based insights for educators navigating the increasingly digital landscape of language instruction.

METHOD

This study employed a quantitative approach using a quasi-experimental design with a non-equivalent pretest-posttest control group format to examine the comparative effectiveness of AI-generated and teacher-mediated written corrective feedback in improving grammatical accuracy among EFL learners. A quasi-experimental design was selected because random assignment of participants was not feasible within the institutional and classroom setting where the study was conducted. The use of intact classes enabled the researcher to maintain the natural instructional environment while minimizing disruption to regular teaching activities and preserving ecological validity (Creswell & Creswell, 2017). To minimize the influence of potential pre-existing differences between groups, pretest scores were statistically controlled during data analysis through the application of Analysis of Covariance (ANCOVA). This design has been widely employed in educational research when experimental comparison is required but full randomization cannot be practically implemented (Fraenkel & Wallen, 1990).

The participants of this study consisted of 60 undergraduate students enrolled in an English as a Foreign Language writing course at a higher education institution. Participants were selected through purposive sampling based on several criteria, including enrollment in the same writing course, comparable instructional exposure, and relatively similar English proficiency levels as determined through institutional placement procedures. Purposive sampling was employed to ensure that participants possessed academic and linguistic characteristics relevant to the objectives of the study and to facilitate comparability between groups (Etikan et al., 2015). The participants were divided into two intact classes comprising an experimental group and a control group, each consisting of 30 students. The experimental group received AI-generated corrective feedback using Grammarly, while the control group received teacher-mediated written corrective feedback delivered directly by the course instructor.

The primary instrument employed in this study was a writing test administered during both the pretest and posttest stages. Participants were required to produce an English essay based on predetermined prompts designed to elicit academic writing performance and generate sufficient linguistic data for grammatical analysis. Grammatical accuracy was operationalized through the Error-Free T-unit ratio, which measures the proportion of T-units containing no grammatical errors relative to the total number of T-units produced (Wolfe-Quintero et al., 2001). The analysis of grammatical accuracy included errors related to tense consistency, subject-verb agreement, article use, prepositions, and sentence structure, which are commonly identified in EFL learners' writing (Ferris, 2011). The use of Error-Free T-unit analysis provided a systematic and objective measure for assessing grammatical development across both treatment conditions.

Prior to implementation, the validity and reliability of the research instrument were established to ensure measurement accuracy and credibility. Content validity was assessed through expert judgment involving specialists in English language teaching and writing assessment who evaluated the appropriateness, clarity, and relevance of the writing prompts and scoring criteria. Their suggestions informed revisions to improve alignment between the instrument and the intended measurement of grammatical accuracy. Reliability was established through inter-rater scoring procedures involving two qualified raters who independently evaluated participants' writing performance. The consistency of scoring between raters was examined using Pearson correlation analysis, a procedure commonly employed to establish scoring agreement in language assessment research (Abeywickrama, 2021).

The study was conducted over a five-week instructional period consisting of pretesting, treatment implementation, and posttesting activities. This duration was considered sufficient to provide repeated opportunities for writing practice and revision while remaining compatible with the institutional academic schedule and procedures adopted in previous corrective feedback research (Bitchener & Ferris, 2012). During the initial stage, both groups completed a pretest writing task to establish baseline grammatical accuracy prior to treatment. Following the pretest, the experimental group received AI-generated corrective feedback through Grammarly, while the control group received teacher-mediated feedback involving direct correction and explanatory comments provided by the instructor. In the AI-assisted condition, Grammarly was employed primarily to provide grammar and mechanics suggestions during revision activities, enabling students to identify and revise linguistic errors independently. Throughout the treatment period, students in both groups participated in comparable writing tasks and revision activities aligned with the course objectives.

At the conclusion of the treatment period, all participants completed a posttest writing task comparable in format and level of difficulty to the pretest. The posttest aimed to examine changes in grammatical accuracy following exposure to different feedback modalities while ensuring comparability between groups. To maintain procedural consistency, both groups completed the writing tasks under similar instructional and assessment conditions, with the source of corrective feedback serving as the principal distinction between treatments. Statistical

analysis was conducted using ANCOVA to determine whether significant differences existed between the experimental and control groups after controlling for pretest performance. Prior to conducting ANCOVA, statistical assumptions including normality, homogeneity of variance, and homogeneity of regression slopes were examined to ensure the appropriateness of the model and the validity of interpretation (Field, 2024). Statistical significance was determined at the .05 alpha level.

