**Comparing AI and Teacher Corrective Feedback on Iranian EFL Learners’ Essay Writing Skills**

.

ABSTRACT

|  |
| --- |
| **Aims**: This study aimed to compare the effectiveness of teacher-generated versus AI-generated corrective feedback on the essay writing skills of Iranian intermediate EFL learners.  **Study design**: Quasi-experimental design.  Place and Duration of Study: Department of TEFL, Islamic Azad University, North Tehran Branch, conducted over a 16-week period in 2025.  **Methodology**: A total of 80 Iranian intermediate EFL learners were selected through convenience sampling and divided into two equal groups: Teacher-Generated Corrective Feedback Group (TGFG, n=40) and AI-Generated Corrective Feedback Group (AGFG, n=40). Both groups participated in 16 instructional sessions, receiving feedback on their essays either from a human instructor or an AI-based application (ChatGPT). Writing performance was assessed using IELTS-based pretests and posttests, evaluated based on content, organization, vocabulary, language use, and mechanics. The 6+1 Traits Writing Rubric was used for scoring, and inter-rater reliability was established.  **Results**: While both groups showed improvement in essay writing performance, the AI-generated feedback group (M = 18.08, SD = 1.32) significantly outperformed the teacher-generated group (M = 16.93, SD = 1.29) on the posttest. An independent-samples t-test indicated a significant difference between the two groups (t(78) = 3.92, P = .000, Cohen’s d = .892), favoring the AI feedback group.  **Conclusion**: The findings suggest that AI-generated corrective feedback is more effective than teacher-generated feedback in improving the essay writing performance of EFL learners. These findings have important implications for future teaching practices, particularly in enhancing learner autonomy and reducing teacher workload in large classrooms. AI tools can serve as a reliable and efficient alternative in writing instruction, particularly in contexts with limited instructional resources. However, their effectiveness may vary depending on learners’ proficiency levels, task complexity, and their familiarity with digital tools. |

*Keywords:* *AI-generated Corrective Feedback, Teacher-generated Corrective Feedback, Writing Skills,*

*vocabulary*

1. INTRODUCTION

Writing is a fundamental skill in language learning, requiring mastery of grammar, structure, and coherence. Proficiency in writing is widely recognized as one of the strongest predictors of academic achievement (Geiser & Studley, 2001). For second language (L2) learners, writing functions not only as a means of communication but also as a tool for organizing thoughts and fostering critical thinking (Raimes, 1983). However, writing in a foreign language presents unique challenges, as learners must simultaneously develop linguistic accuracy, fluency, and coherence while adhering to the conventions of the target language (Hyland, 2003). Research has shown that writing proficiency is closely associated with cognitive and metacognitive skills, highlighting the critical role of corrective feedback in effective writing instruction (Arnold et al., 2017; Altun, 2022). Corrective feedback is widely acknowledged as an essential strategy for enhancing writing accuracy and fluency (Bitchener & Knoch, 2009). Teacher-generated feedback offers explicit, individualized guidance that targets learners’ specific writing weaknesses (Hyland & Hyland, 2006; Ellis, 2009). Nevertheless, providing timely and personalized feedback can be challenging for teachers, particularly in large classroom settings (Calker & Tweedly, 2016). Teacher corrective feedback plays a crucial role in writing development. Muchemwa et al. (2019) found that when teachers provided structured feedback aligned with curriculum goals including Feed Up, Feed Back, and Feed Forward—it helped students improve their composition writing. However, they also noted that unclear correction codes limited some learners’ ability to benefit fully, emphasizing the need for feedback to be both informative and accessible. In addition to providing accurate feedback, teachers need to ensure that their comments are clear, accessible, and aligned with students’ cognitive and digital abilities. Developing this skill, often referred to as feedback literacy, requires not only pedagogical awareness but also strong digital and writing competence, especially for future educators (Emidar et al., 2023).With the advancement of artificial intelligence (AI), technology-driven corrective feedback has emerged as a promising alternative, offering immediate and consistent feedback on grammar, organization, and coherence (Godwin-Jones, 2022). AI-powered tools such as Grammarly and ChatGPT have demonstrated the potential to support learners by delivering instant corrections and promoting learner autonomy (Shi & Deng, 2024). Despite these advantages, AI-generated feedback has notable limitations. Studies suggest that while AI tools are effective in improving grammatical accuracy, they often struggle with higher-order writing components such as argumentation, rhetorical structure, and context-sensitive revision (Kuteeva & Andersson, 2024; Wang, 2024). Furthermore, learners in contexts such as Iran—where teacher-centered instruction predominates—may initially resist or lack trust in AI-generated feedback (Derakhshan & Shirejini, 2020). Given these contrasting perspectives, the present study aims to compare the effectiveness of AI-generated versus teacher-generated corrective feedback on the essay writing performance of Iranian intermediate EFL learners. By examining the impact of each feedback type, this research seeks to offer insights into the integration of AI in EFL writing instruction and its potential to complement traditional pedagogical approaches.

