

Exploring Vocabulary Gains Through ChatGPT-Supported Instruction in Indonesian Vocational Education

Received: April 03, 2026; Revised: May 03, 2026; Accepted: 04 June 2026

Eni Damayanti^{1*}, Hafirah Patang²

¹²Cokroaminoto Palopo University, Indonesia

✉ * Corresponding author: enimarwel@gmail.com

Abstract: Vocabulary knowledge is essential for English communication and vocational learning, yet many English-as-a-foreign-language learners receive limited opportunities for contextualized lexical practice and immediate feedback. This study examined vocabulary gains following ChatGPT-supported instruction among Grade XI Computer and Network Engineering students at an Indonesian public vocational school. A quantitative pre-experimental, one-group pre-test–post-test design was employed with 25 students selected through class-level random sampling from a population of 567 Grade XI students. Vocabulary mastery was measured using a 20-item multiple-choice test covering nouns, verbs, adjectives, and adverbs related to personal money management. The intervention comprised four teacher-guided meetings in which ChatGPT supported dialogue generation, sentence construction, fill-in-the-blank exercises, vocabulary games, word searches, explanations, examples, and feedback. Data were analyzed using descriptive statistics, Shapiro–Wilk tests, and a paired-samples *t* test. The mean score increased from 54.20 (SD = 14.045) on the pre-test to 77.20 (SD = 12.754) on the post-test, yielding a mean gain of 23.00 points. The paired difference was statistically significant, $t(24) = -11.049$, $p < .001$, with a 95% confidence interval of $[-27.296, -18.704]$ for the pre-test-minus-post-test difference. Furthermore, 88% of students achieved good or very good post-test classifications, compared with 32% at pre-test. These findings demonstrate substantial within-group improvement and extend evidence on generative-AI-supported vocabulary instruction to an underrepresented Indonesian vocational context. However, the absence of a control group prevents exclusive causal attribution to ChatGPT. Future controlled studies should examine retention and productive vocabulary use.

Keywords: ChatGPT-supported instruction; English as a foreign language; generative AI; vocabulary mastery; vocational education.

1 Introduction

Vocabulary knowledge is a central component of English proficiency because learners must recognize word forms, access meanings, and use lexical items appropriately before they can comprehend or produce extended discourse. Recent syntheses show that technology-assisted vocabulary instruction can generate meaningful learning gains when digital activities provide repeated exposure, retrieval, and opportunities to connect form, meaning, and use (Hao et al., 2021; Yu & Trainin, 2022). This requirement is especially consequential in English-as-a-foreign-language settings, where classroom instruction may constitute learners' principal source of sustained English input and lexical practice. Mobile and computer-supported

environments can extend such practice beyond teacher-fronted explanation by offering multimodal materials, flexible pacing, immediate response, and frequent rehearsal, although outcomes vary according to task design and learner characteristics (Burston & Giannakou, 2022; Okumuş Dağdeler, 2023). Consequently, vocabulary development should be approached not as the memorization of isolated translations but as a cumulative process requiring contextual encounters, active production, corrective information, and deliberate recycling across meaningful communicative tasks.

The need for purposeful lexical instruction is particularly visible in vocational education, where students must develop general English while also encountering terminology associated with occupational knowledge, workplace communication, and specialized study. Observations in upper-secondary vocational classrooms indicate that vocabulary receives substantial attention, yet instruction may remain dominated by translation and provide limited opportunities to reuse target words productively across language skills (Gay, 2022; Skarpaas & Rødnes, 2022). Digital learning can diversify this experience through interactive practice, contextual examples, games, and learner-controlled repetition that make lexical work more responsive to heterogeneous proficiency levels. Meta-analytic evidence suggests that digitally mediated language activities can improve second-language outcomes, but their effectiveness depends on pedagogical integration, feedback, duration, and alignment between activities and learning objectives rather than technology alone (Dixon et al., 2022; Shahiwala & Rahul, 2026). For Indonesian vocational learners, an instructional approach should therefore combine curricular vocabulary targets with accessible interaction, guided production, and timely feedback while remaining feasible within ordinary classroom conditions.

