Journal of English Language and Education
ISSN 2597- 6850 (Online), 2502-4132 (Print)
Journal Homepage: https://jele.or.id/index.php/jele/index

OPEN ACCESS

Article

ChatGPT-Assisted Vocabulary Learning in a Coastal Junior High School

https://doi.org/10.31004/jele.v11i6.3276

Muh. Alfachrezy, Dinda Ramadani, Nonawati, Aura Ariani Azila, Rais Abin, Bimas Reskiawan[abcdef]

[123456]English Education Study Program, Faculty of Teacher Training and Education, Universitas Sembilanbelas November Kolaka, Indonesia
Corresponding Author: dindaramadanii088@gmail.com

A B S T R A C T

Vocabulary knowledge is essential for English learning, but students in resource-limited educational settings may have fewer opportunities to access interactive learning support. This study examined changes in vocabulary achievement following guided ChatGPT-assisted learning among 19 Grade IX students at SMPN 5 Tanggetada Terapung, Kolaka, Indonesia. A quantitative pre-experimental one-group pre-test and post-test design was employed, with participants selected through convenience sampling. The intervention consisted of six 90-minute meetings, during which students used smartphones collaboratively to learn vocabulary related to descriptive texts under instructor guidance. The mean score increased from 72.63 (SD = 17.27) to 90.79 (SD = 6.92), representing an average gain of 18.16 points. The Shapiro-Wilk test did not reject the normality assumption for the paired difference scores (W = 0.931, p = .180). A paired-samples t-test revealed a statistically significant difference, t(18) = 3.729, p = .002, with a large within-group effect size (Cohen's dz = 0.855). These findings indicate improved vocabulary achievement following the intervention. However, the small sample and absence of a control group prevent attributing the improvement solely to ChatGPT. Guided ChatGPT use may serve as supplementary vocabulary support in this classroom context.

Keywords: *Chatgpt, Vocabulary Learning, Descriptive Text, Junior High School, Coastal Education*

Article History:
Received 07th September 2026
Accepted 05th November 2026
Published 11th November 2026

INTRODUCTION

Vocabulary knowledge is a central component of learning English as a foreign language because learners depend on words to understand input and to express meaning. Reading comprehension, writing, listening, and speaking all require access to a sufficiently broad and usable vocabulary. At the junior high school level, vocabulary learning is particularly important because students are moving from basic word recognition toward the use of English in short texts and communicative tasks. When learners know only a limited number of words, they may understand classroom explanations but still struggle to interpret texts, construct sentences, or participate confidently in interaction. Limited vocabulary knowledge can also restrict the development of broader language proficiency, as students need to recognize and use words appropriately across different communicative situations. For that reason, vocabulary instruction needs to move beyond isolated memorization and provide repeated, contextualized opportunities to notice, understand, and use new words (Nation, 2001; Webb & Nation, 2017).

The challenge becomes more pronounced when learning resources are limited. In the context of the present study, the partner school is located in a coastal area of Tanggetada, Kolaka, Southeast Sulawesi. The final SIM-B report describes SMPN 5 Tanggetada Terapung as the site of the activity and records that the program was implemented from 14 July to 12 August 2026. Initial observation and coordination with the school were used to identify the condition of English learning, students' needs, and practical constraints before the

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

instructional activities were implemented. One practical consideration in organizing the intervention was the use of smartphones collaboratively, allowing students to access ChatGPT through group activities rather than relying exclusively on individual interaction with the application. This approach provided a way to organize technology-supported vocabulary practice within the classroom setting. The instructional concern was therefore how to provide accessible and contextualized vocabulary learning opportunities while maintaining teacher guidance and student interaction. This context is important because the usefulness of a digital learning tool cannot be evaluated independently from the learning environment in which it is introduced.

Digital technology offers one possible way to extend students' access to learning support. Mobile devices are already familiar to many learners and can provide access to dictionaries, multimedia, online exercises, and conversational applications. However, access to a tool does not automatically produce learning. The quality of the task, the guidance provided by the teacher, the appropriateness of the content, and students' ability to use the tool purposefully remain decisive. In this respect, generative artificial intelligence has attracted attention because it can respond to natural-language prompts, generate examples, explain word meanings, create short exercises, and provide immediate responses that can be adapted to the learner's request (Kohnke et al., 2023).

ChatGPT is one of the most widely discussed generative AI tools in language education. For vocabulary learning, its potential lies in the possibility of presenting target words in multiple contexts rather than only as word-translation pairs. A learner can request a simple definition, an example sentence, a contrast with a related word, a short descriptive paragraph, or a practice question. This kind of interaction may support repeated exposure to vocabulary and may encourage learners to ask for clarification when they do not understand an item. Such features are particularly relevant to vocabulary instruction because students need opportunities to understand word meanings and recognize how words function in meaningful contexts. At the same time, the technology has limitations. AI-generated explanations can be inaccurate, too difficult, or overly general, and students may accept responses passively if teachers do not require verification and reflection. Therefore, a pedagogically meaningful implementation should position ChatGPT as supplementary support rather than as a replacement for instruction. In the present study, ChatGPT was integrated into instructor-guided activities to provide additional vocabulary explanations, contextual examples, and opportunities for collaborative practice (Kasneci et al., 2023).

