



PROCEEDINGS

International Conference on Education
Faculty of Tarbiyah | UIN Syekh Wasil Kediri
E-ISSN: 3025-6828

Volume 4, 2026

The Effectiveness of Using ChatGPT in Teaching Writing Skills for Vocational High School Students

Ignosia Adi Irawan,¹; Sri Wahyuni^{2*};

^{1, 2} English Education Department, Faculty of Tarbiyah and Teacher Training
State Islamic University Syekh Wasil Kediri

Correspondence e-mail: sriwahyuni@uinkediri.ac.id;

Abstract:

The present study aims to examine the effectiveness of ChatGPT in teaching writing skills for vocational high school students. This study employed a quantitative research method with a quasi-experimental design to investigate the effectiveness of ChatGPT in teaching writing skills. The subjects of this study were the eleventh-grade students of SMK AL Huda Kota Kediri. The sample consisted of two classes: the experimental group, which used ChatGPT during the learning process, and the control group, which used peer-correction. The data were collected through writing tests in the form of pre test and post-test. The students' writing was assessed based on content, organization, grammar, vocabulary, and mechanics. The collected data were analyzed using IBM SPSS Statistics 27, due to the data were not normally distributed, Mann Whitney U test was used. The results revealed a significant difference in writing achievement between the experimental and control groups. The significance value of the Mann-Whitney U test was 0.036 ($p < 0.05$), indicating that the use of ChatGPT had a positive effect on students' writing skills. The experimental group achieved higher post-test scores compared to the control group. This finding indicates that students who used ChatGPT improved their writing more effectively through instant feedback, corrections, and suggestions provided during the writing process. In conclusion, the study highlights that ChatGPT is effective at enhancing students' writing skills and can serve as a supportive learning tool in English writing classes at vocational high school level.

Keywords: Writing skills; Recount text; AI in education; ChatGPT; peer correction

Introduction

Writing is a crucial skill in learning the English language. Writing is an English language skill that the students must learn because writing activity can improve their thinking skills and help them to develop other language skills (Mustofa & Wahyuni, 2024). In addition, it also can help students to improve their creativity. Writing is recognized as an essential educational ability that is important in both academic and professional contexts (Wahyuni, Fahruriza, & Ningrum, 2025). Teaching writing requires an organized and systematic approach, and not just by putting words together. Writing involves not only grammar and vocabulary but also the organization of ideas in a clear and logical way. As Harmer (2004) points out, writing is a process that includes planning, drafting, revising, and editing, making it a skill that needs time and guidance to develop. Teachers play a crucial role in helping students through this process, offering feedback and support to improve both clarity and expression.

Furthermore, a common students' problem in learning writing skill is the lack of vocabulary and creativity, which limits their ability to express ideas clearly and widely. Without a solid vocabulary knowledge, students tend to repeat words and phrases, making their writing monotonous and less engaging. Based on an observation results in SMK AL-Huda Kota Kediri, the main challenges in learning writing skill were students' limited vocabulary and difficulties in grammar/tenses. As mentioned before, the lack of vocabulary knowledge often leads to repetitive writing and limited expression, making it hard for students to communicate their ideas effectively. As Nation (2009) points out, a rich vocabulary is essential for successful writing, as it allows students to convey their thoughts more precisely and with greater depth. Additionally, many students struggle with using correct tenses, which impacts on their writing. According to Azar (2002), mastering verb tenses is crucial for students to write coherent sentences and avoid confusion in their writing. Therefore, writing becomes the most challenging and complex skill to master (Wahyuni & Umam, 2017; Anggunsari & Wahyuni, 2023).

The presence of AI like ChatGPT is really helping for students enhancing their writing skills, because AI like ChatGPT provides feedback, and gives individual learning needs for students. Moreover, some studies believe that AI can help students improve their academic skills, have a deeper understanding of subject matter, knowledges, and improve problem solving skills (Alenezi, 2023). Although, the role of AI in classroom might give a big impact when its deployed, for example the students might get more interested in learning or vice versa, the students might get to attached on the AI or the gadget and the interactions between student and teacher might disturbed over time. Harry et al (2023), highlights the benefits and the potential of using AI in education such as the potential in personalize learning, improving student engagement etc. According to Gupta et al (2024), AI technologies could remake the way traditional method of teaching and student-teacher interactions, it also has a lot of positive impact such as time efficiency, personalized learning, and access to resources for learning subject. Furthermore, according to Alsaedi (2024), AI provides individual learning which is important for students to improve their academic writing skills.