Despite its importance, vocabulary learning frequently becomes difficult when students meet unfamiliar forms without sufficient contextual support, repeated retrieval, or chances to test their understanding through interaction. Reviews of language-learning research involving ChatGPT report that most studies have concentrated on university students, writing, and learner perceptions, leaving school-level instruction and discrete language outcomes comparatively underexamined (Lo et al., 2024; Yang & Li, 2024). This imbalance matters because adolescents in vocational schools differ from tertiary language majors in proficiency, curricular demands, digital autonomy, and the amount of teacher scaffolding they require. Evidence from K–12 research remains limited and methodologically uneven, while early language-education publications call for broader participant groups, classroom-based interventions, and direct assessments of learning rather than reliance on self-reported usefulness (Li et al., 2024; Zhang & Tur, 2024). A defensible response is therefore to investigate ChatGPT as a teacher-guided instructional resource whose contribution is evaluated through observable vocabulary performance in an authentic secondary vocational classroom.

ChatGPT offers a general route for addressing these constraints because its conversational interface can generate definitions, examples, explanations, questions, and reformulations in response to learners' prompts. Language-education scholarship identifies personalization, dialogic practice, rapid feedback, and adaptable materials as promising affordances, but it also stresses that productive use requires carefully framed prompts, explicit learning goals, and teacher mediation (Kasneci et al., 2023; Kohnke et al., 2023). In vocabulary lessons, these capabilities can support movement from noticing a word to understanding its meaning and applying it in sentences or short exchanges. However, generated responses may be inaccurate, excessively confident, culturally inappropriate, or mismatched to students' proficiency, and unguided use can encourage superficial completion or dependence rather than durable learning (Farrokhnia et al., 2024; Tlili et al., 2023). Classroom integration should therefore position ChatGPT as a structured practice partner within planned instruction, with the teacher selecting target items, sequencing tasks, checking outputs, clarifying errors, and connecting interaction to assessment.

A specific instructional solution is to organize ChatGPT-supported vocabulary work around elicitation, contextualization, production, and feedback. Chatbot-assisted dynamic assessment has demonstrated that graduated prompts and contingent assistance can diagnose vocabulary difficulty while helping learners progress toward independently constructed responses, whereas studies with younger EFL learners show that conversational affordances can create opportunities for low-pressure interaction and sustained engagement (Jeon, 2021, 2022). Such principles permit a teacher to prompt definitions and examples, ask learners to identify parts of speech, require them to reorder or complete sentences, and then compare their answers with immediate explanations. Experimental evidence further indicates that large-language-model chatbots can benefit receptive and productive vocabulary acquisition, while text-based pedagogical chatbots are perceived as intelligible and supportive when interactions remain comprehensible and instructionally directed (Alrajhi, 2024; Zhang & Huang, 2024). Accordingly, ChatGPT-supported tasks should not merely display word meanings; they should require learners to retrieve, manipulate, and use nouns, verbs, adjectives, and adverbs in contexts connected to lesson content.

Implementation requires a coherent relationship among teacher guidance, learner activity, and the AI-generated response. Learners report that ChatGPT can expand vocabulary and offer real-time explanations, adaptive practice, and a psychologically supportive space, while evidence shows that interactivity, enjoyment, usefulness, and trust influence continued adoption for self-directed English learning (Rezai et al., 2024; Tram et al., 2024). In a classroom intervention, these affordances can be translated into short cycles: the teacher introduces a lexical category, students query or respond to ChatGPT individually or collaboratively, generated examples are evaluated, and the class consolidates accurate usage through follow-up exercises. Nevertheless, positive perceptions cannot substitute for measured achievement, because novelty, assessment design, and weak controls may inflate apparent benefits. Meta-analytic findings indicate generally favorable effects of ChatGPT on achievement but also identify methodological limitations, while chatbot research shows that learning effects vary with educational level, intervention design, and the roles assigned to teachers and conversational agents (Deng et al., 2025; Wang et al., 2025).

The closest literature supports conversational AI as a useful language-learning partner, yet it does not establish how a curriculum-aligned ChatGPT intervention affects vocabulary scores among Indonesian vocational students. Systematic reviews reveal that ChatGPT research has been concentrated in higher education and often addresses acceptance, writing, or conceptual opportunities rather than controlled measurement of school learners' lexical gains (Baig & Yadegaridehkordi, 2024; Crompton & Burke, 2023). Research on technology-enhanced vocabulary learning likewise confirms the value of deliberate task design, but it spans heterogeneous tools and populations that cannot be assumed to represent generative conversation in a vocational classroom. Recent reviews therefore call for diversified educational settings, skill-specific outcomes, transparent intervention procedures, and evidence concerning feedback quality and sustained learning (Aljohani, 2026; Zhou et al., 2024). These limitations leave a contextual and empirical gap: little is known about whether teacher-structured ChatGPT activities correspond with measurable changes in English vocabulary mastery for upper-secondary vocational learners studying nouns, verbs, adjectives, and adverbs.