Previous research provides a basis for investigating this possibility. Abdelhalim & Alsehibany (2025) examined the integration of ChatGPT for vocabulary learning and retention among Saudi EFL learners. Their classroom-based study used a quasi-experimental design and reported benefits for vocabulary acquisition and retention when ChatGPT was integrated into guided learning activities. Their findings are relevant because they suggest that generative AI can be used not only for open-ended conversation but also for structured vocabulary practice. The study further illustrates that the contribution of ChatGPT to vocabulary learning depends on how it is embedded within classroom tasks, feedback, and learner interaction.

Sapan & Uzun (2024) also investigated ChatGPT-integrated English teaching, focusing on high school EFL learners' writing skills and vocabulary development. Their work is particularly relevant to the present study because the participants were school-level learners rather than university students, and the intervention connected ChatGPT use with regular English instruction. The study demonstrates the possibility of integrating AI into language lessons instead of treating it as an independent self-study application. Nevertheless, the educational context, age group, instructional objectives, and technological conditions differ from those of a coastal junior high school in Indonesia. This creates a need for context-sensitive evidence rather than assuming that findings from one educational setting will automatically transfer to another.

Evidence is also not uniformly positive. Zahran (2025), for example, investigated a Poe ChatGPT-based TPACK model involving EFL teachers and their students. The study found improvement in teachers' performance, but the vocabulary results did not show a significant

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

difference between experimental and control groups. This result is important because it prevents an overly deterministic interpretation of AI in language learning. The presence of ChatGPT does not guarantee vocabulary gains. Learning outcomes may depend on task design, duration, teacher mediation, learner characteristics, and the relationship between the test and the instructional content. A balanced interpretation of ChatGPT-assisted learning therefore requires attention to both promising and non-significant findings.

The present study is also connected to the longer tradition of computer-assisted vocabulary learning. Ma & Kelly (2006) argued that computer-assisted vocabulary learning should be designed around principles of vocabulary acquisition rather than around technology for its own sake. Their work on the design and evaluation of computer-assisted vocabulary learning is relevant because it emphasizes purposeful interaction, learner engagement, and the design of tasks that support vocabulary processing. Although generative AI is technologically different from earlier computer-assisted vocabulary systems, the underlying pedagogical issue remains similar: the tool must create conditions in which students actively process and use vocabulary. This emphasis on active processing is also consistent with the Involvement Load Hypothesis proposed by Hulstijn & Laufer (2001).

The instructional focus in the present study was descriptive text. This choice provides a practical context for vocabulary development because students need adjectives, nouns, and descriptive expressions to describe people, places, and animals. The final report specifies that the material focused on describing people, places, and animals, while the learning process guided students to use ChatGPT to understand word meanings, obtain examples of word use, and practice vocabulary in sentence contexts. By linking vocabulary to descriptive text, the learning activities aimed to avoid treating words as isolated items and instead connect them with communicative meaning.

A further feature of the intervention was collaborative smartphone use. Students accessed ChatGPT through smartphones and worked in groups. Collaborative access is pedagogically significant in a context where individual device availability may not be equal. Group work also changes the nature of the interaction: students must discuss what to ask, interpret the response, compare ideas, and decide how to use the information. In this arrangement, ChatGPT becomes one component in a broader learning interaction involving peers and instructors. This differs from a purely individual model in which a learner interacts privately with an AI system without classroom mediation. The combination of collaborative access, contextualized vocabulary activities, and instructor guidance therefore represents an important pedagogical characteristic of the intervention.

The research gap addressed by this study is primarily contextual and pedagogical. Previous studies have examined ChatGPT-assisted vocabulary learning and other aspects of EFL education in different educational settings. However, the studies reviewed above provide limited evidence concerning the implementation of guided, collaborative ChatGPT-assisted vocabulary learning among junior high school students in coastal Indonesian contexts. In particular, the application of ChatGPT through shared smartphone activities, combined with instructor guidance and descriptive-text vocabulary instruction, warrants further investigation in this educational setting. The SIM-B report also identifies the need to provide more interactive and accessible technology-supported learning in the partner school. Accordingly, the research problem addressed in the present study concerns whether students demonstrate measurable changes in vocabulary achievement after participating in structured ChatGPT-assisted learning activities. Rather than assuming that ChatGPT independently produces vocabulary gains, the study examines students' achievement following an intervention that integrates ChatGPT with instructor explanation and collaborative classroom practice.