This research aims to investigate the effectiveness of using AI in the classroom for academic writing, whether it is effective or not effective. The researchers examine how AI tool ChatGPT impact students' writing skills, focusing on aspects such as grammar, vocabulary, and structure. The research question is formulated as follow "Is using ChatGPT effective in teaching writing skills to vocational high school students?"

Methods

This study utilized a quasi-experimental research design to evaluate the impact of using ChatGPT on students' writing of recount texts. Both a pre-test and a post-test were conducted for the experimental group and the control group. As noted by Cook (2015), quasi-experiments often assess the causal effects of enduring interventions outside of laboratory settings to determine significant differences. Quasi-experimental research is carried out when researchers cannot randomly select a sample from the population. Subjects are assigned to different groups, each receiving different treatments. This design was appropriate given the natural classroom setting, where randomization was not feasible due to administrative and logistical constraints (Creswell & Creswell, 2018).

Two classes consisting from the eleventh grade students of a public vocational high school in Kediri participated in this research. The experimental group uses ChatGPT as a feedback source, and the control group uses peer feedback source. The research instrument was writing test: pretest and post-test. Students were asked to write a recount text based on the theme given by the researcher. The students' writing was scored using the writing rubric developed by Brown (2007) which consists of five components: content, organization, grammar, vocabulary and mechanics. Then the data collected were analyzed using ANCOVA.

Findings

The findings of this study were divided into some subsections. Those are the pre-test result, the post-test results, assumptions testing, hypothesis testing.

1. The Pre-Test Result

The pre-test was administered to students in experiment and control groups to know their initial writing skill before the treatment. Moreover, the result of pre-test was analyzed using descriptive statistics in SPSS 27. The following are the complete results of the pre-test.

Table 1 Pre-test Scores of Experimental Group

Group	N	Mean	Min.	Max.	SD
Experimental	31	48.3548	26	78	15.23493
Control	21	49.2857	24	77	15.94096

From table 1, it can be seen that mean scores of both groups, experimental and control groups are still relatively in low level. Although the mean score of the control group (49.3) is higher than the experimental group (48.4), those are almost in similar level. This condition is also same with the minimum dan maximum scores of both groups. The highest score was 78 for the experimental group and 77 for the control group. Meanwhile, the lowest score was 26 for the experimental group and 24 for the control group. In addition, the standard deviation was 15.23 for the experimental group, and 15.94 for the control group, indicating the initial students' writing ability are varied. Overall, the pre-test score from both groups showed that the students' initial writing skill was in low level.

2. The Post-Test Result

The post-test was administered to students in the experimental and control groups after the treatment where students from experimental group used ChatGPT while control group used peer-correction method in order to review their writing in teaching learning process. Table 2 presents the results of post-test from both groups.

Table 2 The Post-test Scores

Group	N	Mean	Min.	Max.	Std. Deviation
Experimental	31	74.0968	60	83	4.47478
Control	21	68.5238	51	82	10.03304

From Table 2 above, it can be seen that the mean score of the experimental group (74.09) was higher than the mean score of the control group (68.52). The experimental group mean score was indicating students' writing ability considerably in high level, then for the control group mean score was in moderate level. In addition, the standard deviation of the experimental group was 4.47, indicating the students' writing scores were spread in low range around the mean score. Furthermore, the results showed that even the mean score was not very high, the data distribute around the mean score. This is indicating the post-test score are not varied. In contrast, the standard deviation of the control group was 10.033, indicating the post-test score were kind of vary among individuals. Overall, the result of the post-test score from both groups showed slight difference comparing with pre-test score.

3. The Assumption Testing

Before employed the main statistical testing which is ANCOVA (Analysis of Covariance), researcher had conducted some assumption testing such as the normality test, homogeneity of variance, linearity and homogeneity of regression slopes. These tests are prerequisite before conducting the main statistical analysis. The assumption test ensures the validity and accuracy of findings in this study. Moreover, if the assumption testing is not fulfilled, researcher have to conduct non-parametric testing such as Mann-Witney U test.

a). Normality Test

Normality of the data must be checked first. This test is important to know whether the data are normally distributed or not, because parametric tests like ANCOVA require normal data distribution. In this study, the Shapiro-Wilk test was used to test the normality of the residual data with a 0.05 level of significance. If the significance value is higher than 0.05, it means the data are normally distributed.