This study examined vocabulary gains following ChatGPT-supported instruction among Grade XI Computer and Network Engineering students at an Indonesian vocational school. It specifically compared learners' performance before and after four classroom treatment sessions in which ChatGPT supported guided activities involving nouns, verbs, adjectives, and adverbs within the topic of personal money management. The study is novel in placing a directly assessed lexical outcome at the center of a teacher-structured generative-AI intervention in an underrepresented vocational secondary context, rather than examining

university learners' perceptions or writing performance. It was hypothesized that posttest vocabulary scores would be significantly higher than pretest scores because the intervention combined contextual examples, interactive questioning, productive manipulation of target forms, and immediate explanatory feedback. The scope was deliberately limited to one intact class of 25 students, a 20-item multiple-choice vocabulary measure, and a one-group pretest-posttest design; accordingly, the findings indicate within-group change during the intervention period and are not intended to establish comparative effectiveness or unrestricted causality.

2 Methods

This quantitative study employed a pre-experimental, one-group pretest-posttest design to examine changes in students' English vocabulary mastery following ChatGPT-supported instruction. The research was conducted at SMK Negeri 1 Palopo, a public vocational secondary school in Palopo, Indonesia, during the 2025/2026 academic year, specifically from April to May 2026. The target population comprised all 567 Grade XI students enrolled at the school. Class XI TKJ 4, consisting of 25 students from the Computer and Network Engineering program, was selected as the research sample. The thesis described the selection procedure as simple random sampling because every Grade XI class had an equal opportunity to be chosen. More precisely, the procedure constituted class-level random sampling because an intact class, rather than individual students from across the population, was selected as the experimental unit. All students in the selected class participated in the pretest, instructional treatment, and posttest, with no separate control or comparison group included in the research design.

Vocabulary mastery was measured using a researcher-developed test comprising 20 multiple-choice items, each with four response options. The instrument assessed students' ability to recognize word meanings, select vocabulary appropriate to a given context, and distinguish among nouns, verbs, adjectives, and adverbs associated with personal money management. Its content was developed from the English learning materials prescribed for Grade XI. Before the intervention, students completed the pretest to establish their initial vocabulary performance. They subsequently received four instructional meetings integrating ChatGPT as an interactive learning resource. The first meeting used an AI-generated dialogue to identify nouns related to money, banking, shopping, and payments. The second involved arranging jumbled words, identifying verbs, and requesting meanings and example sentences from ChatGPT. During the third meeting, students completed an AI-generated fill-in-the-blank activity and an adjective quiz in pairs. The fourth meeting addressed adverbs through a word-search activity, followed by ChatGPT-generated definitions, examples, and explanatory feedback. A comparable posttest was administered after the four treatment sessions.

Each correct response on the vocabulary tests was awarded one point, and the resulting raw score was converted to a percentage by dividing the number of correct responses by 20 and multiplying the result by 100. The scores were subsequently classified into five achievement levels: very good (81–100), good (61–80), average (41–60), poor (21–40), and very poor (0–20). Descriptive statistics, including frequencies, percentages, means, and standard deviations, were calculated to summarize the students' vocabulary performance before and after the intervention. Gain scores were obtained by subtracting each participant's pretest score from the corresponding posttest score. The normality of the pretest, posttest, and gain-score distributions was assessed using the Shapiro–Wilk test. After the assumptions for parametric analysis had been evaluated, a paired-samples *t* test was conducted in IBM SPSS Statistics to determine whether the mean posttest score differed significantly from the mean pretest score. Statistical significance was evaluated using a two-tailed alpha level of .05.

3 Results

This section reports the quantitative findings obtained from the vocabulary pre-test and post-test administered to 25 Grade XI Computer and Network Engineering students. The analysis was organized in four stages. First, descriptive statistics were used to summarize the distribution, central tendency, and variability of the students' scores before and after the ChatGPT-supported lessons. Second, Shapiro–Wilk tests were applied to evaluate the normality of the pre-test, post-test, and gain-score distributions. Third, a paired-samples t test examined whether the difference between the two measurements was statistically significant. Finally, the magnitude of the observed within-group change was considered to clarify its practical meaning. The findings are presented sequentially through tables followed by concise interpretations. Because the study employed one group without a comparison condition, the results describe changes observed during the intervention period; they do not independently establish that ChatGPT was the sole cause of those changes. Inferential conclusions were evaluated using a two-tailed significance level of .05.