Corrective feedback is a fundamental component of second language acquisition, helping learners refine their linguistic accuracy and writing skills (Lyster et al., 2013). Scholars categorize corrective feedback as explicit, which provides direct corrections, and implicit, which encourages self-correction (Ellis, 2009). Teacher-generated feedback has traditionally been valued for its ability to address both surface-level grammatical errors and deeper issues related to coherence, organization, and argumentation (Hyland, 2003). Teachers provide detailed explanations, adapt feedback to students’ proficiency levels, and offer encouragement, which can enhance learners’ motivation and engagement (Ellis, 2009). In contrast, machine-generated corrective feedback has gained attention due to its ability to deliver immediate and automated responses. AI-powered tools such as Grammarly and ChatGPT use natural language processing (NLP) algorithms to detect errors and provide corrective suggestions on grammar, word choice, and sentence structure (Godwin-Jones, 2022). Research conducted in Iran has shown that AI feedback significantly improves grammatical accuracy, but it is less effective in addressing complex writing issues such as organization and coherence (Bagheri Nevisi & Arab, 2023; Borna et al., 2024). Furthermore, AI-generated feedback can enhance learner autonomy and engagement by providing instant feedback, which facilitates self-correction and revision processes (Shi & Deng, 2024). Despite its efficiency, AI feedback has several limitations. Studies suggest that AI tools often focus on lower-order concerns and struggle with higher-order writing aspects such as critical thinking, logical argumentation, and cultural appropriateness (Kuteeva & Andersson, 2024). For example, while Grammarly effectively identifies grammatical errors, it does not provide meaningful feedback on argument structure or idea development (Steiss et al., 2024). Additionally, AI lacks the human elements of encouragement, motivation, and pedagogical support inherent in teacher-generated feedback (Hyland, 2003). Several studies have examined the effectiveness of AI-generated feedback compared to traditional teacher feedback. Wang (2024) found that AI-generated feedback was particularly effective in improving writing fluency and reducing writing anxiety, whereas teacher feedback was more beneficial for enhancing coherence and idea development. Similarly, Steiss et al. (2024) concluded that while AI feedback excels in identifying grammatical errors, it lacks the depth and contextual awareness that human feedback provides. Research in Iran by Derakhshan & Shirejini (2020) indicates that EFL learners may require training to effectively engage with AI-generated feedback, as cultural and contextual factors influence their perceptions of AI tools. In the same vein, Borna et al. (2024) studied the effect of AI writing assistance tools on Iranian intermediate EFL Learners’ writing performance. They focused on prewriting aid and Grammarly. The findings showed that prewriting aid and Grammarly tools helped students write better by giving quick, clear feedback. This type of feedback made their writing more accurate and smoother. Given the strengths and limitations of both feedback types, scholars suggest that a hybrid approach combining AI-generated and teacher-generated feedback may be the most effective strategy for writing instruction (Luo et al., 2025).

The present study compared the effect of AI-generated and teacher-generated corrective feedback on Iranian EFL learners’ essay writing. The following research question helped the researchers accomplish the purpose of this study.

RQ: Do AI-generated corrective feedback and teacher-generated corrective feedback have different effects on the essay writing skills of Iranian intermediate EFL learners?

2. METHODOLOGY

**2.1 Participants**

The study involved 80 Iranian EFL students at the intermediate level from Islamic Azad University, North Tehran Branch. The participants were between 18 and 35 years old, came from different social and economic backgrounds, and spoke Persian as their first language. All participants had approximately seven years of experience learning English. They were selected from two intact classes.