Accordingly, this study aimed to examine whether Grade IX students at SMPN 5 Tanggetada Terapung demonstrated a statistically significant change in vocabulary achievement following ChatGPT-assisted vocabulary learning. The change was measured by comparing students' pre-test and post-test vocabulary scores. The study addressed the following research question: Is there a statistically significant difference between students'

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

vocabulary achievement before and after ChatGPT-assisted vocabulary learning? The hypothesis test was complemented by an effect-size estimate so that the interpretation would consider not only statistical significance but also the magnitude of the observed change. Given the one-group pre-test and post-test design, the findings were intended to describe changes in students' achievement within this particular educational context rather than establish the independent causal effect of ChatGPT.

METHOD

Research Design

This study used a quantitative pre-experimental method with a one-group pre-test and post-test design (Campbell & Stanley, 1963; Knapp, 2016). This design was used to examine changes in students' vocabulary achievement before and after learning vocabulary with the assistance of ChatGPT. The design was selected because the research was conducted within an existing classroom setting as part of an educational program, allowing the researchers to examine vocabulary achievement before and after the instructional activities. The absence of a control group was related to the implementation of the program within a single participating class and the practical constraints of the school setting. The research activities consisted of initial observation, pre-test, ChatGPT-assisted vocabulary learning, post-test, and analysis of the students' scores.

The research design can be illustrated as O1 – X – O2. O1 refers to the vocabulary pre-test, X refers to the ChatGPT-assisted vocabulary learning activities, and O2 refers to the vocabulary post-test (Campbell & Stanley, 1963). Since this study involved only one group without a control group, the results were used to examine changes in the participating students rather than to make a direct comparison with another group. Consequently, the design could identify differences in vocabulary achievement before and after the intervention but could not establish that ChatGPT independently caused the observed changes (Marsden & Torgerson, 2012; Shadish et al., 2002). The intervention consisted of six meetings, with approximately 90 minutes for each meeting.

Respondents

The participants of this study were 19 Grade IX students of SMPN 5 Tanggetada Terapung, located on Jalan Perkampungan Bajo, Kelurahan Anaiwoi, Kecamatan Tanggetada, Kabupaten Kolaka, Southeast Sulawesi, Indonesia. The participants were selected using convenience sampling, as the research involved an existing Grade IX class that was accessible during the implementation of the educational program. All students in the selected class participated in the intervention, and all 19 students had complete pre-test and post-test scores and were included in the analysis. The study focused specifically on the learning outcomes of this class within the context of the school.

The participants were junior high school students aged approximately 14–16 years who were receiving English instruction at the ninth-grade level. Their selection was based on classroom accessibility and participation in the scheduled educational program. Since the research involved a single class from one school, the sample was not intended to represent the wider population of junior high school students. Therefore, the findings were interpreted within the educational context of the participating class.

Before the learning activities were conducted, the team carried out an initial observation and communicated with the English teacher. This was done to identify the students' needs and difficulties in learning English as well as the resources available at the school. The team also coordinated with the school and the English teacher to adjust the activities to the students' needs and classroom conditions. Since the activities involved the use of smartphones and internet access to ChatGPT, these conditions were also considered when planning the implementation.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

Instruments

The main instrument used in this study was a 20-item multiple-choice vocabulary test, which was administered as both the pre-test and post-test. The test focused on vocabulary related to descriptive texts, particularly words used to describe people, places, and animals. The items were designed to assess students' understanding of word meanings and their ability to recognize appropriate vocabulary usage in descriptive contexts. The pre-test was given before the treatment to measure students' initial vocabulary achievement, while the post-test was given after the treatment to measure their achievement following the learning activities (Downing, 2006).

The test employed a dichotomous scoring system, in which each correct answer received one point and each incorrect answer received zero points. The total number of correct answers was converted into a score ranging from 0 to 100 by dividing the raw score by 20 and multiplying the result by 100. Therefore, each correct answer contributed five points to the final score, with a maximum possible score of 100. Higher scores indicated better vocabulary achievement.

The test items were prepared by the research team and consulted with the field supervisor before being used. Content validity was examined through expert judgment involving the field supervisor and an English teacher, who evaluated the relevance of the vocabulary items to the instructional objectives, the clarity of the questions, and their appropriateness for Grade IX students. The evaluations were conducted using a four-point relevance scale and analyzed using Aiken's V coefficient (Aiken, 1985; Kania et al., 2024). The analysis produced an average coefficient of 0.85, with individual item coefficients ranging from 0.67 to 1.00, indicating generally strong content relevance. Minor revisions were made to improve the clarity and suitability of several items based on the evaluators' feedback.