3.1 Descriptive Results of Students' Vocabulary Mastery

Descriptive analysis was conducted to establish the students' vocabulary performance at both measurement points before undertaking inferential testing. The analysis included the number of participants, minimum and maximum scores, total scores, means, standard deviations, mean gain, and distributions across the five achievement classifications specified in the study. Table 1 summarizes these statistics and classifications for the pre-test and post-test. The computations were based on the 25 paired observations reported in the thesis. The post-test total is presented as 1,930 because this value is obtained by summing the individual scores and is consistent with the reported post-test mean of 77.20; the narrative total of 1,940 in the thesis appears to be an arithmetic transcription error. The score categories were very good (81–100), good (61–80), average (41–60), poor (21–40), and very poor (0–20). Together, the descriptive indicators provide an initial account of the direction, size, and distribution of the observed vocabulary-score changes.

Table 1: Descriptive Statistics and Achievement Classification of Vocabulary Scores

Measure	Pre-test	Post-test
Number of students	25	25
Minimum score	15	50
Maximum score	85	100
Total score	1,355	1,930
Mean	54.20	77.20
Standard deviation	14.045	12.754
Mean gain	—	23.00
Very good (81–100)	1 (4%)	9 (36%)
Good (61–80)	7 (28%)	13 (52%)
Average (41–60)	14 (56%)	3 (12%)
Poor (21–40)	2 (8%)	0 (0%)
Very poor (0–20)	1 (4%)	0 (0%)

Note. Percentages are based on 25 students. Mean gain was calculated by subtracting the pre-test mean from the post-test mean.

Table 1 shows a clear upward shift in vocabulary performance between the two measurement occasions. The mean increased from 54.20 on the pre-test to 77.20 on the post-test, producing an observed mean gain of 23.00 points. The minimum score rose from 15 to 50,

while the maximum increased from 85 to 100. Score variability decreased moderately, as indicated by the reduction in the standard deviation from 14.045 to 12.754. The achievement distributions changed accordingly. Before the intervention, 14 students (56%) were classified as average, whereas only eight students (32%) occupied the good or very good categories. At the post-test, 22 students (88%) were classified as good or very good, three (12%) remained in the average category, and none were categorized as poor or very poor. Thus, the descriptive results indicate that improvement occurred across the score distribution rather than only among the highest-performing students. Statistical significance, however, requires the inferential analysis reported subsequently.

3.2 Assessment of Normality

Before testing the difference between students' pre-test and post-test scores, the distributions of both measurements were examined for normality. The Shapiro–Wilk test was selected because the analysis involved 25 paired observations and was intended to assess whether substantial departures from normality were evident in the score distributions. The null hypothesis for each test stated that the corresponding scores were drawn from a normally distributed population. A probability value greater than the predetermined alpha level of .05 was interpreted as insufficient evidence to reject this assumption, whereas a value below .05 would indicate a statistically significant departure from normality. Table 2 presents the Shapiro–Wilk statistic, degrees of freedom, and probability value for the pre-test and post-test scores. This assessment preceded the paired-samples comparison and was used in the thesis to determine whether parametric analysis was considered appropriate. The available statistical output reported normality results for the two measurement occasions but did not provide a separate test for the paired difference scores.

Table 2: Shapiro–Wilk Normality Test Results

Measurement	Shapiro–Wilk statistic	<i>df</i>	<i>p</i>
Pre-test	.939	25	.144
Post-test	.966	25	.535

Note. The significance criterion was set at $\alpha = .05$.

As shown in Table 2, the pre-test scores yielded a Shapiro–Wilk statistic of .939 and a probability value of .144. Because this probability exceeded the .05 criterion, the null hypothesis of normality was not rejected for the pre-test distribution. The post-test scores produced a Shapiro–Wilk statistic of .966 with a probability value of .535, which likewise exceeded .05 and provided no statistically significant evidence of non-normality. Thus, neither set of observed scores demonstrated a significant departure from a normal distribution according to the reported tests. On this basis, the thesis proceeded with a parametric paired-samples *t* test to evaluate the difference between the mean scores. Nevertheless, the statistical assumption underlying a paired-samples *t* test specifically concerns the distribution of the within-participant difference scores. Because a separate Shapiro–Wilk result for those differences was not included in the available output, the reported analysis supports normality of each measurement but does not independently document normality of the paired differences.

