



Enhancing EFL Students' Reading Comprehension through ChatGPT-Assisted Instruction: A Quasi-Experimental Study

Adelia Citra Ramadani¹, Bradiansyah Tri Suryanto²

^{1,2} Pendidikan Bahasa Inggris, Fakultas Sosial dan Humaniora, Universitas Nurul Jadid
adeliacitraramadaniiii@gmail.com, brady@unuja.ac.id

Abstract

The use of ChatGPT in English language learning has created opportunities for providing interactive and individualized instructional support. However, empirical evidence regarding its effectiveness in improving the reading comprehension of junior high school EFL students in Indonesia remains limited. This study aimed to examine the effectiveness of ChatGPT-assisted instruction in improving the reading comprehension of eighth-grade students at SMPN 1 Paiton. A quantitative approach with a quasi-experimental design was employed. The participants consisted of 54 eighth-grade students, with 28 students in the experimental group and 26 students in the control group. Data were collected through pre-tests and post-tests consisting of 20 multiple-choice reading comprehension items. The data were analyzed using descriptive statistics, the Shapiro-Wilk normality test, Levene's homogeneity test, paired sample t-tests, an independent sample t-test, and N-Gain analysis. The results showed that the experimental group improved from a pre-test mean of 50.96 to a post-test mean of 85.77, while the control group improved from 49.46 to 72.14. The paired sample t-test showed significant improvement within both groups ($p < 0.05$). The experimental group also achieved a higher N-Gain score of 72.71, categorized as high, compared with 46.37, categorized as medium, in the control group. However, the independent sample t-test yielded $p = 0.117$, indicating that the post-test difference between the groups was not statistically significant at the 0.05 level. These findings suggest that ChatGPT-assisted instruction was associated with greater descriptive improvement in reading comprehension, although further research with larger samples and longer interventions is needed.

Keywords: ChatGPT, Reading Comprehension, EFL Students, Quasi Experimental

1. Introduction

The rapid advancement of Artificial Intelligence (AI) has significantly transformed educational practices, particularly in English as a Foreign Language (EFL) learning. Among various AI technologies, ChatGPT has emerged as one of the most widely adopted generative AI applications because it provides immediate feedback, personalized explanations, interactive communication, and adaptive learning support. These features enable students to access learning assistance beyond conventional classroom instruction while encouraging greater engagement and autonomous learning (Wang & Fan, 2025). Consequently, ChatGPT has attracted increasing attention from researchers and educators as an instructional tool capable of supporting language learning. (Kainta & Umarella, 2026), through a meta-analysis, reported that ChatGPT generally has a positive effect on students' learning outcomes, although its effectiveness depends on instructional design and learning context. Likewise, (Balci, 2024) concluded that ChatGPT contributes to learner engagement, motivation, language development, and autonomous learning in EFL education. (Ardiyani et al., 2025) explained that ChatGPT facilitates students' understanding of learning materials by providing contextual explanations, summaries, vocabulary assistance, and interactive feedback. Similarly, (Ali et al., 2025). emphasized that integrating AI into language instruction creates more flexible and student-centered learning experiences. These findings indicate that ChatGPT has considerable potential to improve English language learning when integrated into appropriate pedagogical practices.

Despite these opportunities, reading comprehension remains one of the most challenging language skills for EFL learners. Reading comprehension requires students to identify main ideas, interpret contextual vocabulary, locate detailed information, make inferences, and construct meaning from written texts. However, numerous studies have reported that many students continue to experience difficulties in these aspects. (Mallipa et al., 2024) found that secondary school students demonstrated weaknesses in understanding main ideas, identifying detailed information, and applying effective reading strategies. (Siregar, 2025) similarly reported that limited student participation during reading activities contributes to lower comprehension achievement. (Hidayah et al., 2026) emphasized that reading strategies such as skimming, scanning, and inferencing play an important role in improving students' comprehension of English texts. Furthermore, (Wang & Fan, 2025) argued that reading comprehension requires not only linguistic knowledge but also strategic learning support that enables students to actively construct meaning from texts. These findings are consistent with the preliminary observation conducted at SMPN 1 Paiton, where many eighth-grade students experienced difficulties in identifying the main idea, interpreting unfamiliar vocabulary, understanding explicit and implicit information, and drawing conclusions from English reading

Enhancing EFL Students' Reading Comprehension through ChatGPT-Assisted Instruction: A Quasi-Experimental Study

passages. Reading instruction was still predominantly conducted through teacher explanation, translation, and comprehension exercises, providing limited opportunities for students to receive individualized guidance during the reading process. Therefore, more innovative instructional approaches are needed to support students in developing better reading comprehension. (Duignan et al., 2025)