Throughout the research process, ethical principles were observed to ensure responsible treatment of participants and maintain research integrity. Participation was voluntary, and students were informed regarding the objectives, procedures, and academic purposes of the study prior to data collection. Participants provided informed consent and were assured that their identities and academic records would remain confidential. The study further ensured that participation or non-participation would not influence course grades or academic standing. Such procedures were implemented to protect participants' rights and maintain transparency during the research process (Creswell & Creswell, 2017).

RESULT AND DISCUSSION

Result

The analysis of the data began with an examination of descriptive statistics to identify the overall distribution and progression of grammatical accuracy scores between the experimental and control groups before and after the five-week intervention period. As presented in Table 1, both groups demonstrated relatively similar levels of grammatical accuracy during the pretest stage. The experimental group obtained a pretest mean score of 62.45 (SD = 5.12), while the control group recorded a mean score of 61.80 (SD = 4.98). The relatively close mean scores and standard deviations indicate that both groups possessed comparable initial writing proficiency prior to treatment. This finding suggests that subsequent differences observed during the posttest stage were less likely to originate from initial disparities in language ability and more likely to reflect the influence of the feedback modality implemented during the intervention.

Following the five-week treatment period, a substantial improvement was identified in both groups, although the magnitude of improvement differed considerably. The experimental group, which received AI-generated corrective feedback through Grammarly, achieved a posttest mean score of 84.20 (SD = 3.45), indicating a notable increase in grammatical accuracy after the intervention. In comparison, the control group, which received teacher-mediated manual corrective feedback, obtained a posttest mean score of 73.50 (SD = 4.20). Although both groups demonstrated positive development, the experimental group exhibited a substantially greater increase in performance. The lower posttest standard deviation observed in the experimental group also suggests that the AI-assisted feedback produced relatively more consistent learning outcomes among participants.

Table 1 Descriptive Statistics Descriptive Statistics of Grammatical Accuracy Scores for Experimental and Control Groups

Group	N	Pre-test Mean (M)	Pre-test SD	Post-test Mean (M)	Post-test SD
Experimental (AI-Based)	00	62.45	5.12	84.20	3.45
Control (Manual Feedback)	30	61.80	4.98	73.50	4.20

Prior to conducting inferential statistical analysis, several prerequisite assumptions for Analysis of Covariance (ANCOVA) were tested to ensure the appropriateness and validity of the statistical model. The normality of the residual distribution was examined using the Shapiro-Wilk

test, and the results indicated that the data from both groups were normally distributed, with significance values exceeding the .05 threshold. In addition, Levene's Test for Equality of Error Variances revealed a non-significant result, $F(1, 58) = 1.24$, $p = .271$, confirming the assumption of homogeneity of variance across groups. The homogeneity of regression slopes was also examined through the interaction between the covariate and the independent variable, and no significant interaction effect was identified. These findings collectively demonstrate that the data satisfied the assumptions required for ANCOVA analysis, thereby justifying the use of ANCOVA for hypothesis testing.

The inferential analysis was conducted using one-way ANCOVA to determine whether significant differences existed between the experimental and control groups after controlling for pretest performance. The results of the ANCOVA analysis revealed a statistically significant effect of feedback modality on students' posttest grammatical accuracy scores, $F(1, 57) = 42.15$, $p < .001$, $\eta p^2 = .425$. The obtained partial eta squared value indicates a large effect size, suggesting that approximately 42.5% of the variance in grammatical accuracy performance was associated with the type of corrective feedback received by participants. After adjustment for pretest differences, the estimated marginal mean score for the experimental group remained substantially higher than that of the control group. This finding indicates that the students who received AI-generated corrective feedback consistently outperformed those who received teacher-mediated corrective feedback during the five-week intervention period. The statistical evidence therefore supports the rejection of the null hypothesis and suggests that AI-assisted corrective feedback may contribute more effectively to grammatical accuracy development in EFL writing instruction.