The internal consistency of the vocabulary test was examined using the Kuder–Richardson Formula 20 (KR-20), which was appropriate for the dichotomously scored multiple-choice items (Kuder & Richardson, 1937). The reliability analysis was conducted using the item-level responses obtained from the 19 students during the pre-test. The analysis produced a KR-20 coefficient of 0.79, indicating acceptable internal consistency for the vocabulary assessment. Together, the expert judgment results and reliability analysis provided supporting evidence regarding the suitability and consistency of the instrument for measuring students' vocabulary achievement.

Observation and documentation were also used to support the data collection. Observation was conducted during the learning activities to see students' participation and the implementation of the activities. Documentation was used to record the implementation of the program and classroom activities. These data were used as supporting information and were not analyzed as separate quantitative data.

Procedures

The research began with observation and coordination with the partner school. The team met with school representatives and the English teacher to explain the SIM-B program, identify students' learning needs, and discuss the technical preparation for the activities. The initial observation and coordination were conducted on 14 July 2026, while the program was completed on 12 August 2026. During this period, six instructional meetings of approximately 90 minutes each were conducted.

Before the treatment began, the students completed the vocabulary pre-test to measure their initial vocabulary achievement. After the pre-test, the learning activities focused on vocabulary used in descriptive texts. The materials covered three main topics: vocabulary for describing people, places, and animals.

During the learning activities, the instructors introduced and explained the target vocabulary. ChatGPT was used as a supplementary learning tool to provide word meanings, examples of vocabulary use, and sentences containing the target words. Students accessed ChatGPT through smartphones and worked collaboratively in groups. They discussed the prompts and responses with their group members, while the instructors guided the activities

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

and explained unfamiliar words or responses when necessary. This allowed the learning activities to combine instructor explanation, group discussion, and the use of ChatGPT. In this process, ChatGPT functioned as supplementary learning support rather than as a replacement for instructor-led vocabulary instruction.

Students' participation and difficulties were also observed throughout the learning process. The results of this monitoring were used to make adjustments when necessary during the following activities. After completing the six instructional meetings, the students took the post-test using the same vocabulary instrument administered during the pre-test. Their post-test scores were then matched with their pre-test scores for statistical analysis.

Data Analysis

The students' pre-test and post-test scores were analyzed using descriptive and inferential statistics. The descriptive statistics included the number of participants, mean, median, standard deviation, minimum score, and maximum score. Gain scores were also calculated by subtracting each student's pre-test score from the post-test score.

Before conducting the paired-samples t-test, the normality of the difference scores was examined using the Shapiro-Wilk test. The difference scores were used for the normality test because the analysis compared two measurements from the same participants. The Shapiro-Wilk test was selected considering the small number of participants in this study. A significance value greater than .05 was interpreted as insufficient evidence to reject the assumption of normality rather than as definitive proof that the difference scores were normally distributed. Since the normality assumption was not rejected, a paired-samples t-test was used to determine whether there was a significant difference between the students' pre-test and post-test scores.

The level of statistical significance was set at .05. Cohen's dz was also calculated by dividing the mean difference score by the standard deviation of the difference scores. The effect size was interpreted using the commonly used criteria of 0.20 as a small effect, 0.50 as a medium effect, and 0.80 as a large effect. This calculation was used to show the magnitude of the difference between the students' scores before and after the intervention. Given the absence of a control group, the statistical results were interpreted as evidence of changes in vocabulary achievement within the participating group rather than definitive evidence that ChatGPT independently caused the improvement.

FINDINGS AND DISCUSSION

Descriptive Findings

The descriptive results showed an increase in students' vocabulary achievement after the ChatGPT-assisted learning activities. A total of 19 paired pre-test and post-test scores were analyzed. The mean score increased from 72.63 (SD = 17.27) on the pre-test to 90.79 (SD = 6.92) on the post-test, with an average increase of 18.16 points. The median score also increased from 75.00 to 95.00. Meanwhile, the minimum score increased from 20 to 75, while the maximum score remained 100 in both tests.

The lower standard deviation in the post-test indicates that students' scores became less varied after the intervention. This can also be seen from the increase in the minimum score. Although students started with different levels of vocabulary achievement, their post-test scores were generally closer to the higher end of the score range. These findings indicate an overall improvement in vocabulary test performance, although the extent of improvement differed among individual students.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

Table 1. Descriptive Statistics of Students' Vocabulary Scores

Statistic	Pre-test	Post-test
N	19	19
Mean	72.63	90.79
Median	75.00	95.00
Standard deviation	17.27	6.92
Minimum	20	75
Maximum	100	100
Mean difference	–	+18.16

The individual scores show that most students obtained higher scores on the post-test, although the amount of improvement varied. Of the 19 students, 14 students (73.7%) demonstrated positive gains, two students (10.5%) obtained the same score on both tests, and three students (15.8%) received lower scores on the post-test. Some students showed considerable gains, particularly those who had relatively low pre-test scores. The largest individual improvement was 75 points, while the greatest decrease was 10 points. Therefore, the improvement found in this study refers to the overall performance of the group rather than to every individual student.