**2.2 Instruments**

The researcher used the following instruments to achieve the objectives of the study. The Oxford Placement Test (OPT) (Allen,

2004) is a validated placement test published by Oxford University Press and was used to homogenize the learners regarding general English proficiency. It included 60 items on vocabulary and grammar. Additionally, an IELTS-based writing samples were used as both the pretest and posttest to evaluate students writing performance. Two instructors rated the writings based on 6+1 Traits Writing Rubric that was employed to assess students’ writing, focusing on ideas, organization, voice, word choice, sentence fluency, conventions, and presentation. The correlation between the two ratings was computed through the Pearson product-moment correlation coefficient formula, and the results presented a high inter-rater reliability index for the pretest and posttest. This study utilized ChatGPT, powered by OpenAI’s GPT-4, an advanced language model renowned for its enhanced capabilities in understanding and generating text. Accessible via both web and mobile applications, ChatGPT provided feedback aligned with a predefined rubric and occasionally offered suggestions and revisions aimed at improving the quality of writing.

**2.3 Procedure**

To intact classes met one session a with a 90 minutes duration within 16 weeks. The participants studied “Think to Write, Write to Think” (Vosough-Yazdani, 2018), through which they learned about writing skills, paragraph writing, essay writing, cohesion, coherence and mechanics of writing. It is worthy to pointing that both groups studied the same course material. An IELTS-based writing pretest and posttest were administered at the beginning and end of the study to assess their writing skills.

**2.4 Placement test**

The Oxford Placement Test (OPT) was administered to 100 students. Based on their scores, 80 homogeneous intermediate-level learners were selected and randomly assigned into two groups: the AI-Generated Corrective Feedback Group (AIGCFG) and the Teacher-Generated Corrective Feedback Group (TGCFG).

**2.5 Pre-Test**

The instructor selected IELTS-based writing samples as a pretest. Besides, the 6+1 traits rubric was used for the correction procedure. The writings were rectified by two qualified instructors and the mean of the two sets of scores was the student’s final score. Following the pretest, the students were randomly assigned to either the Teacher-Generated Corrective Feedback Group (TGCFG) or the AI-Generated Corrective Feedback Group (AICFG), ensuring a balanced distribution of writing skills across the groups.

**2.6 AI-Generated Corrective Feedback Group (AGFG)**

After administering the Oxford placement test and pretest in the writing, the treatment phase started and the students learned how to receive feedback from AI. Each 90-minute session was divided into two parts: 50 minutes of direct instruction from the teacher on writing conventions and skills, followed by 40 minutes of individual writing practice where students received AI feedback. Students in the AIGCFG uploaded their written essays to ChatGPT and received real-time feedback based on 6+1 traits rubric. AI also offered suggestions for revision, helping students to self-correct and improve their writing over time. Also, it increased students’ autonomy and engagement.

**2.7 Teacher-Generated Corrective Feedback Group (TGFG)**

In the other cluster TGCFG students exclusively received feedback from teacher in a traditional manner. The teacher explicitly stated the writing conventions, cohesion, coherence and word choice. In TGCFG teacher taught the lesson in each session and students only benefited from teacher’s written corrective feedback.

**2.8 Post-Test**

Finally, after the treatment phase was completed, both TGCFG and AIGCFG participants took a post-test to assess whether the treatment had any impact on their essay writing skills. The post-test, which was based on IELTS writing samples, was administered to both groups under timed conditions. Students had 40 minutes to write about a specific topic. The pretest and post-test results were evaluated by two experienced writing teacher, and the scores were compared to determine any improvements in students’ writing abilities.

3. results

The following section displays the results and findings of the research and the data analysis employed in this research. The students’ pretest and posttest scores in both groups were used to answer the research question. The data collected for this study were analyzed through Independent-Samples t-test which requires normality of data, and homogeneity of variances of groups. The skewness indices examine symmetry of the data, while the kurtosis indices probe their relative height. In an ideally normal distribution, the skewness and kurtosis indices are equal to zero as shown in Table 1. Thus, it was concluded that the present data did not show any significant deviation from normality.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Skewness and Kurtosis Indices of Normality** | | | | | | |
|  |  | **N** | **Skewness** | | **Kurtosis** | |
| Group |  | **Statistic** | **Statistic** | **Std. Error** | **Statistic** | **Std. Error** |
|  | OPT | 40 | .146 | .374 | -1.030 | .733 |
| Experimental | Pretest | 40 | -.236 | .374 | -.729 | .733 |
|  | Posttest | 40 | .021 | .374 | -.919 | .733 |
|  | OPT | 40 | .255 | .374 | -.738 | .733 |
| Control | Pretest | 40 | .013 | .374 | -.695 | .733 |
|  | Posttest | 40 | .362 | .374 | -.478 | .733 |