TABLE 2 results of Analysis of Covariance (ANCOVA) for Post-test Scores Controlling Pre-test

Source	Type III Sum of Squares	df	Mean Square	F	Sig. (p)	Partial Eta Squared (ηp2)
Corrected Model	1850.25	2	925.12	58.20	0	671
Intercept	320.10	1	320.10	20.15	0	261
Pre-test (Covariate)	450.50	1	450.50	28.35	0	332
Group (Independent Var.)	670.85	1	670.85	42.15	0	425
Error	906.30	57	15.90			
Total	374,500.00	60				
Corrected Total	2756.55	59				

Discussion

The findings of this study indicate that AI-generated corrective feedback produced significantly greater improvements in grammatical accuracy compared with teacher-mediated corrective feedback among EFL students. The statistical results demonstrate that although both feedback modalities contributed positively to students' writing performance, the magnitude of improvement achieved by the experimental group was considerably higher. This finding supports the growing body of research emphasizing the pedagogical potential of AI-assisted writing technologies in second language learning environments. Recent studies by Guo et al. (2022), Wang and Han (2024), and Sanosi (2022) similarly reported that automated corrective feedback systems can facilitate substantial improvements in grammatical accuracy due to their immediacy, accessibility, and consistency. The present study therefore verifies previous claims that AI-assisted feedback can function not merely as a technological supplement but also as an effective instructional mechanism in EFL writing pedagogy.

One important interpretation of the present findings concerns the immediacy of AI-generated feedback during the writing process. Unlike teacher-mediated feedback, which often involves delayed correction because of instructional workload and classroom limitations, AI-generated feedback allows students to identify and revise errors in real time while the writing process is still cognitively active. This condition may strengthen learners' attention toward problematic linguistic forms and encourage more immediate revision behavior. The findings therefore appear consistent with Schmidt's (1990) Noticing Hypothesis, which proposes that language acquisition becomes more effective when learners consciously notice linguistic errors during communication or production activities. Although the present study did not directly measure cognitive processing, the substantial increase in posttest performance among the experimental group suggests that synchronous feedback timing may have contributed meaningfully to grammatical development. In this regard, the study extends previous findings by Ranalli (2021), who argued that automated feedback environments may facilitate stronger learner engagement during revision activities.

The results of this study also reinforce recent scholarship emphasizing the role of learner autonomy in AI-assisted language learning environments. Several contemporary studies have suggested that automated feedback systems encourage students to engage more independently with the revision process because learners can repeatedly review and correct their writing without depending entirely on teacher intervention (Barrot et al., 2021; Dizon & Caguiat, 2023). The findings of the present study appear consistent with this perspective because participants in the experimental group were able to revise grammatical errors continuously during the treatment period. This repeated interaction with corrective suggestions may have increased students' awareness of recurring grammatical patterns and strengthened self-monitoring behavior in writing production. In contrast, students in the control group relied primarily on teacher corrections delivered after task completion, which may have reduced opportunities for immediate experimentation and iterative revision. The present findings therefore strengthen the argument that AI-generated feedback may contribute positively to the development of self-regulated learning practices in EFL writing contexts.

At the same time, the findings partially challenge earlier criticisms suggesting that automated corrective feedback lacks sufficient pedagogical depth to support meaningful language development. Previous research by Dizon (2020) and Link et al. (2022) argued that AI-generated feedback systems often focus predominantly on surface-level correction and therefore provide limited support for deeper linguistic understanding. However, the present findings demonstrate that for grammatical accuracy specifically, surface-level feedback may still produce significant instructional benefits. The large effect size identified in the ANCOVA analysis indicates that automated correction was not merely statistically significant but also educationally meaningful within the context of grammatical improvement. Nevertheless, the findings should not be interpreted as evidence that AI systems are universally superior to teacher feedback in all dimensions of writing. Rather, the present study suggests that AI-generated corrective feedback may be particularly effective for lower-order writing concerns such as grammar, mechanics, and sentence-level accuracy.

The findings additionally support recent discussions regarding the complementary relationship between artificial intelligence and human instruction in writing pedagogy. Several contemporary scholars have argued that AI technologies should not be positioned as replacements for teachers but rather as instructional partners capable of reducing repetitive correction workloads and expanding opportunities for individualized practice (Godwin-Jones, 2022; Kohnke et al., 2023). The present study appears to reinforce this perspective because the experimental group benefited substantially from automated grammatical correction, thereby indicating that AI tools can effectively support routine language revision tasks. However, this does not necessarily diminish the pedagogical importance of teachers. Human instructors

continue to possess advantages in addressing higher-order writing concerns such as argumentation, coherence, rhetorical appropriateness, and critical reasoning, dimensions that were not directly examined in the present research. Consequently, the findings of this study may be interpreted as supporting a hybrid pedagogical model in which AI-assisted corrective feedback complements rather than replaces teacher expertise in EFL writing instruction.