Table 2. Individual Pre-test, Post-test, and Gain Scores

Student	Pre-test	Post-test	Gain
1	70	95	+25
2	55	95	+40
3	75	75	0
4	80	90	+10
5	90	80	-10
6	100	95	-5
7	60	100	+40
8	80	95	+15
9	65	95	+30
10	75	90	+15
11	20	95	+75
12	70	100	+30
13	55	95	+40
14	85	85	0
15	75	95	+20
16	80	90	+10
17	80	85	+5
18	75	90	+15
19	90	80	-10

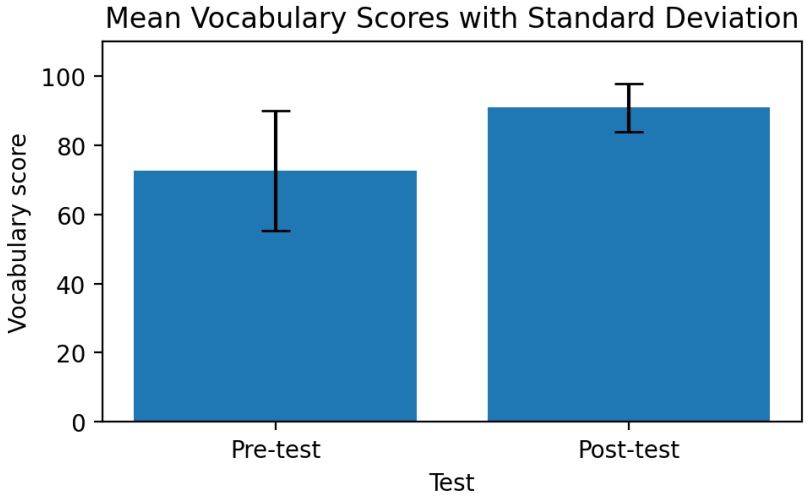


Figure 1. Mean Pre-test and Post-test Vocabulary Scores with Standard Deviation

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

Normality and Hypothesis Testing

Since the study involved paired measurements, the normality test was conducted on the difference between the pre-test and post-test scores. The Shapiro-Wilk test produced W = 0.931 and p = .180. Since the p-value was greater than the significance level of .05, there was insufficient evidence to reject the assumption that the difference scores were normally distributed. Therefore, the normality assumption was not rejected, supporting the use of the paired-samples t-test for further analysis.

The paired-samples t-test showed a statistically significant difference between the pre-test and post-test scores, t(18) = 3.729, p = .002. The mean difference between the two tests was 18.16 points. Since the p-value was lower than .05, the null hypothesis of no mean difference was rejected. This indicates that students obtained significantly higher vocabulary scores after participating in the ChatGPT-assisted learning activities. However, the statistical significance of the difference does not establish that ChatGPT independently caused the improvement, as the study involved a single group without a comparison condition.

The effect size was also calculated using Cohen's dz and resulted in a value of 0.855. Based on the conventional criteria, this value represents a large effect. Therefore, the difference between the students' pre-test and post-test scores was not only statistically significant but also showed a relatively large standardized within-group change. Nevertheless, the magnitude of this effect should be interpreted in relation to the study design, as it represents the observed change in the participating group rather than an isolated treatment effect attributable exclusively to ChatGPT.

Table 3. Inferential Analysis of Pre-test and Post-test Differences

Analysis	Result	Interpretation
Shapiro-Wilk difference scores	W = 0.931; p = .180	Normality assumption met
Paired-samples t-test	t(18) = 3.729; p = .002	Significant difference
Mean difference	+18.16	Post-test higher
Cohen's dz	0.855	Large within-group effect size

Discussion

The findings showed that students achieved higher vocabulary scores after participating in the ChatGPT-assisted learning activities. The mean score increased by 18.16 points, from 72.63 on the pre-test to 90.79 on the post-test. The statistical analysis also showed a significant difference between the two scores, with a large within-group effect size. These findings suggest that the learning activities involving ChatGPT were associated with improvement in students' vocabulary achievement. However, the improvement should be understood within the context of an instructional intervention that combined ChatGPT use, instructor explanations, collaborative activities, and repeated vocabulary practice.

This finding is in line with Abdelhalim & Alsehibany (2025), who reported positive results from the use of ChatGPT in vocabulary learning and retention. One possible reason for the improvement in the present study is that ChatGPT gave students additional opportunities to encounter vocabulary in context. During the learning activities, students used ChatGPT to obtain word meanings, examples, and sentences related to descriptive texts. Instead of learning vocabulary only as isolated words, students could see how the words were used in sentences and descriptions. Repeated exposure to vocabulary in meaningful contexts may have supported students' understanding of word meanings and appropriate usage. Nevertheless, the present study measured vocabulary achievement immediately after the intervention and did not examine long-term retention, making its findings more directly relevant to short-term learning outcomes.