Table 2 the results of inter-rater reliability for pretest and posttest of essay writing. As shown in Table 2, there were significant agreements between the two raters on pretest (r (78) = .935 representing a large effect size, P < .05), and posttest (r (78) = .980 representing a large effect size, P < .05) of essay writing.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2. Pearson Correlations for Inter-Rater Reliability Indices for Pretest and Posttest of Essay Writing** | | | |
|  |  | **Pretest Rater2** | **Posttest Rater2** |
|  | Pearson Correlation | .779\*\* |  |
| Pretest  Rater1 | Sig. (2-tailed) | .000 |  |
|  | N | 40 |  |
|  | Pearson Correlation |  | .916\*\* |
| Posttest  Rater1 | Sig. (2-tailed) |  | .000 |
|  | N |  | 40 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |

An independent-samples t-test was run to probe whether the two groups were homogenous in terms of the general language proficiency before the treatment. Table 3 shows the results of descriptive statistics for the two groups on the OPT. The results indicated that the AI-generated Corrective Feedback (M=43.63, SD=1.21) and Teacher-generated Corrective Feedback (M=43.40, SD=1.35).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3. Descriptive Statistics for Oxford Placement T*est by Groups*** | | | | | |
| **Group** |  | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| Experimental |  | 40 | 43.63 | 1.213 | .192 |
| OPT |  |  |  |  |  |
| Control |  | 40 | 43.40 | 1.355 | .214 |

Table 4 displays the results of the Independent-Samples t-test. This study reports the results in two rows. The results displayed across the first row; i.e., “Equal variances assumed” were reported due to the fact that the assumption of homogeneity of variances was retained (F = .433, *P* > .05). The results of Independent Samples t-test; (t (78) = .783, *P* > .05, Cohen’s d = .182 representing a weak effect size) indicated that there was not any significant difference between the two groups’ means on OPT. Thus, it can be concluded that the two groups were homogeneous in terms of their general language proficiency prior to the administration of the treatment. Figure 1 shows the two groups’ means on the OPT test. As it was discussed above, both groups had almost the same means on the OPT test.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 4. Independent-Samples t-test for Oxford Placement Test by Groups** | | | | | | | | | |
|  | **Levene's Test for Equality of Variances** | |  |  |  | **t-test for Equality of Means** | |  |  |
|  | **F** | **Sig.** | **T** | **Df** | **Sig. (2-**  **tailed)** | **Mean Difference** | **Std. Error Difference** | **95% Confidence Interval of the Difference** | |
|  |  |  |  |  |  |  |  | **Lower** | **Upper** |
| Equal variances assumed | .433 | .513 | .783 | 78 | .436 | .225 | .288 | -.347 | .797 |
| Equal variances not assumed |  |  | .783 | 77.063 | .436 | .225 | .288 | -.348 | .798 |



**Figure 1. Means of Oxford Placement Test by Groups**

Table 5 examines the assumption of homogeneity of regression slopes. The significant interaction effect between the independent variable and the pretest (F(1,76) = 4.65, p > .05) indicates that this assumption was violated.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 5. Tests of Between-Subjects Effects for Testing Homogeneity of Regression Slopes** | | | | | | |
| **Source** | **Type III Sum**  **of Squares** | **df** | **Mean Square** | **F** | **Sig.** | **Partial Eta**  **Squared** |
| Group | .754 | 1 | .754 | 2.521 | .117 | .032 |
| Pretest | 109.764 | 1 | 109.764 | 366.822 | .000 | .828 |
| Group \* Pretest | 1.393 | 1 | 1.393 | 4.655 | .034 | .058 |
| Error | 22.741 | 76 | .299 |  |  |  |
| Total | 24678.000 | 80 |  |  |  |  |