Recent studies have increasingly investigated the integration of ChatGPT and other AI technologies in English language learning, particularly to improve students' reading performance. (Usman, 2017) reported that ChatGPT promotes independent learning by providing adaptive learning materials and immediate responses to students' questions. Similarly, (Unus & Un, 2025) highlighted that AI-assisted learning environments improve reading comprehension through personalized feedback, vocabulary support, and interactive learning experiences. Balci (2024), through a systematic review, further concluded that ChatGPT positively influences learner engagement, motivation, language development, and autonomous learning, although teachers remain essential in guiding students' use of AI. These findings are supported by (Iqbal Tawakkal & Apriono, 2026) whose meta-analysis demonstrated that ChatGPT generally improves learning outcomes across different educational contexts when integrated with appropriate instructional strategies. At the classroom level, et al. That ChatGPT effectively assisted students in understanding English texts through explanations, summaries, and interactive discussions, while (Riyani, 2025) showed that AI-assisted formative assessment enabled teachers to provide more timely feedback and monitor students' reading progress. Furthermore, (Çelik et al., 2024) revealed that AI-supported text simplification significantly improved students' reading comprehension and inferencing abilities. More recent studies also reinforce these findings. (Irons & Bostanci, 2026) demonstrated that integrating generative AI into reading instruction significantly enhanced students' reading comprehension and learner autonomy, whereas (Unus & Un, 2025) reported that ChatGPT-mediated mind mapping effectively improved EFL learners' reading comprehension by encouraging more organized processing of textual information. Collectively, these studies indicate that ChatGPT possesses considerable potential to support reading instruction; however, its effectiveness depends on learners' characteristics, instructional design, and educational context.

Although previous studies have reported encouraging findings, several important gaps remain. First, most empirical studies have focused on university students, adult EFL learners, or general language learning rather than junior high school students. Second, several investigations have examined students' perceptions of ChatGPT, AI-assisted assessment, or systematic reviews without directly evaluating the effectiveness of ChatGPT-assisted reading instruction in authentic secondary school classrooms. Third, differences in participants, intervention duration, learning environments, and outcome measures make it difficult to generalize previous findings to Indonesian junior high school contexts. Moreover, while previous studies generally reported positive impacts of ChatGPT, concerns regarding students' overreliance on AI, information accuracy, ethical considerations, and the role of teachers continue to be highlighted (Syafuddin et al., 2025). Consequently, empirical evidence examining whether ChatGPT can effectively improve reading comprehension among Indonesian junior high school students through controlled classroom instruction remains limited.

Based on these gaps, the present study offers several contributions to the existing literature. Unlike previous studies that primarily investigated higher education settings, students' perceptions, or AI-assisted assessment, this study focuses specifically on junior high school EFL students in Indonesia using a quasi-experimental design. Furthermore, ChatGPT is employed only during the instructional treatment, whereas the post-test is administered without AI assistance to ensure that the measured outcomes reflect students' actual reading comprehension achievement rather than their ability to use ChatGPT. This approach provides stronger empirical evidence regarding the instructional effectiveness of ChatGPT in improving students' reading comprehension within an authentic classroom setting. Therefore, this study aims to investigate the effectiveness of ChatGPT-assisted instruction in improving the reading comprehension of eighth-grade students at SMPN 1 Paiton. The findings are expected to contribute to the growing body of knowledge on AI-assisted language learning while providing practical implications for English teachers seeking to integrate ChatGPT into reading instruction in a pedagogically meaningful, effective, and responsible manner.

2. Research Method

This study employed a quantitative research approach using a **quasi-experimental method with a Nonequivalent Control Group Design** (Sugiyono, 2020). The research was conducted at SMPN 1 Paiton during the second semester of the 2025/2026 academic year to examine the effectiveness of ChatGPT-assisted instruction in improving students' reading comprehension.

The participants were 54 eighth-grade students selected using purposive sampling. The experimental group consisted of 28 students, while the control group consisted of 26 students. Both classes were selected based on comparable academic characteristics, the same English curriculum, and the teacher's recommendation. Prior to the implementation of the study, permission was obtained from the school, and all participants' data were treated confidentially and used exclusively for research purposes.