The findings are also related to the ideas presented by Ma & Kelly (2006) regarding computer-assisted vocabulary learning. Digital tools can support vocabulary learning when students actively process the language rather than simply receive information. In this study, ChatGPT was used together with group discussion and instructor guidance. Students discussed prompts and responses with their peers, while instructors helped explain unfamiliar

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

vocabulary and kept the activities connected to the lesson objectives. These activities provided opportunities for students to engage with vocabulary through explanation, contextual examples, and collaborative interpretation. Therefore, ChatGPT functioned as one part of the learning process rather than as a replacement for instruction. This interpretation supports the importance of pedagogical integration, suggesting that the potential contribution of generative AI depends not only on its technological features but also on how teachers incorporate it into meaningful learning activities.

The collaborative use of smartphones was also an important part of the intervention. Students did not interact with ChatGPT individually throughout the activities. Instead, they worked in groups and discussed what to ask and how to understand the responses. This arrangement provided opportunities for interaction among students and made the activity more practical because students could share devices. In this context, ChatGPT served as a source of additional vocabulary information within a learning environment that also involved peer discussion and instructor assistance. Such collaboration may have supported the interpretation and application of vocabulary by allowing students to exchange ideas and clarify unfamiliar expressions. However, because collaborative learning was implemented alongside ChatGPT use, its independent contribution to the observed improvement could not be determined.

The focus on descriptive text may also have supported the learning process. The students learned vocabulary for describing people, places, and animals, which gave them a clear context for using the target words. Vocabulary such as nouns and adjectives could be directly connected to familiar descriptions. Keeping the activities focused on a specific topic also helped make the responses generated by ChatGPT more relevant to the lesson. This connection between the instructional materials and the vocabulary assessment may have supported students' performance by providing repeated opportunities to encounter and recognize words within a consistent learning context. Consequently, the observed improvement may reflect the combined influence of contextualized vocabulary instruction and technology-assisted practice.

However, the results do not mean that the use of ChatGPT will always produce the same effect. Zahran (2025), for example, reported no significant difference in vocabulary achievement between experimental and control groups in a study involving a Poe ChatGPT-based TPACK intervention. The different findings may be related to differences in participants, learning activities, duration, assessment, and the way AI was integrated into instruction. Furthermore, the present study examined changes within a single group, whereas Zahran's study involved a comparison between experimental and control groups. These methodological differences should be considered when interpreting the contrasting findings. While the present study demonstrated a significant improvement between two measurements, it could not determine whether the improvement exceeded what students might have achieved through conventional instruction without ChatGPT. This suggests that the effectiveness of ChatGPT-assisted vocabulary learning should be evaluated in relation to the instructional context and research design rather than assumed from the presence of the technology alone.

A similar point can be found in Sapan & Uzun (2024), who investigated the use of ChatGPT among high school EFL learners. Although their study involved a different educational level and also examined writing, both studies placed ChatGPT within planned classroom activities. In the present study, instructors provided learning objectives, guided students' use of the tool, and discussed its responses with them. Such guidance is important because students may otherwise use ChatGPT only for direct answers or translations without actively processing the vocabulary. The similarity between these instructional approaches highlights the relevance of structured teacher-guided AI integration in school-level EFL learning. Nevertheless, differences in participants, language skills, and assessment procedures limit direct comparisons between the studies.

Several alternative explanations should also be considered when interpreting the improvement in vocabulary achievement. First, instructor guidance may have contributed to

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

students' understanding of the target vocabulary, as the learning activities involved explanations, clarification, and assistance from the teaching team. Second, repeated vocabulary practice across six instructional meetings may have increased students' familiarity with the words and descriptive expressions included in the lessons. Third, the use of the same vocabulary test for both the pre-test and post-test introduced the possibility of a testing effect, in which students' prior exposure to the questions may have influenced their subsequent performance. Increased familiarity with the instructional materials may also have contributed to the higher post-test scores. Since these factors occurred alongside ChatGPT-assisted learning, their individual contributions could not be separated. Accordingly, the findings should be interpreted as evidence of improved vocabulary performance following the overall instructional intervention rather than as definitive proof of the independent effectiveness of ChatGPT.

Another interesting result was the reduction in score variation. The standard deviation decreased from 17.27 on the pre-test to 6.92 on the post-test, while the minimum score increased from 20 to 75. This indicates that the students' post-test scores were more concentrated within the higher score range. The pattern may suggest that some students with lower initial achievement experienced considerable improvement during the intervention. For instance, students who began with relatively low pre-test scores demonstrated some of the largest positive gains. However, because the study involved only 19 students, this pattern should be interpreted carefully. The reduction in score variation also does not necessarily indicate that differences in students' overall vocabulary proficiency were eliminated, as the test measured a specific set of vocabulary items related to descriptive texts.