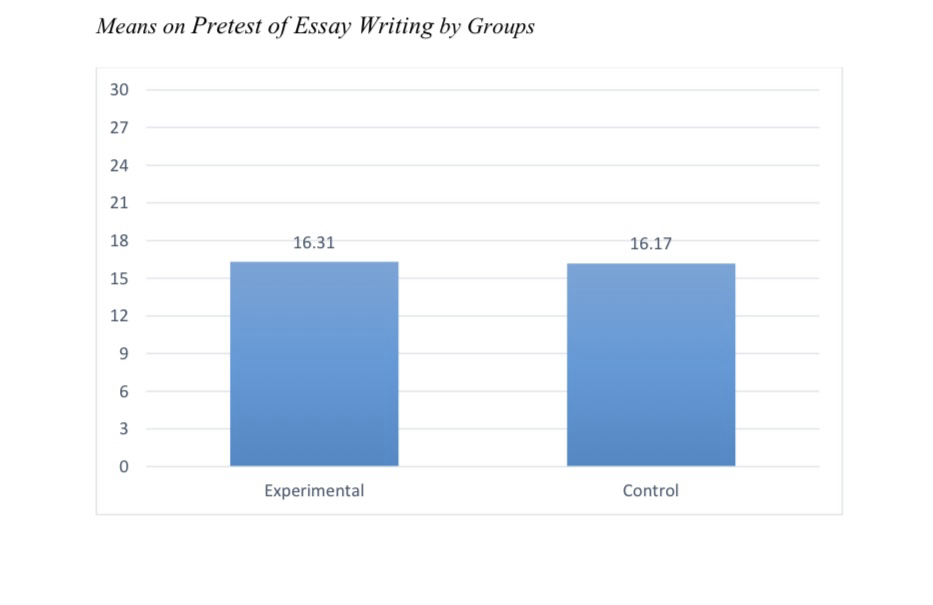
Table 6 shows the descriptive statistics for the essay writing pretest, demonstrating that the experimental (M = 16.31, SD = 1.02) and control groups (M = 16.17, SD = 1.29) had comparable writing proficiency levels at the outset of the study.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 6. Descriptive Statistics for Pretest of Essay Writing by Groups** | | | | | |
| **Group** |  | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| Experimental |  | 40 | 16.31 | 1.025 | .162 |
| Pretest |  |  |  |  |  |
| Control |  | 40 | 16.17 | 1.293 | .204 |

Independent-Samples t-test used for the pretest scores. This study reports the results in two rows. The results displayed across the first row; i.e., “Equal variances assumed” were reported due to the fact that the assumption of homogeneity of variances was retained (F = 2.28, *P* > .05). The results of Table 7, Independent Samples t-test; (t (78) = .552, *P* > .05, Cohen’s d = .122 representing a weak effect size) indicated that there was not any significant difference between the two groups’ means on pretest of essay writing. Thus, it can be concluded that the two groups were homogeneous in terms of their essay writing ability prior to the administration of the treatment. Figure 2 shows the two groups’ means on pretest. As it was discussed above, both groups had almost the same means on pretest of essay writing.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 7. Independent-Samples t-test for Pretest of Essay Writing by Groups** | | | | | | | | | |
|  | **Levene's Test for Equality of Variances** | |  |  |  | **t-test for Equality of Means** | |  |  |
|  | **F** | **Sig.** | **T** | **Df** | **Sig. (2-**  **tailed)** | **Mean Difference** | **Std. Error Difference** | **95% Confidence Interval of the Difference** | |
|  |  |  |  |  |  |  |  | **Lower** | **Upper** |
| Equal variances assumed | 2.285 | .135 | .551 | 78 | .583 | .144 | .261 | -.376 | .663 |
| Equal variances  not assumed |  |  | .551 | 74.144 | .583 | .144 | .261 | -.376 | .663 |

**Figure 2. Means on Pre-test of Essay Writing by Groups**

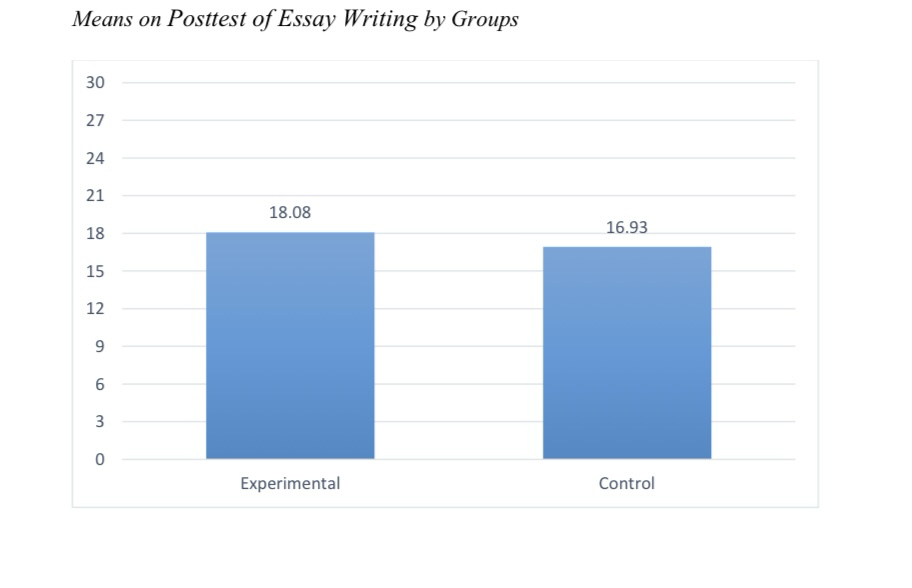
The experimental and control groups’ means on posttest of essay writing were compared through Independent-Samples t-test in order to probe the only null- hypothesis raised in this study. As shown in Table 8 the experimental group (M = 18.08, SD = 1.32) had a higher mean than the control group (M = 16.39, SD = 1.29) on posttest of essay writing.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Table 8. Descriptive Statistics for Posttest of Essay Writing by Groups** | | | | | |
| **Group** | |  | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| Experimental | |  | 40 | 18.08 | 1.325 | .209 |
| Posttest | |  |  |  |  |  |
| Control | |  | 40 | 16.93 | 1.297 | .205 |