The study was conducted in three stages: pre-test, treatment, and post-test. At the beginning of the study, both groups completed the same pre-test to determine their initial reading comprehension ability. The experimental group then received ChatGPT-assisted reading instruction, whereas the control group received conventional reading instruction using teacher explanation, guided reading, vocabulary discussion, and reading comprehension exercises. During the treatment, ChatGPT was integrated into classroom activities to help students identify the main ideas of texts, understand unfamiliar vocabulary, interpret difficult sentences, answer comprehension questions, and receive immediate feedback under the teacher's supervision. Both groups studied the same reading materials, learning objectives, and instructional duration. At the end of the treatment, both groups completed the same post-test without ChatGPT assistance to ensure that the obtained scores reflected students' actual reading comprehension achievement.

The research instrument consisted of a reading comprehension test comprising 20 multiple-choice questions developed according to the eighth-grade English learning objectives. The test measured five reading comprehension indicators: identifying the main idea, locating detailed information, understanding vocabulary in context, making inferences, and drawing conclusions from reading passages. Prior to data collection, the instrument was examined through validity and reliability testing to ensure its suitability for measuring students' reading comprehension.

The collected data were analyzed using IBM SPSS Statistics version 26. Descriptive statistics were employed to summarize students' pre-test and post-test scores. Before testing the research hypothesis, the data were examined using the Shapiro–Wilk normality test and Levene's homogeneity test to verify the assumptions of parametric analysis. The effectiveness of the treatment within each group was analyzed using the paired sample *t*-test, while differences in learning outcomes between the experimental and control groups were analyzed using the independent sample *t*-test at a significance level of 0.05. In addition, the N-Gain score was calculated to determine the magnitude of students' improvement in reading comprehension after the instructional treatment.

3. Findings and Discussion

This section presents the findings of the study concerning the effectiveness of ChatGPT-assisted instruction in improving students' reading comprehension. The findings are presented systematically through descriptive statistics, normality testing, homogeneity testing, paired sample *t*-tests, independent sample *t*-test, and N-Gain analysis. Each analysis is presented in a separate table to provide a clearer description of the students' reading comprehension performance before and after the instructional treatment. The results are subsequently discussed by relating the statistical findings to the learning process and previous studies concerning the use of ChatGPT and artificial intelligence in English language learning.

3.1 Descriptive Statistics

Descriptive statistics were first conducted to provide an overview of the students' reading comprehension performance in the experimental and control groups. The analysis focused on the students' pre-test and post-test scores, including the number of participants, mean scores, and standard deviations. The pre-test results were used to identify the students' initial reading comprehension abilities before the instructional treatment, while the post-test results were used to examine the students' achievement after the treatment.

Table 1. Descriptive Statistics of Students' Reading Comprehension Scores

Group	N	Pre-test		Post-test	
		Mean	± SD	Mean	± SD
Experimental	26	50.96	± 8.249	85.77	± 7.168
Control	28	49.46	± 8.643	72.14	± 9.372

The research method reports 26 students in the control group, while the descriptive-statistics table in the manuscript reports 28. This number should be checked against the original dataset before submission.

The descriptive results show that the experimental and control groups had relatively similar reading comprehension performance before the treatment. The experimental group obtained a pre-test mean score of 50.96, while the control group obtained a mean score of 49.46. The difference between the two pre-test means was relatively small, indicating that the two groups started the instructional process with comparable levels of reading comprehension based on the reported descriptive statistics.

After the instructional process, both groups demonstrated an increase in their reading comprehension scores. The experimental group obtained a post-test mean score of 85.77, representing an increase from its initial mean score of 50.96. Meanwhile, the control group obtained a post-test mean score of 72.14, increasing from its pre-test mean score of 49.46. Therefore, the experimental group demonstrated a larger descriptive improvement than the control group.

The standard deviation results also provide information concerning the distribution of students' scores within each group. The experimental group had a post-test standard deviation of 7.168, while the control group had a post-test standard deviation of 9.372. The lower standard deviation in the experimental group suggests that the students' post-test scores were relatively more concentrated around the group mean. Nevertheless, descriptive statistics alone are insufficient to determine whether the observed differences were statistically significant. Therefore, assumption tests and inferential statistical analyses were subsequently conducted.

3.2 Normality Test

Before conducting parametric hypothesis testing, the normality of the data was examined using the Shapiro–Wilk test. The test was conducted for the pre-test and post-test scores of both the experimental and control groups. A significance value greater than 0.05 indicates that the distribution does not significantly differ from a normal distribution and can therefore be considered normally distributed for the purpose of the analysis.

Table 2. Results of the Shapiro–Wilk Normality Test

Group	Test	Sig.	Decision
Experimental	Pre-test	0.368	Normal
Experimental	Post-test	0.260	Normal
Control	Pre-test	0.224	Normal
Control	Post-test	0.264	Normal

As presented in Table 2, all significance values obtained from the Shapiro–Wilk test were greater than 0.05. The experimental group obtained significance values of 0.368 for the pre-test and 0.260 for the post-test. Similarly, the control group obtained significance values of 0.224 for the pre-test and 0.264 for the post-test.