3.3 Difference Between Pre-test and Post-test Vocabulary Scores

A paired-samples *t* test was conducted to determine whether the students' vocabulary scores differed significantly between the pre-test and post-test measurements. This procedure was

appropriate because the same 25 participants completed both tests, producing two related sets of observations. The paired difference was calculated in the order pre-test minus post-test; consequently, a negative mean difference indicates that post-test performance was higher. Table 3 presents the means and standard deviations for both measurements, together with the mean paired difference, standard deviation and standard error of the differences, 95% confidence interval, *t* statistic, degrees of freedom, and two-tailed probability value. The null hypothesis stated that the population mean difference between the two measurements was zero, whereas the alternative hypothesis predicted a statistically significant difference. The decision criterion was set at $\alpha = .05$. A probability value below this threshold was therefore interpreted as sufficient evidence to reject the null hypothesis and conclude that vocabulary performance changed significantly across the two measurement occasions.

Table 3

*Paired-Samples *t*-Test for Pre-test and Post-test Vocabulary Scores*

Comparison	Pre-test <i>M</i> (<i>SD</i>)	Post-test <i>M</i> (<i>SD</i>)	Mean difference	<i>SD</i> of difference	<i>SE</i>	95% CI	<i>t</i>	<i>df</i>	<i>p</i>
Pre-test – Post-test	54.20 (14.045)	77.20 (12.754)	-23.000	10.408	2.082	[-27.296 , -18.704]	-11.049	24	<.001

Note. *N* = 25. The mean difference was calculated by subtracting the post-test score from the pre-test score. CI = confidence interval.

Table 3 indicates that the mean vocabulary score increased from 54.20 (*SD* = 14.045) on the pre-test to 77.20 (*SD* = 12.754) on the post-test. Because the paired difference was computed as pre-test minus post-test, the resulting mean difference was -23.00 points. The 95% confidence interval ranged from -27.296 to -18.704 and did not include zero, indicating that the direction of the difference was consistently negative under the specified subtraction order. The paired-samples analysis produced $t(24) = -11.049$ and $p < .001$. The probability value was below the .05 criterion; therefore, the null hypothesis of no mean difference was rejected, and the alternative hypothesis was supported. The students consequently obtained significantly higher vocabulary scores after the four ChatGPT-supported instructional meetings than before them. This finding demonstrates a statistically reliable within-group increase during the study period. However, because no comparison group was included, the analysis alone cannot isolate ChatGPT-supported instruction from other possible explanations for the observed improvement.

4 Discussion

The findings revealed a substantial within-group increase in vocabulary mastery, with the mean rising by 23 points from 54.20 on the pre-test to 77.20 on the post-test and the paired comparison reaching statistical significance. This pattern is consistent with recent meta-analytic evidence showing that AI-supported vocabulary interventions generally produce positive effects, although the magnitude varies according to learner characteristics, intervention duration, and instructional design (Gao et al., 2026; Şimşek et al., 2026). The result also corresponds with research indicating that ChatGPT-enhanced vocabulary instruction can improve immediate achievement and that well-designed conversational agents can expand opportunities for individualized language practice, feedback, and lexical retrieval (Albaqami & Alahmadi, 2026; Koç & Savaş, 2025). Nevertheless, the present gain should be interpreted as improvement observed during four teacher-guided sessions rather than evidence that unrestricted ChatGPT use is inherently beneficial. Its educational significance lies in

demonstrating that generative dialogue can be incorporated into ordinary vocational instruction while maintaining a clear focus on assessed vocabulary outcomes.

The improvement may be related to the way the intervention combined contextual exposure with active manipulation of vocabulary rather than presenting words as isolated translations. Research on ChatGPT-supported foreign-language instruction indicates that prompting can provide examples and repeated practice, while meta-analytic evidence on computer-mediated feedback shows that timely responses can strengthen vocabulary learning when aligned with purposeful tasks (Karataş et al., 2024; Li, 2023). In this study, students identified nouns in a financial dialogue, reconstructed sentences containing verbs, completed adjective items, and examined adverbs through a word-search activity, thereby encountering lexical items through several forms of processing. Studies of adolescent application use and multimodal vocabulary environments emphasize curriculum alignment, immediate feedback, social support, and combined verbal or contextual cues as conditions that help learners engage with and reuse new words (Lin et al., 2022; Wei, 2023). The score increase is therefore plausibly associated with the instructional sequence as a whole, including teacher direction, task variation, peer activity, and AI-generated assistance.