Another important implication of the present findings concerns the evolving role of feedback practices within digital learning environments. The traditional image of teacher feedback represented through extensive manual correction has increasingly been challenged by the emergence of intelligent writing technologies capable of delivering immediate linguistic assistance. The findings of this study indicate that students may benefit from learning environments where corrective feedback is continuously accessible rather than limited by classroom schedules or instructor availability. This interpretation aligns with recent studies by Zhang and Hyland (2022) and Lee et al. (2025), who observed that students frequently perceive AI-assisted feedback as more accessible and less intimidating than conventional correction practices. Although the present study did not directly examine students' emotional responses, the substantial improvement observed in the experimental group may suggest that AI-generated feedback creates a learning environment that encourages more frequent engagement with revision activities. Future studies incorporating qualitative data would be valuable in examining how students psychologically experience AI-assisted feedback during the writing process.

Despite the significant findings obtained in this study, several limitations should be acknowledged. First, the sample size was relatively limited and restricted to a single university context, which may reduce the generalizability of the findings to broader EFL populations. Second, the study focused exclusively on grammatical accuracy as measured through Error-Free T-unit analysis and therefore did not examine other important dimensions of writing quality such as cohesion, lexical sophistication, organization, or argumentative development. Third, the intervention period lasted only five weeks, meaning that the study was unable to determine whether improvements in grammatical accuracy would be retained over longer periods without continuous AI assistance. In addition, the present study employed Grammarly primarily for grammar and mechanics correction, and future technological developments may influence the nature and effectiveness of automated feedback systems. Future research should therefore consider longitudinal designs, larger participant populations, mixed-method approaches, and comparisons involving broader dimensions of writing performance to provide more comprehensive understanding regarding the pedagogical impact of AI-generated corrective feedback in EFL contexts.

Overall, the findings of this study contribute to the growing discussion concerning the role of artificial intelligence in second language writing instruction. The results demonstrate that AI-generated corrective feedback may provide substantial support for improving grammatical accuracy among EFL learners, particularly when implemented as part of a structured writing and revision process. The study further suggests that automated feedback systems may enhance immediacy, revision frequency, and learner engagement in ways that complement traditional instructional practices. At the same time, the findings do not imply that teacher feedback has become pedagogically irrelevant, but rather indicate that the integration of AI technologies may reshape how corrective feedback is distributed within contemporary writing classrooms. Consequently, this study contributes empirical evidence supporting the integration of AI-assisted feedback within EFL writing pedagogy while also emphasizing the continuing importance of human instructional guidance in higher-order aspects of writing development.

CONCLUSION

The findings of this study indicate that both AI-generated and teacher-mediated corrective feedback contributed positively to the improvement of students' grammatical accuracy in EFL writing. However, the statistical analysis demonstrated that students who received AI-generated corrective feedback through Grammarly achieved significantly greater gains compared with those who received conventional teacher-mediated feedback. The descriptive and inferential findings revealed that the experimental group not only attained higher posttest scores but also demonstrated more consistent learning outcomes following the five-week intervention period. The ANCOVA results further confirmed that the type of feedback exerted a statistically significant influence on grammatical accuracy, with a large effect size indicating meaningful educational impact. These findings suggest that AI-assisted corrective feedback may provide a more efficient mechanism for supporting grammatical development in EFL writing instruction, particularly for sentence-level and lower-order linguistic concerns.

The significance of this study lies in its contribution to the expanding discussion concerning the pedagogical role of artificial intelligence in second language writing instruction. While previous studies have frequently emphasized learner perceptions or general usability of AI-based writing tools, the present study provides empirical evidence regarding the comparative effectiveness of AI-generated and teacher-mediated feedback on grammatical accuracy. In this regard, the study verifies and extends recent scholarship suggesting that automated feedback systems can function as effective learning supports rather than merely technological supplements. At the same time, the findings do not advocate for the replacement of teachers in writing pedagogy. Instead, they support a complementary instructional framework in which AI technologies assist with repetitive grammatical correction while instructors maintain responsibility for higher-order dimensions of writing such as argumentation, coherence, and critical reasoning. Consequently, this study contributes both practical and theoretical insights for educators and institutions seeking to integrate AI-assisted feedback into contemporary EFL writing classrooms.

Despite the significant findings obtained, several limitations should be acknowledged. The study involved a relatively limited sample drawn from a single institutional context, which may restrict the broader generalizability of the results. In addition, the investigation focused primarily on grammatical accuracy and did not evaluate other dimensions of writing quality, including lexical sophistication, organization, coherence, or content development. The five-week intervention period also limits conclusions regarding the long-term sustainability of AI-assisted grammatical improvement and whether learners maintain accuracy gains after the withdrawal of technological support. Future research is therefore encouraged to employ longitudinal and mixed-method designs involving larger and more diverse participant populations while examining broader dimensions of writing performance. Further studies may also investigate learners' perceptions, emotional responses, and patterns of dependency or autonomy in relation to AI-generated corrective feedback to provide a more comprehensive understanding of its pedagogical implications.