Table 9 provides the results of the Independent-Samples t-test for posttest scores. This study reports the results in two rows. The results displayed across the first row; i.e., “Equal variances assumed” were reported due to the fact that the assumption of homogeneity of variances was retained (F = .027, p > .05).

The results of Independent Samples t-test; (t (78) = 3.92, *P* < .05, Cohen’s d = .892 representing a large effect size) indicated that the experimental group significantly outperformed the control group on posttest of essay writing. Thus, the null-hypothesis as “AI-generated corrective feedback and teacher-generated corrective feedback did not have significantly different effects on the essay writing skills of Iranian intermediate EFL learners” was rejected. Figure 3 shows the two groups’ means on posttest. As it was discussed above, the experimental group had a higher mean than the control group on posttest of essay writing.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 9. Independent-Samples t-test for Posttest of Essay Writing by Groups** | | | | | | | | | |
|  | **Levene's Test for Equality of Variances** | |  |  |  | **t-test for Equality of Means** | |  |  |
|  | **F** | **Sig.** | **T** | **Df** | **Sig. (2-**  **tailed)** | **Mean Difference** | **Std. Error Difference** | **95% Confidence Interval of**  **the Difference** | |
|  |  |  |  |  |  |  |  | **Lower** | **Upper** |
| Equal variances assumed | .027 | .869 | 3.923 | 78 | .000 | 1.150 | .293 | .566 | 1.734 |
| Equal variances  not assumed |  |  | 3.923 | 77.966 | .000 | 1.150 | .293 | .566 | 1.734 |

**Figure 3. Means on Posttest of Essay Writing by Groups**

4. DISCUSSION

The present study compared the effect of AI-generated Corrective Feedback versus Teacher-generated Corrective Feedback on essay writing skills of Iranian intermediate EFL learners. The findings of this study indicate that AI-generated Corrective Feedback significantly improved essay writing skills among Iranian intermediate EFL learners. The findings align with prior studies (e.g., Wang, 2024; Evenddy, 2024) that highlight the effectiveness of AI tools in delivering immediate and consistent feedback on writing accuracy. AI tools offer quicker and more tailored feedback than traditional methods, helping learners improve their grammar, vocabulary, and writing proficiency more efficiently (Evenddy, 2024). Similarly, Zou et al. (2023) found that AI-powered tools significantly enhanced speaking and writing abilities by providing more diverse and practical feedback to learners, leading to improvements in their test scores. This study supports the notion that AI can offer effective, diverse feedback to enhance language learning, not only in writing but also in speaking. Furthermore, Rad et al. (2023) showed that AI tools, like Wordtune, were beneficial in improving students’ writing by offering valuable feedback that helped learners understand the corrections and apply them effectively. This aligns with the present study, reinforcing the idea that AI tools are instrumental in enhancing writing performance, providing immediate and user-friendly feedback. Moreover, recent research in an Indonesian EFL context revealed that students perceived AI-based tools positively in writing classes, reporting benefits in terms of idea generation, writing structure, and vocabulary development (Sumakul et al., 2022). However, the findings of this study contradict the results of Han (2021), whose research revealed that teacher feedback was more effective than AI feedback in improving essay revisions. Han’s study suggested that the motivating, encouraging language used in teacher feedback had a more significant impact on students, leading to higher scores in essay revision. Conversely, participants who received online automated feedback expressed frustration with the AI tools, particularly the Scoring Network, citing difficulties in understanding the feedback they received. These differences could stem from the varying educational contexts or the specific nature of the tasks involved, such as the complexity of the writing assignments or the level of learners’ language proficiency. Moreover, the findings are inconsistent with those of Escalante et al. (2023), who evaluated the effectiveness of AI tools in providing writing feedback to English as a Native Language (ENL) learners. While students appreciated the speed and clarity of AI feedback, they found human feedback more personalized, particularly for addressing nuanced issues like tone and style. This suggests that while AI can provide effective feedback for simpler corrections, it may struggle with more complex writing issues, which may explain the differences observed in this study compared to Escalante et al.’s findings. Additionally, the study is inconsistent with Taşkıran and Gökşel’s (2022) research, which found that teacher-provided feedback led to higher writing scores compared to automated feedback. Students in their study reported enjoying the teacher feedback more, suggesting that the interpersonal interaction and motivational aspects of teacher feedback may play a crucial role in students’ writing improvements.