These results indicate that the pre-test and post-test scores in both groups were normally distributed. Because all significance values exceeded the predetermined significance level of 0.05, there was no statistical evidence indicating a violation of the normality assumption. Consequently, the data were considered appropriate for subsequent parametric statistical analyses, including the paired sample t-test and independent sample t-test.

3.3 Homogeneity Test

In addition to normality testing, a homogeneity test was conducted to determine whether the variances of the experimental and control groups were statistically comparable. The Levene's test was applied to the post-test scores of the two groups. A significance value greater than 0.05 indicates that the variances can be considered homogeneous.

Table 3. Results of the Levene's Homogeneity Test

Variable	Test	Sig.	Decision
Post-test scores	Levene's Test	0.117	Homogeneous

Table 3 shows that the significance value obtained from Levene's test was 0.117. Because this value was greater than 0.05, the variance of the post-test scores between the experimental and control groups can be considered homogeneous. In other words, the assumption of equal variances was satisfied.

The fulfillment of both normality and homogeneity assumptions provided statistical justification for continuing with parametric hypothesis testing. The paired sample t-test was subsequently used to examine whether there was a significant difference between students' pre-test and post-test scores within each group. Meanwhile, the independent sample t-test was used to examine whether the post-test scores differed significantly between the experimental and control groups.

3.4 Paired Sample t-Test

The paired sample t-test was conducted separately for the experimental and control groups. This analysis was intended to determine whether students' reading comprehension scores changed significantly from the pre-test to the post-test after the instructional process.

Table 4. Results of the Paired Sample t-Test

Group	Comparison	Sig. (2-tailed)	Decision	Interpretation
Experimental	Pre-test vs. Post-test	0.000	Significant	Significant
Control	Pre-test vs. Post-test	0.000	Significant	Significant

The results in Table 4 indicate that both groups experienced statistically significant improvements between the pre-test and post-test scores. The experimental group obtained a significance value of 0.000, which is lower than 0.05. Therefore, the null hypothesis stating that there was no difference between the experimental group's pre-test and post-test scores was rejected.

The control group also obtained a significance value of 0.000. This result indicates that the conventional instructional process also produced a statistically significant improvement in students' reading comprehension. Therefore, the improvement observed in the experimental group cannot simply be interpreted as evidence that only ChatGPT-assisted instruction produced learning gains. Both instructional approaches were associated with significant improvements from pre-test to post-test.

However, the descriptive statistics show that the magnitude of improvement was greater in the experimental group. The experimental group increased from a mean of 50.96 to 85.77, whereas the control group increased from 49.46 to 72.14. This difference in the magnitude of descriptive improvement was further examined through the N-Gain analysis.

3.5 Independent Sample t-Test

An independent sample t-test was conducted to compare the post-test reading comprehension scores of the experimental and control groups. This analysis was important because the primary purpose of the study was to determine whether the reading comprehension achievement of students receiving ChatGPT-assisted instruction differed statistically from that of students receiving conventional instruction.

Table 5. Results of the Independent Sample t-Test

Comparison	Sig. (2-tailed)	Decision at $\alpha = 0.05$	Interpretation
Experimental vs. Control Post-test	0.117	Not significant	No statistically significant difference

As shown in Table 5, the significance value of the independent sample t-test was 0.117. Because 0.117 is greater than the significance level of 0.05, the difference between the experimental and control groups' post-test scores was not statistically significant based on the reported result.

This finding requires careful interpretation. Although the experimental group obtained a considerably higher descriptive post-test mean than the control group, the reported independent sample t-test does not provide sufficient statistical evidence to conclude that the difference between the two groups was significant at the 0.05 level. Therefore, the result should not be reported as a statistically significant difference.

Nevertheless, the higher mean score of the experimental group and its larger improvement from pre-test to post-test indicate a favorable descriptive pattern for ChatGPT-assisted instruction. The experimental group reached a post-test mean of 85.77, compared with 72.14 for the control group. However, because the reported significance value was 0.117, the observed difference should be described as a descriptive advantage rather than a statistically significant treatment effect.

This result also demonstrates the importance of distinguishing between descriptive improvement and inferential statistical significance. A higher mean score does not automatically demonstrate that one instructional treatment is statistically more effective than another. In this study, both groups improved significantly within their respective groups, while the between-group comparison did not reach the conventional 0.05 significance threshold.