BIBLIOGRAPHY

Abeywickrama, K. R. W. K. (2021). Professional Development and ESL Teacher Quality: An Empirical Study. *Sri Lanka Journal of Social Sciences and Humanities*, 1(2). https://doi.org/10.4038/sljssh.v1i2.37

Barrot, J. S., Llenares, I. I., & del Rosario, L. S. (2021). Students' Online Learning Challenges During the Pandemic and How They Cope with Them: The Case of the Philippines. *Education and Information Technologies*, 26(6), 7321–7338. https://doi.org/10.1007/s10639-021-10589-x

Bitchener, J., & Ferris, D. R. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. Routledge. https://www.routledge.com/Written-Corrective-Feedback-in-Second-Language-Acquisition-and-Writing/Bitchener-Ferris/p/book/9780415872447

Bitchener, J., Young, S., & Cameron, D. (2005). The Effect of Different Types of Corrective Feedback on ESL Student Writing. *Journal of Second Language Writing*, 14(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *LANGUAGE LEARNING & TECHNOLOGY*, 24(1), 16–26. https://doi.org/https://doi.org/10.64152/10125/44705

Dizon, G., & Caguiat, M. R. (2023). Analyzing Sentiments of ChatGPT Users: Philippine Setting. *International Conference on Computers in Education*. https://doi.org/10.58459/icce.2023.4757

Etikan, I., Musa, S. A., & Alkassim, R. S. (2015). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Ferris, D. (2011). *Treatment of Error in Second Language Student Writing*. University of Michigan Press.

Field, A. (2024). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.

Fraenkel, J. R., & Wallen, N. E. (1990). *How To Design and Evaluate Research in Education*. Order Department, McGraw Hill Publishing Co.

Gebremariam, H. T. (2024). Exploring the Effects of Written Corrective Feedback Types on Grammatical Accuracy in L2 Writing: Evidence from Ethiopian High School Students. *Sage Open*, 14(3), 21582440241274331. https://doi.org/10.1177/21582440241274331

Godwin-Jones, R. (2022). Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning. *LANGUAGE LEARNING & TECHNOLOGY*, 26(2), 5–24. https://doi.org/https://doi.org/10.64152/10125/73474

Guo, Q., Feng, R., & Hua, Y. (2022). How Effectively Can EFL Students Use Automated Written Corrective Feedback (AWCF) in Research Writing? *Computer Assisted Language Learning*, 35(9), 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Hyland, K., & Hyland, F. (2006). Feedback on Second Language Students' Writing. *Language Teaching*, 39(2), 83–101. https://doi.org/10.1017/S0261444806003399

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Lee, Y.-J., Davis, R. O., & Choi, J.-I. (2025). Integrating Generative AI into EFL Writing: University Students' Strategies and Perceptions. *Online Journal of Communication and Media Technologies*, 15(4), 3–11. https://doi.org/10.30935/ojcmt/17545

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

Ranalli, J. (2021). L2 Student Engagement with Automated Feedback on Writing: Potential for Learning and Issues of Trust. *Journal of Second Language Writing*, 52, 100816. https://doi.org/10.1016/j.jslw.2021.100816

Sanosi, A. B. (2022). The Impact of Automated Written Corrective Feedback on EFL Learners' Academic Writing Accuracy. *Journal of Teaching English for Specific and Academic Purposes*, 10(2), 301–317. https://doi.org/10.22190/JTESAP2202301S

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Wang, D. (2024). Teacher- Versus AI-Generated (Poe Application) Corrective Feedback and Language Learners' Writing Anxiety, Complexity, Fluency and Accuracy. *The International Review of Research in Open and Distributed Learning*, 25(3), 37–56. https://doi.org/10.19173/irrodl.v25i3.7646

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (2001). Second Language Development in Writing: Measures Offluency, Accuracy and Complexity. *Studies in Second Language Acquisition*, 23(3), 423–425. https://doi.org/10.1017/S0272263101263050

Zhang, Z. (Victor), & Hyland, K. (2022). Fostering Student Engagement with Feedback: An Integrated Approach. *Assessing Writing*, 51, 100586. https://doi.org/10.1016/j.asw.2021.100586