5. Conclusion

The goal of the current study was to determine the effect of AI-generated versus teacher-generated corrective feedback on Iranian intermediate EFL learners’ essay writing performance. The overall results indicate that AI-generated corrective feedback was effective in enhancing EFL learners’ writing skills, providing immediate and consistent corrections that helped students improve their writing accuracy. The findings of this study may enhance EFL learners’ writing competence by providing insights into the relative effectiveness of automated versus traditional teacher-generated feedback. The practical and theoretical implications of this study are significant for the EFL/ESL context, particularly in the pedagogy of writing instruction. From a theoretical perspective, this study contributes to existing literature by highlighting the advantages of AI-generated feedback in providing immediate and consistent corrections. It also calls for further development of a comprehensive model to integrate AI-assisted feedback into L2 writing instruction. Moreover, this study supports the idea of a balanced approach to feedback, where AI tools complement teacher-generated feedback to optimize learning outcomes, aligning with theories like Ellis’s interaction hypothesis, which stresses the importance of both immediate correction and the interpersonal aspects of feedback. From a practical standpoint, language instructors can integrate AI-generated corrective feedback into writing courses to enhance students’ engagement and autonomy in the revision process. AI-powered tools can be employed alongside teacher feedback to create a more efficient and scalable feedback system, allowing students to receive timely corrections while still benefiting from human guidance on higher-order writing skills, such as coherence and argumentation. Moreover, this approach could reduce teachers’ workload while enabling students to take greater responsibility for their learning. This study also has implications for curriculum designers and material developers, who may want to integrate automated-feedback systems into writing courses, especially in digital learning environments. For language educators and policymakers, the findings suggest the potential for large-scale implementation of AI feedback systems, making high-quality writing instruction more accessible to learners and enabling more flexible, self-directed learning experiences. This could have a transformative effect on writing instruction by extending beyond the classroom, allowing students to engage with AI-supported feedback outside of class hours. Despite its contributions, this study had some limitations. One limitation was the failure to control individual differences among participants, such as motivation, cognitive styles, and prior writing experience, which could have influenced the results. Another limitation was the non-random selection of participants which limits the generalizability of the findings. Future research could address these limitations by exploring the long-term impact of AI-generated feedback on writing proficiency across different levels of language learners. Additionally, investigating students’ perceptions and engagement with AI-assisted feedback tools in diverse learning environments would provide deeper insights into how these tools can best support writing development.

**DISCLAIMER**

Authors hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References

Allen, D. (2004). Oxford Placement Test. Oxford University Press.

Altun, M. (2023). Developing writing abilities in L2: A review. *International Journal of Social Sciences and Educational Studies, 10*(2), 175-177. <https://doi.org/10.23918/ijsses.v10i2p175>

Arnold, K. M., Umanath, S., Thio, K., Reilly, W. B., McDaniel, M. A., & Marsh, E. J. (2017).Understanding the cognitive processes involved in writing to learn. *Journal of Experimental Psychology:Applied, 23*(2), 115–127. <https://doi.org/10.1037/xap0000119>

Bagheri Nevisi, R., & Arab, N. (2023). Computer-Generated vs. Direct Written Corrective Feedback and Iranian EFL Students’ Syntactic Accuracy and Complexity: A Multi-Method Probe into their Effects and Students’ Perceptions. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 42*(2),111-148. <https://doi.org/10.22099/tesl.2023.46955.3177>

Bitchener, J. and Knoch, U. (2008) The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409-431.