3.6 N-Gain Analysis

To provide additional information about the magnitude of students' learning improvement, an N-Gain analysis was conducted. N-Gain analysis measures the proportion of the possible improvement that was achieved by students from the pre-test to the post-test. The analysis was used to compare the magnitude of improvement between students who received ChatGPT-assisted instruction and those who received conventional instruction.

Table 6. N-Gain Scores of the Experimental and Control Groups

Group	N-Gain Score	Category
Experimental	72.71	High
Control	46.37	Medium

Table 6 shows that the experimental group obtained an N-Gain score of 72.71, which was categorized as high. In contrast, the control group obtained an N-Gain score of 46.37, which was categorized as medium. These findings indicate that the magnitude of learning improvement was greater in the experimental group than in the control group.

The high N-Gain score obtained by the experimental group suggests that students who participated in ChatGPT-assisted instruction experienced substantial improvement in their reading comprehension relative to their initial performance. The students were able to move from a relatively low initial level of reading comprehension to a substantially higher level following the instructional process.

The control group also demonstrated meaningful improvement, although its N-Gain category was lower. This finding indicates that conventional instruction remained capable of improving students' reading comprehension. Therefore, the role of ChatGPT should not be interpreted as replacing teacher instruction. Rather, the result suggests that ChatGPT-assisted instruction may provide additional support that can complement classroom teaching.

Discussion

The findings of this study provide a nuanced picture of the use of ChatGPT-assisted instruction in EFL reading comprehension. Overall, both the experimental and control groups demonstrated improvement from the pre-test

to the post-test. However, the experimental group showed a greater descriptive increase in reading comprehension and obtained a higher N-Gain score than the control group. At the same time, the independent sample t-test did not show a statistically significant difference between the two groups. Therefore, the findings should not be interpreted as definitive evidence that ChatGPT-assisted instruction was statistically superior to conventional instruction. Rather, they indicate that ChatGPT-assisted instruction was associated with a greater magnitude of descriptive improvement during the intervention.

The initial descriptive results indicate that the two groups started from relatively comparable levels of reading comprehension. The experimental group obtained a pre-test mean score of 50.96, while the control group obtained a mean score of 49.46. The relatively small difference between the initial mean scores suggests that the groups had similar starting points before the treatment. This is important because the subsequent improvement can be considered within the context of their initial reading performance rather than being attributed simply to a substantial difference in baseline ability. The similarity of the initial scores also provides a reasonable basis for comparing the learning development of the two groups after the instructional process.

Following the treatment, both groups demonstrated improvement. The experimental group increased from a mean score of 50.96 in the pre-test to 85.77 in the post-test, representing a descriptive increase of 34.81 points. The control group also improved, increasing from 49.46 to 72.14, or 22.68 points. Thus, the experimental group demonstrated a larger increase in its mean score than the control group. This pattern suggests that the instructional activities involving ChatGPT may have provided additional learning support for students as they processed English reading texts.

The improvement within the experimental group can be interpreted in relation to the functions of ChatGPT that were incorporated into the learning activities. During the treatment, students used ChatGPT to obtain explanations of unfamiliar vocabulary, clarify difficult sentences, identify main ideas, and check their understanding of comprehension questions. These activities potentially provided students with immediate assistance when they encountered difficulties during reading. Instead of stopping their reading process when they did not understand a word or sentence, students could obtain an explanation and continue working with the text. Such interaction may have reduced some of the linguistic barriers that commonly make English reading difficult for EFL learners.

The use of ChatGPT may also have supported students in moving through different stages of the reading process. Vocabulary assistance could help students understand individual words, while sentence clarification could support their understanding of relationships between ideas. Identifying the main idea required students to move beyond individual words and sentences toward understanding the overall message of the text. Similarly, checking answers to comprehension questions provided an opportunity for students to reconsider whether their interpretation of the text was consistent with the information presented. In this sense, ChatGPT functioned as a supplementary source of instructional support rather than simply as an answer-generating tool.

This interpretation is consistent with previous research on ChatGPT in EFL learning. Balci (2024) reported that ChatGPT can support language proficiency, learner engagement, individualized learning, and learner autonomy, while also identifying limitations that require careful implementation. Ardiyani et al. (2025) similarly reported that ChatGPT can support reading instruction by helping learners understand difficult texts and providing flexible feedback. More specifically, Kainta and Umarella (2026) synthesized recent empirical evidence on ChatGPT and EFL reading and reported that ChatGPT use has been associated with improvements in reading comprehension, main-idea comprehension, inferencing, contextual understanding, and related reading competencies. These findings provide a broader context for interpreting the higher descriptive improvement observed in the experimental group in the present study.