ChatGPT's immediate explanations and example sentences may have enabled students to resolve uncertainty while the target words were still active in working memory. Digital game-based vocabulary research suggests that interactive construction and meaningful contextualization can support engagement and lexical development, whereas augmented-reality studies similarly associate situated cues and learner interaction with improved vocabulary performance (Chowdhury et al., 2024; Khan et al., 2023). However, immediacy is pedagogically useful only when learners attend to the response, compare it with their own understanding, and apply the information in another task. Reviews of mobile vocabulary learning emphasize that technological access must be accompanied by learner responsibility and purposeful input–interaction–feedback cycles, while evidence from blended EFL instruction shows that targeted digital feedback can support achievement and self-regulated learning (Chen, 2024; Klimova, 2021). The four-session procedure reflected this principle by embedding ChatGPT within teacher-planned dialogue, sentence construction, quizzes, and review activities, instead of treating conversational output as a substitute for instruction or independent evidence of learning.

The movement across achievement categories indicates that change was distributed throughout the class: 88% of participants reached the good or very good categories at post-test, compared with 32% at pre-test, and none remained poor or very poor. Studies of ChatGPT-supported feedback show that combining automated suggestions with instructor judgment can make assistance more usable, while comparisons of human and GPT feedback demonstrate that generated responses may differ in focus and consistency (Han & Li, 2024; Lin & Crosthwaite, 2024). The decline in the score standard deviation, from 14.045 to 12.754, likewise suggests that improvement was not confined to a small group of stronger students. Evidence from secondary classrooms indicates that designed LLM feedback can support performance and motivation, while human–AI instructional collaboration can help teachers monitor difficulties and contextualize generated assistance (Meyer et al., 2024; Pahi et al., 2024). These comparisons support a complementary model in which ChatGPT broadens practice opportunities, but the teacher remains responsible for instructional decisions.

Although the statistical result was strong, several features of the evidence require caution. Learners' willingness to use ChatGPT is influenced by perceived usefulness, ease of use, and attitudes, while intervention research shows that favorable outcomes depend on how prompts, feedback, and revision activities are structured rather than on access to the platform alone (Liu & Ma, 2024; Polakova & Ivenz, 2024). ChatGPT can provide inaccurate, inconsistent, or overly elaborate responses, and conversational interfaces may place substantial demands on users' prompting knowledge and ability to judge output quality (Crompton & Burke, 2024;

Pozdniakov et al., 2024). More importantly, this study used one intact class, only 25 participants, a 20-item multiple-choice measure, four treatment sessions, and no comparison group or delayed post-test. The significant difference therefore cannot exclude maturation, repeated testing, concurrent classroom learning, novelty effects, or other uncontrolled influences. It also demonstrates gains in recognition-oriented vocabulary performance, not necessarily long-term retention, spontaneous productive use, or superiority over instruction delivered without ChatGPT.

The study contributes context-specific evidence by examining assessed vocabulary gains among Indonesian vocational secondary students, a population less visible in university-oriented ChatGPT literature. Indonesian EFL learners report using ChatGPT to clarify vocabulary and manage language difficulties, but they emphasize the need to preserve personal agency and evaluate whether generated language remains authentic and appropriate (Werdiningsih et al., 2024a; Werdiningsih et al., 2024b). For practice, the findings support teacher-structured cycles in which students encounter target words, formulate or analyze language, inspect AI feedback, and verify that feedback through classroom discussion. Research on self-regulated learning with generative AI and learner-centered feedback indicates that digital assistance becomes more meaningful when learners monitor goals, interpret feedback, and decide how to act on it (Chiu, 2024; Mandouit & Hattie, 2023). Future studies should employ larger multisite samples, randomized or matched comparison groups, validated measures of receptive and productive vocabulary, delayed post-tests, interaction logs, and qualitative observations to distinguish sustained learning from short-term score improvement..

5 Conclusion

This study identified a statistically significant increase in the English vocabulary scores of Indonesian vocational secondary students following ChatGPT-supported instruction. The mean score increased from 54.20 on the pre-test to 77.20 on the post-test, representing an average gain of 23 points. The paired-samples analysis confirmed that this difference was statistically significant, $t(24) = -11.049$, $p < .001$. The achievement distribution also shifted favorably: 88% of the students were classified in the good or very good categories at post-test, compared with 32% at pre-test, while no student remained in the poor or very poor categories. These findings indicate that vocabulary performance improved during the intervention period. The study contributes to the emerging literature by providing directly assessed classroom evidence from an underrepresented population of Indonesian vocational students and by demonstrating how ChatGPT can be integrated into structured activities addressing nouns, verbs, adjectives, and adverbs rather than being used only for unrestricted independent learning.