Table 3 Normality Test Tests

		Shapiro-Wilk		
	group	Statistic	df	Sig.
pretest	Experimental	.940	31	.081
	Control	.950	21	.335
posttest	Experimental	.951	31	.165
	Control	.888	21	.020

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on table 2, the results show that the pre-test data of the experimental group ($p = 0.081$) and the control group ($p = 0.335$) were normally distributed ($p > 0.05$). In addition, the post-test of the experimental group ($p = 0.165$) also met the normality assumption. However, the post-test result of the control group ($p = 0.020$) did not follow a normal distribution. From these results, it can be concluded that most of the data were normal, but there is one group that does not meet the normality assumption.

b). Homogeneity of Variance Test

Before the main statistical analysis is carried out, the homogeneity of variance was conducted. This test was to check whether the experimental and control groups have similar variances, because parametric tests like t-test or ANCOVA assume equal variance between groups. In this research, Levene's Test is used with a significance level of 0.05. A value higher than 0.05 means the variances are equal, while a value lower than 0.05 shows that the variances are not equal.

Table 4 Homogeneity of Variance Test

Levene's Test for Equality of Variances			
F			Sig.
pretest	Equal variances assumed	.048	.827
	Equal variances not assumed		
posttest	Equal variances assumed	14.787	.000
	Equal variances not assumed		

The homogeneity of variance was conducted using Levene's Test. In the pre-test, the significance value was 0.827, which is higher than 0.05. This showed that the variance of the experimental and control groups was equal. On the other hand, the post-test result showed a significance value of 0.000, which was lower than 0.05, indicating the variance between the two groups was not equal. From these results, the assumption of homogeneity was fulfilled in the pre-test but not fulfilled in the post-test.

c). Linearity Test

Linearity test is important to determine whether there is a linear relationship between the covariate and the dependent variable. In this study, the linearity test was conducted using Curve Estimation in SPSS with a significance level of 0.05. If the significance value is lower than 0.05, it means that the relationship between the variables is linear.

Table 5 Linearity Test ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	395.941	1	395.941	7.594	.008
Residual	2606.829	50	52.137		
Total	3002.769	51			

The independent variable is pretest.

Based on the table above, the significance value was 0.008, which is lower than 0.05. This means that **there was a linear relationship between the pre-test and post-test scores**. Therefore, the assumption of linearity was fulfilled.

d). Homogeneity of Regression Slopes

Before conducting the main hypothesis testing, the homogeneity of regression slopes assumption was conducted. This test is used to check whether the relationship between the covariate and dependent variable was the same in each group. In this research, the test was conducted by analyzing the interaction between pre-test scores and the group variable through the Test of Between-Subjects Effects in SPSS with a significance level of 0.05. The result of the test can be seen in the following table.

Table 6 Homogeneity of Regression Slopes Test Tests of Between-Subjects Effects

Dependent Variable: posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1235.697 ^a	3	411.899	11.189	.000
Intercept	16465.506	1	16465.506	447.262	.000
pretest	550.974	1	550.974	14.966	.000
group	665.101	1	665.101	18.067	.000
group * pretest	426.668	1	426.668	11.590	.001
Error	1767.072	48	36.814		
Total	271420.000	52			
Corrected Total	3002.769	51			

a. R Squared = .412 (Adjusted R Squared = .375)

Based on the result of the test, the significance value of the interaction between group and pretest was 0.001. Since the value is lower than 0.05, it means there was a significant interaction between the variables. Therefore, the assumption of homogeneity of regression slopes was not achieved, which means the data did not completely meet the assumption required for ANCOVA analysis.

4. The Hypothesis Testing

Because the assumptions for parametric tests were not fully satisfied, such as normality and homogeneity of data, a non-parametric test was used in this study. The researchers chose the Mann–Whitney U test to analyze the difference between the post-test scores of the experimental and control groups.