The present findings also have similarities with Çelik et al. (2024), who examined AI-supported simplification of authentic texts and reported benefits for reading-related outcomes. Although the instructional approach in the present study was not identical, both studies indicate that AI can be used to provide additional support when students encounter difficulties in processing English texts. The similarity is particularly relevant to the present study because the students were not simply exposed to AI technology; rather, ChatGPT was incorporated into reading activities in which students interacted with texts and received assistance with specific comprehension difficulties.

The significant improvement found within the experimental group is further supported by the paired sample t-test. The experimental group obtained a significance value of 0.000, which was below the 0.05 significance level. This result indicates that students' reading comprehension scores after the instructional process were significantly different from their scores before the treatment. Therefore, the experimental group experienced statistically significant improvement over time. This finding supports the observation from the descriptive statistics that the students developed their reading comprehension during the intervention.

However, the paired sample t-test for the control group also produced a significance value of 0.000. This means that the control group also experienced statistically significant improvement between the pre-test and post-test. This finding is important because it demonstrates that improvement cannot be attributed exclusively to ChatGPT. Conventional instruction also contributed to students' learning development. Students in the control group still

participated in structured reading activities, teacher explanations, and comprehension exercises, all of which could support the development of reading comprehension.

The improvement in the control group also helps explain why the difference between the two groups was not statistically significant. Although the experimental group obtained a higher post-test mean, the control group also experienced substantial improvement. When both groups receive instruction that supports reading comprehension, the statistical difference between their final scores may remain insufficient to reach the conventional significance level. Therefore, the significant within-group improvement in both groups should be distinguished from the between-group comparison.

The independent sample t-test provides the most important qualification to the interpretation of the treatment effect. The reported significance value was 0.117, which was greater than 0.05. Consequently, the difference between the experimental and control groups' post-test scores was not statistically significant. Although the experimental group achieved a higher post-test mean of 85.77 compared with 72.14 in the control group, the statistical test does not provide sufficient evidence to conclude that the observed difference was caused by the treatment at the 0.05 significance level.

This result indicates that descriptive superiority and statistical significance should not be treated as equivalent. The experimental group clearly demonstrated a higher mean score and a larger numerical improvement, but the independent sample t-test indicates that the difference between the two groups could not be confirmed statistically under the criteria used in this study. Therefore, the finding is better described as a descriptive advantage of the ChatGPT-assisted group rather than as statistically proven superiority over conventional instruction.

The N-Gain results provide another perspective on the magnitude of students' improvement. The experimental group obtained an N-Gain score of 72.71, categorized as high, whereas the control group obtained an N-Gain score of 46.37, categorized as medium. The difference between these values indicates that the experimental group achieved a greater proportion of possible learning improvement relative to its initial performance. Thus, although the between-group t-test did not reach statistical significance, the N-Gain analysis shows a meaningful descriptive difference in the extent of improvement achieved by the two groups.

The higher N-Gain in the experimental group may be related to the individualized assistance available during the reading activities. Students could request explanations according to the difficulties they encountered rather than receiving exactly the same explanation at the same time. For students who struggled with vocabulary, ChatGPT could provide contextual explanations. For students who had difficulty interpreting sentences, it could provide clarification. For students who needed help identifying the main idea, it could guide them toward the central information in the text. Such differentiated support may have contributed to the higher learning gain observed in the experimental group.

Nevertheless, the use of ChatGPT should not be understood as a replacement for teacher instruction. The findings of the present study indicate that conventional instruction also produced significant improvement, while the difference between the two groups was not statistically significant. In addition, previous research has emphasized that ChatGPT can present limitations related to inaccurate information, excessive student dependence, and reduced opportunities for independent thinking when it is not used appropriately (Ardiyani et al., 2025). Therefore, the pedagogical value of ChatGPT depends not only on the technology itself but also on how teachers structure and monitor its use.

Teacher guidance is particularly important in reading instruction because students need to develop the ability to interpret texts independently. If students use ChatGPT merely to obtain direct answers, the technology may reduce rather than strengthen their engagement with the reading process. In contrast, when ChatGPT is used to clarify vocabulary, explain difficult sentences, provide guiding questions, or support students in checking their understanding, it can function as an additional learning resource. This distinction is important for interpreting the present findings because the observed improvement should be understood within the context of teacher-guided ChatGPT-assisted instruction rather than unrestricted use of artificial intelligence.