<http://dx.doi.org/10.1177/1362168808089924>

Borna, P., Mohammadi, R., & Karimi Nia, R. (2024). Investigating the effect of AI writing assistance tools on Iranian intermediate EFL learners’ writing performance: A comparative study of ProWritingAid and Grammarly. *Research in English Language Pedagogy, 12*(3), 478–504. <https://doi.org/10.30486/RELP.2024.897162>

Calver, M. C., & Tweedley, J. R. (2016). A technique for expediting comprehensive written feedback on assignments. *The American Biology Teacher, 78*(8), 684–686. <https://doi.org/10.1525/abt.2016.78.8.684>

Derakhshan, A., & Shirejini, R. K. (2020b). An investigation of the Iranian EFL learners’ perceptions towards the most common writing problems. *SAGE Open, 10*(2), 215824402091952. <https://doi.org/10.1177/2158244020919523>

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal, 1*(1), 3–18. <https://doi.org/10.5070/L2.V1I1.9054>

Emidar, Indriyani, V., & Ningsih, A. G. (2023). The Effect of Digital Literacy and Writing Skills on The Skills of Developing Teaching Materials for Prospective Teacher Students. *JMKSP (Jurnal Manajemen, Kepemimpinan, Dan Supervisi Pendidikan), 8*(2), 666-682. <https://doi.org/10.31851/jmksp.v82.11457>

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), 1-20. <https://doi.org/10.1186/s41239-023-00425-2>

Evenddy, S. S. (2024). Investigating AI’s automated feedback in English language learning. *Foreign Language Instruction Probe, 3*(1), 76–87. <https://doi.org/10.54213/flip.v3i1.401>

Geiser, S., & Studley , R. (2001). UC and the SAT: Predictive validity and differential impact of the SAT and SAT II at the universe of California. *Educational Assessment, 8*(1), 1-26 <https://doi.org/10.1207/S15326977EA0801_01>

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology, 26*(2), 5-24. <https://doi.org/10125/73474>

Han, F., & Wang, Z. (2021). Comparison of English teacher feedback and automated writing feedback on the quality of English language learners’ essay revision. *Open/Technology in Education, Society, and Scholarship Association Conference, 1*(1), 1–5. <https://doi.org/10.18357/otessac.2021.1.1.44>

Hyland, K. (2003). *Second language writing*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511667251>

Hyland, K., & Hyland, F. (2006). Feedback on second language students’ writing. *Language Teaching, 39*(2), 83–101. <https://doi.org/10.1017/S0261444806003399>

Kuteeva, M., & Andersson, M. (2024). Diversity and standards in writing for publication in the age of AI—Between a rock and a hard place. *Applied Linguistics, 45*(3), 561–567. <https://doi.org/10.1093/applin/amae025>

Luo, M., Hu, X., & Zhong, C. (2025). The collaboration of AI and teacher in feedback provision and its impact on EFL learner’s argumentative writing. Education and Information Technologies. <https://doi.org/10.1007/s10639-025-13488-7>

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1-40. <http://dx.doi.org/10.1017/S0261444812000365>

Muchemwa, S., Amimo, C., & Allida, V. (2019). Analysis of teacher-written corrective feedback in second language writing pedagogy in secondary schools, Bulilima, Zimbabwe. *Journal of Education, Society and Behavioural Science, 31*(3), 1–13. <https://doi.org/10.9734/jesbs/2019/v31i330150>

Rad, H. S., Alipour, R., & Jafarpour, A. (2023b). Using artificial intelligence to foster students’writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments, 32*(9), 5020–5040. <https://doi.org/10.1080/10494820.2023.2208170>

Raimes, A. (1983). *Techniques in teaching writing.* Oxford University Press.

Shi, Y., & Deng, B. (2023). Exploring the optimal communication delay for AI feedback tools.*Information Processing & Management, 61*(2), 103572. <https://doi.org/10.1016/j.ipm.2023.103572>

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students’ writing. *Learning and Instruction, 91*, 101894-101894 <https://doi.org/10.1016/j.learninstruc.2024.101894>

Sumakul, D. T. Y. G., Hamied, F. A., & Sukyadi, D. (2022). Students' perceptions of the use of Al in a writing class. In Proceedings of the 67th TEFLIN International Virtual Conference & the 9th ICOELT 2021 (TEFLIN ICOELT 2021) (pp. 52-57). Atlantis Press.

<https://doi.org/10.2991/assehr.k.220201.009>

Taskiran, 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>

Wang, D. (2024b). 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>

Vosough-Yazdani, M. (2018). *Think to Write, Write to Think*. Arshad Sepahan Press.

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting Speaking Practice by Social Network-Based Interaction in Artificial Intelligence (AI)-Assisted Language Learning. Sustainability, 15(4), 2872. <https://doi.org/10.3390/su15042872>