Table 4.9 Mann-Whitney U Test Test Statistics^a

	posttest
Mann-Whitney U	213.500
Wilcoxon W	444.500
Z	-2.094
Asymp. Sig. (2-tailed)	.036

a. Grouping Variable: group

The result of the Mann–Whitney U test showed that the significance value (Asymp. Sig. 2-tailed) was 0.036, which is lower than 0.05. This means that there was a significant difference between the two groups in the post-test results. Moreover, the experimental group obtained a higher mean rank (30.11) compared to the control group (21.17), indicating better achievement after the treatment. From these findings, it can be concluded that there is a significant difference between the experimental and control groups. The experimental group showed better performance than the control group, which means that the treatment used in this study had a positive effect on students' learning outcomes.

Discussion

This study focused on examining the effectiveness of using ChatGPT in teaching students' writing skills at a vocational high school level compared to the peer correction method. The result of the hypothesis testing showed that the significance value was 0.036, which is lower than 0.05. This means that there was a significant difference between the two groups in the post-test results. Moreover, the experimental group had a higher mean rank (30.11) than the control group (21.17). Based on this result, it can be said that students who learned using ChatGPT achieved better writing performance compared to those who learned through peer-correction. ChatGPT is more effective in improving students' writing skills. The better performance in the experimental group may be because ChatGPT provides instant feedback, correction, and suggestions during the writing process. These features help students identify their mistakes quickly and revise their writing more effectively, which leads to better writing results.

Previous studies have also shown similar results about the use of AI tools, especially ChatGPT, in improving students' writing skills. Wu et al. (2022) stated that AI writing support can improve students' writing accuracy because it gives instant feedback and helps students to correct their mistakes continuously. This finding is similar to the result of this study, where students in the experimental group achieved better performance after using ChatGPT. In the same way, Li and Sun (2021) also found that students who learn with AI tools perform better in writing than those who use traditional learning methods.

Furthermore, Zhang and Liu (2020) explained that AI can help students improve the organization of ideas and coherence in writing tasks. This idea can also be connected to the improvement seen in the experimental group in this research. Besides that, Huang

et al. (2023) pointed out that the use of AI makes students more confident and reduces their writing anxiety, which may also be one of the reasons why students in this study showed better writing achievement after the treatment. Overall, these previous studies are consistent with the findings of this research, showing that AI tools such as ChatGPT can give positive effects on students' writing skill development.

Furthermore, the effectiveness of ChatGPT in this study can be understood through scaffolded learning and feedback theory. ChatGPT works as a tool that supports students in the writing process by giving help such as explanations, examples, corrections, and suggestions. This idea is related to Vygotsky's (1978) Zone of Proximal Development (ZPD), which explains that students can do better when they get support from someone or something more capable. In addition, Wood, Bruner, and Ross (1976) describe scaffolding as temporary help that guides learners to complete tasks they cannot finish alone at the beginning. Because of this, ChatGPT can be seen as digital scaffolding that helps students improve their writing step by step. Besides that, feedback theory also explains the role of ChatGPT. Hattie and Timperley (2007) state that feedback is important for improving learning results, especially when it is given clearly and quickly. In this study, ChatGPT gives direct feedback while students are writing, so they can quickly see their mistakes and revise their work. This process helps students reduce errors and improve their writing in a more structured way.

Besides that, ChatGPT also makes the learning process more interactive. Students can revise their writing many times and directly see the improvements. This process encourages students to learn independently and correct their own mistakes. Compared to peer-correction, ChatGPT gives more consistent and detailed feedback, which may be the reason why the experimental group achieved

Conclusion

Based on the result of the data analysis, it can be concluded that there is a significant difference between the experimental group and the control group in students' writing achievement after the treatment. The result of the Mann-Whitney U test showed that the significance value was 0.036, which is lower than 0.05. This indicates that the experimental group, which was taught using ChatGPT, performed better than the control group that was taught using peer-correction. Moreover, the experimental group obtained a higher mean rank (30.11) compared to the control group (21.17), indicating better achievement after the treatment. The use of ChatGPT is effective in improving students' writing skills at vocational high school level. ChatGPT helps students by giving instant feedback, suggestions, and corrections during the writing process. This makes students easier to understand their mistakes and improve their writing quality. Compared to peer-correction, ChatGPT provides more consistent and immediate support, which leads to better writing performance. Furthermore, the use of ChatGPT also increased students' confidence and motivation in learning writing. Many students felt more comfortable during the writing process because they could receive immediate assistance whenever they faced difficulties. Compared to peer-correction, ChatGPT provided more consistent, detailed, and immediate support, which encouraged students to revise their writing more actively and independently. As a result, students became more engaged in the learning process and demonstrated better writing performance after the treatment.