Pedagogically, the findings suggest that ChatGPT may function as a useful supplementary resource when its responses are embedded within teacher-designed activities involving contextual exposure, lexical retrieval, sentence construction, peer collaboration, and immediate feedback. Teachers should nevertheless verify generated content, formulate clear prompts, and ensure that students actively process vocabulary rather than merely reproduce AI-generated answers. The conclusions are limited by the one-group pre-test–post-test design, small sample of 25 students, short four-session intervention, single institutional setting, and reliance on a 20-item multiple-choice test. Without a comparison group, the observed improvement cannot be attributed exclusively to ChatGPT-supported instruction. Future research should employ randomized or matched control groups, larger multisite samples, longer interventions, and delayed post-tests. Validated measures of receptive and productive vocabulary, classroom observations, learner interviews, and interaction-log analysis should also be incorporated to examine retention, actual vocabulary use, engagement, and the specific instructional mechanisms through which ChatGPT may support vocational EFL learning.

References

- Albaqami, S., & Alahmadi, A. (2026). Immediate and delayed effects of ChatGPT-enhanced vocabulary instruction on Saudi EFL learners. *Information*, 17(9), 812. <https://doi.org/10.3390/info17090812>
- Aljohani, N. J. (2026). ChatGPT in language learning: A systematic review of applications and challenges. *Social Sciences & Humanities Open*, 13, 102357. <https://doi.org/10.1016/j.ssaho.2025.102357>
- Alrajhi, A. S. (2024). Artificial intelligence pedagogical chatbots as L2 conversational agents. *Cogent Education*, 11(1), 2327789. <https://doi.org/10.1080/2331186X.2024.2327789>
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Burston, J., & Giannakou, K. (2022). MALL language learning outcomes: A comprehensive meta-analysis 1994–2019. *ReCALL*, 34(2), 147–168. <https://doi.org/10.1017/S0958344021000240>
- Chen, J. (2024). Effects of learning analytics-based feedback on students' self-regulated learning and academic achievement in a blended EFL course. *System*, 124, 103388. <https://doi.org/10.1016/j.system.2024.103388>
- Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 72(4), 2401–2416. <https://doi.org/10.1007/s11423-024-10366-w>
- Chowdhury, M., Dixon, L. Q., Kuo, L.-J., Donaldson, J. P., Eslami, Z., Viruru, R., & Luo, W. (2024). Digital game-based language learning for vocabulary development. *Computers and Education Open*, 6, 100160. <https://doi.org/10.1016/j.caeo.2024.100160>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Crompton, H., & Burke, D. (2024). The educational affordances and challenges of ChatGPT: State of the field. *TechTrends*, 68(2), 380–392. <https://doi.org/10.1007/s11528-024-00939-0>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language gains through digital game-based language learning: A meta-analysis. *Language Learning & Technology*, 26(1), 1–25. <https://doi.org/10.64152/10125/73464>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Gao, K., Shi, X., & Zhang, J. (2026). Student–AI interaction: Unveiling the impact of AI agents on vocabulary learning—A meta-analysis. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2025-0382>
- Gay, F. (2022). Investigating the effects of technology-enhanced vocabulary learning strategy instruction on supporting mixed-ability EMI learners studying a journalism and communication major: An action research project at a university in China. *Journal of English for Academic Purposes*, 55, 101068. <https://doi.org/10.1016/j.jeap.2021.101068>

- Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502. <https://doi.org/10.1016/j.system.2024.103502>
- Hao, T., Wang, Z., & Ardasheva, Y. (2021). Technology-assisted vocabulary learning for EFL learners: A meta-analysis. *Journal of Research on Educational Effectiveness*, 14(3), 645–667. <https://doi.org/10.1080/19345747.2021.1917028>
- Jeon, J. (2021). Chatbot-assisted dynamic assessment for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, 36(7), 1338–1364. <https://doi.org/10.1080/09588221.2021.1987272>
- Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1–2), 1–26. <https://doi.org/10.1080/09588221.2021.2021241>
- Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29(15), 19343–19366. <https://doi.org/10.1007/s10639-024-12574-6>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khan, R. M. I., Ali, A., Kumar, T., & Venugopal, A. (2023). Assessing the efficacy of augmented reality in enhancing EFL vocabulary. *Cogent Arts & Humanities*, 10(1), 2223010. <https://doi.org/10.1080/23311983.2023.2223010>
- Klimova, B. (2021). Evaluating impact of mobile applications on EFL university learners' vocabulary learning: A review study. *Procedia Computer Science*, 184, 859–864. <https://doi.org/10.1016/j.procs.2021.03.108>
- Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, 37(1), 4–21. <https://doi.org/10.1017/S0958344024000168>
- 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>
- Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
- Li, R. (2023). Investigating effects of computer-mediated feedback on L2 vocabulary learning. *Computers & Education*, 198, 104763. <https://doi.org/10.1016/j.compedu.2023.104763>
- Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, 103529. <https://doi.org/10.1016/j.system.2024.103529>
- Lin, V., Yeh, H.-C., Huang, H.-H., & Chen, N.-S. (2022). Enhancing EFL vocabulary learning with multimodal cues supported by an educational robot and an IoT-based 3D book. *System*, 104, 102691. <https://doi.org/10.1016/j.system.2021.102691>
- Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of