The individual gain scores also show that students did not respond to the intervention in exactly the same way. Most students improved, but two students showed no change and three received lower post-test scores. Students who already had high pre-test scores also had less room for improvement because the maximum test score was 100. This suggests the possibility of a ceiling effect, particularly among students whose initial scores were already close to the maximum. The relatively high post-test mean of 90.79 further indicates that many students performed near the upper limit of the assessment. As a result, the instrument may have been less sensitive to additional differences in achievement among higher-scoring students. Therefore, the increase in the group mean should not be interpreted as equal improvement for all students, and the possibility of ceiling effects should be considered when evaluating the magnitude and distribution of the observed gains.

The quality of the vocabulary assessment also deserves consideration. The instrument underwent content review through expert judgment, producing an average Aiken's V coefficient of 0.85, while the internal consistency analysis using KR-20 yielded a coefficient of 0.79. These results provided preliminary support for the relevance of the test items to the instructional objectives and the consistency of the assessment. Nevertheless, the reliability estimate was obtained from a relatively small group of 19 students, which limits the precision and stability of the coefficient. Furthermore, evidence of content relevance and internal consistency does not eliminate other measurement concerns, including the use of identical test items across assessments and the possibility of ceiling effects. The instrument was therefore considered suitable for examining vocabulary achievement within the present classroom context, while broader claims concerning its measurement properties require further evidence.

Several limitations should also be considered. First, the study did not include a control group. As a result, the improvement cannot be attributed entirely to ChatGPT because other parts of the learning process, such as instructor explanation, group discussion, repeated vocabulary practice, and familiarity with the test, may also have contributed to the higher post-test scores. The absence of a comparison group limited the ability to distinguish the contribution of ChatGPT from other instructional influences. Future research could employ a quasi-experimental or experimental design with an appropriate comparison group to provide stronger evidence regarding the specific contribution of ChatGPT-assisted learning.

Second, the study involved a relatively small number of participants from a single junior high school and a short intervention consisting of six meetings. The results therefore mainly

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

reflect short-term vocabulary achievement within one class. The use of the same instrument for both assessments also introduced potential testing effects, while the high post-test scores suggested possible limitations in measuring further improvement among students with stronger initial vocabulary achievement. Moreover, no delayed post-test was conducted to examine whether students retained the vocabulary over a longer period. Future studies could involve larger samples, more varied assessment items, and delayed measurements to investigate the consistency and sustainability of vocabulary development.

Despite these limitations, the findings have several practical implications for English teaching. The study illustrates how smartphones can be used collaboratively to facilitate access to digital vocabulary learning activities in a classroom setting. ChatGPT can provide additional vocabulary examples, meanings, and practice when its use is connected to a clear learning objective. However, teacher guidance remains important in helping students understand, evaluate, and use the generated information appropriately. The findings therefore support considering ChatGPT as a supplementary resource within structured classroom instruction rather than as a replacement for teachers or established vocabulary teaching practices.

Teachers may also use ChatGPT for activities that require students to actively use vocabulary. For example, students can ask for simple definitions, compare similar words, create their own sentences, identify vocabulary in short descriptive texts, or answer short quizzes generated by ChatGPT. In this way, students are not only receiving information from AI but also using the information as part of a learning task. Nevertheless, the selection of activities should consider students' language proficiency, the accuracy of AI-generated responses, and the availability of appropriate guidance.

In the context of a coastal school where access to learning resources may vary, the use of shared smartphones offers one possible approach to integrating interactive digital materials without requiring individual devices for every student. However, technology should be viewed as a supporting tool rather than a solution to broader limitations in educational resources. Its usefulness still depends on the way teachers organize activities, monitor students' understanding, and guide their interaction with the learning materials. The present findings are therefore most applicable to the participating classroom and should not be generalized to all coastal schools or other educational settings without further investigation.

Overall, the findings suggest that guided ChatGPT-assisted learning can be incorporated as a supplementary approach to vocabulary instruction. In this study, its use was associated with higher vocabulary achievement after six instructional meetings. However, the observed improvement occurred within a broader learning process involving instructor guidance, collaborative activities, and repeated vocabulary practice. Therefore, ChatGPT should be integrated with clear learning objectives, appropriate assessment, and active teacher involvement rather than used as an independent source of learning.

CONCLUSIONS

This study investigated the use of ChatGPT to support vocabulary learning among 19 Grade IX students at SMPN 5 Tanggetada Terapung, Kolaka, Southeast Sulawesi. The intervention consisted of six 90-minute meetings focusing on vocabulary for describing people, places, and animals. Students used smartphones collaboratively under instructor guidance. The results showed an increase in mean vocabulary scores from 72.63 (pre-test) to 90.79 (post-test), representing an average gain of 18.16 points. A paired-samples t-test indicated a statistically significant difference ($p = .002$), with a large within-group effect size. These findings suggest improved vocabulary achievement following the intervention. However, the one-group pre-test and post-test design without a control group prevents attributing the improvement solely to ChatGPT. Instructor guidance, collaboration, repeated practice, and test familiarity may also have influenced the results. The small sample from one school further limits generalizability. Therefore, ChatGPT should be considered a supplementary learning tool rather than an independent instructional approach. Future

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

research should involve larger samples, comparison groups, and delayed post-tests to examine long-term vocabulary retention.