The findings also have implications for the role of learner autonomy. ChatGPT can provide students with opportunities to seek clarification when they encounter difficulties, which may encourage them to become more active in identifying what they do not understand. Previous research has associated ChatGPT-supported learning with greater opportunities for individualized learning and learner autonomy (Balci, 2024; Kainta & Umarella, 2026). However, autonomy should not be interpreted as complete independence from teachers. In an EFL classroom, students still require appropriate learning objectives, task guidance, evaluation, and feedback from teachers.

The findings of this study therefore both support and qualify previous positive claims regarding ChatGPT-assisted language learning. On the one hand, the experimental group achieved a higher post-test mean and a high N-Gain score, which is consistent with previous research reporting potential benefits of AI-supported reading instruction. On the other hand, the non-significant independent sample t-test indicates that the present study does not provide sufficient statistical evidence to conclude that ChatGPT-assisted instruction was significantly more effective than

conventional instruction. This difference between the descriptive and inferential results is important and should be explicitly acknowledged rather than overlooked.

One possible explanation for the non-significant between-group result is the relatively limited scope of the intervention. The study was conducted with students from one junior high school and within a limited instructional period. The sample size was also relatively small, consisting of 54 students. Under these circumstances, differences in individual learning ability, classroom interaction, prior knowledge, and students' familiarity with digital tools may influence the final outcomes. A longer intervention and a larger sample may provide greater statistical power to determine whether the descriptive advantage observed in the experimental group can be replicated consistently. Another limitation concerns the focus of the assessment. The study measured reading comprehension through pre-test and post-test scores based on multiple-choice items. Such an assessment provides useful evidence of comprehension achievement but may not fully capture other dimensions of reading development, such as critical reading, inferencing, vocabulary development, reading strategies, or the ability to independently evaluate information. Future research could therefore combine standardized comprehension tests with qualitative evidence, student reflections, classroom observations, or interviews to obtain a more comprehensive understanding of how students use ChatGPT during reading.

The findings also suggest that future studies should investigate not only whether ChatGPT improves reading comprehension but also which specific ChatGPT-supported activities contribute most to students' improvement. For example, future research could separately examine vocabulary explanation, text summarization, main-idea identification, question generation, and feedback activities. Such research would provide more specific evidence about which functions of ChatGPT are most beneficial for EFL reading.

Overall, the present study demonstrates that ChatGPT-assisted instruction was associated with substantial improvement in the experimental group, as reflected in the increase in mean scores and the high N-Gain category. However, because the independent sample t-test did not show a statistically significant difference between the experimental and control groups, the findings should be interpreted cautiously. The contribution of ChatGPT in this study is therefore best understood as a promising supplementary instructional resource that may enhance students' learning opportunities when integrated with structured reading activities and teacher guidance. The results provide encouraging evidence while also demonstrating the need for further empirical investigation using larger samples, longer interventions, and more comprehensive measures of reading comprehension.

4. Conclusion

This study aimed to examine the effectiveness of ChatGPT-assisted instruction in improving the reading comprehension of eighth-grade students at SMPN 1 Paiton. The findings showed that both the experimental and control groups experienced improvement after the instructional process. The experimental group increased from a pre-test mean score of 50.96 to a post-test mean score of 85.77, while the control group increased from 49.46 to 72.14. The paired sample t-test results showed statistically significant improvement within both groups, with a significance value of 0.000. These results indicate that both ChatGPT-assisted instruction and conventional instruction were associated with improvement in students' reading comprehension.

The experimental group demonstrated a greater descriptive improvement than the control group. This was reflected in both the difference between the pre-test and post-test mean scores and the N-Gain results. The experimental group obtained an N-Gain score of 72.71, categorized as high, whereas the control group obtained an N-Gain score of 46.37, categorized as medium. These results indicate that students who participated in ChatGPT-assisted instruction achieved a greater magnitude of learning improvement during the intervention. However, the independent sample t-test produced a significance value of 0.117, which was higher than the 0.05 significance level. Therefore, the difference in post-test performance between the experimental and control groups was not statistically significant. The study consequently does not provide sufficient statistical evidence to conclude that ChatGPT-assisted instruction was significantly more effective than conventional instruction.

The findings suggest that ChatGPT can be considered a supplementary instructional resource for EFL reading instruction, particularly for providing students with additional support in understanding vocabulary, difficult sentences, main ideas, and comprehension questions. Its application is more appropriately positioned as a complement to teacher-guided instruction rather than as a replacement for the teacher. The use of ChatGPT should be accompanied by clear instructional objectives and appropriate guidance so that students use the technology to support comprehension rather than simply obtain answers.