References

- Anggunsari, P. & Wahyuni, S. 2023. Direct Written Corrective Feedback for Tenth Graders Recount Text: Adequate Practice to Boost Sentential Accuracy. *JET (Journal of English Teaching)*, 9(3): 310-322.
- Alenezi, A. (2023). Teacher perspectives on AI-driven gamification: Impact on student motivation, engagement, and learning outcomes. *Information Technologies and Learning Tools*, 97(5), 138–148.
- Alsaedi, B. (2024). AI in learning English: Enhancing English skills through the use of AI-powered language learning tools. *International Journal of Education and Social Science Research*, 7(5), 82–94.
- Basha, J. Y. (2024). The negative impacts of AI tools on students in academic and real-life performance. *International Journal of Social Sciences and Commerce*, 3, 1–16.
- Cook, T. D. (2015). Quasi-experimental design. In *Wiley encyclopedia of management* (pp. 1–2). Wiley.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Fazel, I., & Ahmadi, A. (2011). On the relationship between writing proficiency and instrumental/integrative motivation among Iranian IELTS candidates. *Theory and Practice in Language Studies*, 1(7), 747–757.
- Gupta, R., Nair, K., Mishra, M., Ibrahim, B., & Bhardwaj, S. (2024). Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda. *International Journal of Information Management Data Insights*, 4(1), Article 100232.
- Harmer, J. (2004). *How to teach writing*. Pearson Education Limited.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Huang, X., Chen, Y., & Zhang, L. (2023). The role of AI in reducing writing anxiety among high school EFL learners. *Journal of Applied Language Studies*, 18(2), 220–235.
- Hui, Z., Qian, Y., & Zheng, Q. (2023). The influence of ChatGPT participation in teaching on students' learning interest. *Lecture Notes in Education Psychology and Public Media*, 25(1), 166–171. <https://doi.org/10.54254/2753-7048/25/20230704>
- Lentez, A., & Mager, G. (2023). Exploring the potential of ChatGPT in enhancing user experience (UX) writing. In *Proceedings of the 10th International Conference on Human Interaction and Emerging Technologies (IHET 2023)*. <https://doi.org/10.54941/ahfe1004030>
- Li, J., & Sun, Y. (2021). Exploring the impact of AI-based tools on EFL learners' writing performance: A quantitative analysis. *International Journal of Language Learning and Technology*, 14(4), 113–128.
- Mustofa, A.Y., & Wahyuni, S. (2024). "Whisper and Write Game" as Pre-Writing Activity to Enhance Descriptive Writing Proficiency of Indonesian Seventh Graders. *JALL (Journal of Applied Linguistics and Literacy)*, 8(1), 61-70. <http://dx.doi.org/10.25157/jall.v8i1.12738>
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Routledge.
- Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences*, 13(10), Article 6039. <https://doi.org/10.3390/app13106039>
- Shidiq, M. (2023). The use of artificial intelligence-based ChatGPT and its challenges for the world of education: From the viewpoint of the development of creative writing skills. *Proceedings of the International Conference on Education, Society and*

Humanity, 1(1).

- Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*. <https://doi.org/10.1177/20965311231168423>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wahyuni, S., & Umam, M.K. (2017). An Analysis on Writing Anxiety of Indonesian EFL College Learners. *Journal of English Education and Linguistics Studies (JEELS)*, 4 (1), 105-128. <http://dx.doi.org/10.30762/jeels.v4i1.333>
- Wahyuni, S., Fahrurizha, S.E., Ningrum, A.S.B. (2025). Enhancing Students' Writing Skill Through Threads Application in SILC (Skapsa International Language Class) at Kediri. *Proceeding of International Conference on Education*. 3, 495-504. <https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/proceedings/article/view/6710>
- Widayanti, T., Rustyana, N., & Haryudin, A. (2022). Students' perception in writing procedure text. *PROJECT (Professional Journal of English Education)*, 2(5), 687–691.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.
- Wu, L., Lin, Z., & Wang, H. (2022). Enhancing EFL writing through AI: The effect of ChatGPT on student writing accuracy and fluency. *Computer-Assisted Language Learning*, 35(3), 345–360.
- Zhang, Y., & Liu, Q. (2020). Fostering creativity in EFL writing: The role of AI-powered tools in enhancing student creativity and expression. *Language Teaching Research*, 22(5), 425–442.