ACKNOWLEDGEMENTS

The authors would like to thank SMPN 5 Tanggetada Terapung, including the school leadership, English teacher, and participating students, for their support during the implementation of the SIM-B program. The authors also express their gratitude to the English Education Study Program, Faculty of Teacher Training and Education, Universitas Sembilanbelas November Kolaka, for its institutional support and assistance throughout the preparation and implementation of the program.

DECLARATION

Author Contributions: Muh. Alfachrezy contributed to the research conceptualization, development of the study design, implementation of the learning intervention, and preparation of the manuscript. Dinda Ramadani was involved in coordinating research activities, organizing data collection, conducting data analysis, and drafting the manuscript. Nonawati contributed to the development of instructional procedures, classroom implementation, and interpretation of the research findings. Aura Ariani Azila assisted with participant coordination, data documentation, and research administration. Rais Abin contributed to data processing, statistical examination, and interpretation of the quantitative results. Bimas Reskiawan participated in the review of the research methodology, manuscript editing, and refinement of the final paper. All authors examined and approved the final manuscript.

Conflict of Interest: The authors report no conflicts of interest or personal affiliations that may have influenced the conduct of the research, the interpretation of its findings, or the preparation of this publication.

Funding: No external financial support was received for this research. All research activities were conducted independently without grants or sponsorship from funding institutions or other organizations.

Ethical Approval: The research was conducted with consideration for ethical standards applicable to educational research involving school students. Permission to conduct the study was obtained from the relevant school authorities. Appropriate measures were taken to protect students' identities and maintain the confidentiality of the collected data.

Informed Consent: Permission and informed consent were obtained from the relevant parties before the research activities commenced. Students and their parents or legal guardians were informed about the purpose, procedures, and voluntary nature of participation. The participants' involvement in the study was carried out with appropriate consent.

Use of Generative AI: ChatGPT was employed as a supplementary learning tool during the vocabulary instruction sessions investigated in this study. Its use formed part of the research intervention and was guided by the instructor. Generative AI tools were also used to assist with English language refinement and manuscript readability. The research design, data collection, statistical analysis, interpretation of results, and conclusions remained the responsibility of the authors. The authors reviewed the AI-assisted content and accept full responsibility for the accuracy and integrity of the published work.

REFERENCES

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: Effects on accuracy and L2 motivation. *Computer Assisted Language Learning*, 1–25. https://doi.org/10.1080/09588221.2025.2534015

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.*

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Muh. Alfachrezy, et al.

Houghton Mifflin.

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), Ha*ndbook of test development* (pp. 3–25). Lawrence Erlbaum Associates.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. Language Learning, 51(3), 539–558. *https://doi.org/10.1111/0023-8333.00164*

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. R*EID (Research and Evaluation in Education),* 10*(*1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Günnemann, S., Hüllermeier, E., & Krusche, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learnin*g and Individual Differences, 103,* Artic*l*e 102274. https://doi.org/10.1016/j.lindif.2023.102274

Knapp, T. R. (2016). Why is the one-group pretest-posttest design still used? Clini*cal Nursing Research, 25(5*), 467–472. https://doi.org/10.1177/1054773816666280

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

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. P*sychometrika,* 2(*3*), 151–160. https://doi.org/10.1007/BF02288391

Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation. *Computer Assisted Language Learning*, *19*(1), 15–45. https://doi.org/10.1080/09588220600803998

Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. Oxford Review of Education, 38(5), *583–616 https://doi.org/10.1080/03054985.2012.731208*

Nation, I. S. P. (2001). *Learning vocabulary in another language.*

Sapan, M., & Uzun, L. (2024). The effect of ChatGPT-integrated English teaching on high school EFL learners' writing skills and vocabulary development. *International Journal of Education in Mathematics, Science and Technology*, *12*(6), 1679–1699. https://doi.org/10.46328/ijemst.4655

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). E*xperimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Webb, S., & Nation, I. S. P. (2017). How vocabulary is le*arned. Oxford University Press https://doi.org/10.25170/ijelt.v12i1.1458*

Zahran, F. A. (2025). The impact of using Poe ChatGPT-based TPACK model on English as a foreign language teachers' performance and their students' vocabulary learning. *Higher Learning Research Communications*, *15*(1), 1–11. https://doi.org/10.18870/hlrc.v15i1.1622

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.