This study was limited by its relatively small sample, its implementation in a single junior high school, and its limited instructional period. In addition, the assessment focused on reading comprehension measured through pre-test and post-test scores. Future research should involve larger and more diverse samples, longer intervention periods, and different educational contexts. Further studies may also examine specific ChatGPT-supported reading activities and use additional measures, such as reading strategies, inferencing, critical reading, learner autonomy, or qualitative classroom data, to provide a broader understanding of the role of ChatGPT in EFL reading instruction.

Reference

- Ali, M. M., Alaa, A. M., Alharbi, W., & Alqurashi, I. (2025). Investigating The Impact Of Chatgpt On Reading Proficiency In EFL Context. *International Journal Of Technology In Education (IJTE)*, 8(4), 1222–1244.
- Ardiyani, R. T., Kristen, U., & Wacana, S. (2025). *Teaching English And Language Learning English Journal (TELLE) Vol 05 ; No : 03 ; December 2025 301 Teaching English And Language Learning English Journal (TELLE). 05(03)*, 301–313.
- Bacı, Ö. (2024). The Role Of Chatgpt In English As A Foreign Language (EFL) Learning And Teaching: A Systematic Review. *International Journal Of Current Educational Studies*, 3(1), 66–82. <https://doi.org/10.46328/ijces.107>
- Çelik, F., Ersanlı, C. Y., & Arslanbay, G. (2024). Does AI Simplification Of Authentic Blog Texts Improve Reading Comprehension, Inferencing, And Anxiety? A One-Shot Intervention In Turkish EFL Context. *International Review Of Research In Open And Distributed Learning*, 25(3 Special Issue), 287–303. <https://doi.org/10.19173/irrod.V25i3.7779>
- Hidayah, L. W., Maulida, P. A., & Rahmawati, S. (2026). *Reading Strategies In EFL Classrooms : Overcoming Obstacles In Drawing*. 4778, 35–45.
- Iqbal Tawakkal, M., & Apriono, D. (2026). Enhancing University Students' Academic Reading Comprehension Through Chatgpt-Assisted Instruction In Indonesian Higher Education. *Journal Of English Language Teaching, Literatures & Applied Linguistics (JELTLAL)*, 4(1), 69–78.
- Ironsi, C. S., & Bostanci, H. B. (2026). Enhancing The Reading Comprehension And Autonomy Of Foreign Language Students Using Generative AI-Integrated Assistant. *Frontiers In Education*, 11(March). <https://doi.org/10.3389/Feduc.2026.1763238>
- Kainta, E., & Umarella, R. F. (2026). Empirical Evidence On The Use Of Chatgpt In Developing EFL Students' Reading Comprehension: A Systematic Review. *English Education:Journal Of English Teaching And Research*, 11(1). <https://doi.org/10.29407/Jetar.V11i1.28305>
- Mallipa, I., Sitompul, E., & Sitorus, P. A. (2024). The English Reading Comprehension, Skill, And Strategy Of Senior High School Students. *IDEAS Journal Of Language Teaching And Learning, Linguistics And Literatur*, 12(1), 714–728. <https://doi.org/10.24256/Ideas>
- Riyani, N. (2025). *The Effectiveness Of AI-Assisted Reading Tools*. 4778, 8970–8980. <https://doi.org/10.24256/Ideas>
- Siregar. (2025). Enhancing EFL Reading Comprehension Via An AI-Chatbot-Guided Toulmin Mapping In Viat-Map. *Utamax : Journal Of Ultimate Research And Trends In Education*, 7(2), 144–158. <https://doi.org/10.31849/9nwj8r52>
- Sugiyono. (2020). *Metode Penelitian Kuantitatif, Kualitatif Dan Kombinasi (Mix Methods)*.
- Syafruddin, S., Hidayati, S., & Triastuti, A. (2025). A Systematic Literature Review: The Trends And Challenges Of AI (Chat GPT) For Reading Skills In ELT/L Using The HHH Framework. *Indonesian Journal Of Integrated English Language Teaching*, 11(1), 32. <https://doi.org/10.24014/Ijiet.V11i1.36613>
- Unus, A. Y. Y., & Un, M. M. L. (2025). *ENCHANCING ENGLISH READING SKILLS THROUGH AI- BASED TOOLS : A SYSTEMATIC REVIEW*. 6(6), 373–390.
- Usman, K. Et Al. (2017). *Efl Students' Reading Comprehension Problems: Linguistic And Non Linguistic Complexities*. 8 (3), 308–321.
- Wang, J., & Fan, W. (2025). The Effect Of Chatgpt On Students' Learning Performance, Learning Perception, And Higher-Order Thinking: Insights From A Meta-Analysis. *Humanities And Social Sciences Communications*, 12(1), 1–21. <https://doi.org/10.1057/S41599-025-04787-Y>