- empirical studies. *Smart Learning Environments*, 11, 50. <https://doi.org/10.1186/s40561-024-00342-5>
- Mandouit, L., & Hattie, J. (2023). Revisiting “The power of feedback” from the perspective of the learner. *Learning and Instruction*, 84, 101718. <https://doi.org/10.1016/j.learninstruc.2022.101718>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students’ text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- Okumuş Dağdeler, K. (2023). A systematic review of mobile-assisted vocabulary learning research. *Smart Learning Environments*, 10, 19. <https://doi.org/10.1186/s40561-023-00235-z>
- Pahi, K., Hawlader, S., Hicks, E., Zaman, A., & Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. *Computers and Education Open*, 6, 100183. <https://doi.org/10.1016/j.caeo.2024.100183>
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students’ writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- Pozdniakov, S., Brazil, J., Abdi, S., Bakharia, A., Sadiq, S., Gašević, D., Denny, P., & Khosravi, H. (2024). Large language models meet user interfaces: The case of provisioning feedback. *Computers and Education: Artificial Intelligence*, 7, 100289. <https://doi.org/10.1016/j.caeai.2024.100289>
- Rezai, A., Namaziandost, E., & Hwang, G.-J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports*, 16, 100510. <https://doi.org/10.1016/j.chbr.2024.100510>
- Shahiwala, S., & Rahul, D. R. (2026). A Bayesian meta-analysis of digital game-enhanced vocabulary learning. *Journal of Computer Assisted Learning*, 42(3), e70246. <https://doi.org/10.1002/jcal.70246>
- Şimşek, B., Direkci, B., Koparan, B., Akbulut, S., Şimşek, E. E., Gülmez, M., & Ökmen, D. N. (2026). Artificial intelligence-assisted vocabulary teaching: A meta-analysis study. *Frontiers in Artificial Intelligence*, 9, 1806700. <https://doi.org/10.3389/frai.2026.1806700>
- Skarpaas, K. G., & Rødnes, K. A. (2022). Vocabulary teaching practices of L2 English in upper-secondary vocational classrooms. *Languages*, 7(1), 55. <https://doi.org/10.3390/languages7010055>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System*, 127, 103528. <https://doi.org/10.1016/j.system.2024.103528>
- Wang, F., Cheung, A. C. K., Neitzel, A. J., & Chai, C. S. (2025). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*, 95(4), 623–660. <https://doi.org/10.3102/00346543241255621>

- Wei, W. (2023). Understanding and supporting the use of feedback from mobile applications in the learning of vocabulary among young adolescent learners. *Studies in Educational Evaluation*, 78, 101264. <https://doi.org/10.1016/j.stueduc.2023.101264>
- Werdiningsih, I., Marzuki, & Rusdin, D. (2024a). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities*, 11(1), 2392388. <https://doi.org/10.1080/23311983.2024.2392388>
- Werdiningsih, I., Marzuki, Indrawati, I., Rusdin, D., Ivone, F. M., Basthomi, Y., & Zulfahreza. (2024b). Revolutionizing EFL writing: Unveiling the strategic use of ChatGPT by Indonesian master's students. *Cogent Education*, 11(1), 2399431. <https://doi.org/10.1080/2331186X.2024.2399431>
- Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, 103351. <https://doi.org/10.1016/j.system.2024.103351>
- Yu, A., & Trainin, G. (2022). A meta-analysis examining technology-assisted L2 vocabulary learning. *ReCALL*, 34(2), 235–252. <https://doi.org/10.1017/S0958344021000239>
- Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K–12 education. *European Journal of Education*, 59(2), e12599. <https://doi.org/10.1111/ejed.12599>
- Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second-language vocabulary acquisition. *Heliyon*, 10(3), e25370. <https://doi.org/10.1016/j.heliyon.2024.e25370>
- Zhou, K., Jin, F., Li, W., Song, Z., Huang, X., & Lin, C.-H. (2024). The design of technology-enhanced vocabulary learning: A systematic review. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12423-y>