*Ed Process Int J*

**www.edupij.com**

ISSN 2147-0901 (Print), 2564-8020 (Online)

EDUCATIONAL PROCESS
INTERNATIONAL JOURNAL

# Digital Scaffolding as Learning Opportunities: Enhancing Vocabulary Knowledge Through Copilot AI

**Miew Luan NG, Ali Al Ghaithi, Behnam Behforouz**

**Research Article**

**Cite this article:** NG, M. L., Al Ghaithi, A., & Behforouz, B. (2026). Digital scaffolding as learning opportunities: Enhancing vocabulary knowledge through Copilot AI. *Educational Process: International Journal, 20*, e2026012. https://doi.org/10.22521/edupij.2026.20.12

**Received** July 5, 2025
**Accepted** December 16, 2025

**Keywords:** Artificial Intelligence, learning opportunities and empowerment, Copilot AI, vocabulary learning

**Author for correspondence:**
Miew Luan NG
miewluan.ng@newinti.edu.my
INTI International University, Malaysia[1]; Shinawatra University, Thailand[2]

## Abstract

**Background/purpose.** The primary objective of this paper was to measure the effect of Copilot AI on vocabulary improvement among language learners.

**Materials/methods**. The study involved 90 Omani students with pre-intermediate English proficiency, randomly divided into three groups: control, experimental A, and experimental B, with equal allocation. While all groups received in-class training, experimental A used Copilot in the classroom setting, and experimental B used Copilot outside classes for extra vocabulary practice and feedback. Some researcher-developed tests were prepared as pretests, posttests, and delayed posttests, with high validity and reliability, to measure the impact of the treatment and compare results promptly.

**Results**. The study's findings revealed that all groups made significant progress in their posttests; however, experimental group A performed the best, followed by experimental group B and the control group. Later, the analysis of the delayed posttest revealed progress in experimental group A with no sign of improvement in the other two groups. Additionally, in this delayed posttest, the control group's scores decreased dramatically.

**Conclusion**. As an educational facilitator, Copilot helped improve students' vocabulary. Although the study's results emphasize the advantages of Copilot AI in the learning process, further studies could examine students with varying proficiency levels and other subskills, such as speaking, reading, and writing. Additionally, qualitative studies could give a clear picture of students' satisfaction in using AI tools as learning facilitators.

OPEN ACCESS

© The Author(s), 2025. This is an Open Access article, distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction, provided the original article is properly cited.

## 1. Introduction

Vocabulary competence is often regarded as a crucial aspect of learning a foreign language, and a lack of it can hinder clear communication (Susanto, 2017). Schmitt (2000) emphasizes that knowing words is essential for learning a second language and effective communication. Nation (2001) describes the link between understanding words and practising a language as complementary. Tellier (2008) and Susanto (2016) have both found that learning new words is crucial for effectively utilizing a foreign language and for producing complete texts, both written and spoken. It is essential to learn new words to enhance all language abilities, including reading, writing, speaking, and listening (Nation, 2001). Altyari (2017) found that EFL students with insufficient vocabulary knowledge perform worse in English language learning. Khan (2011) examined vocabulary areas that are challenging for those learning English as a foreign language (EFL), including word meanings, spelling, and the use of synonyms, prefixes, and suffixes.

Researchers have explored the use of digital and AI-driven technologies as alternatives to traditional language teaching methods to address these issues. Using computers in the classroom helps students learn new words, which in turn opens more educational options beyond conventional classrooms. Recent research on vocabulary acquisition indicates that technology enhances language learning by making it more engaging, dynamic, and practical in real-life applications (Chen et al., 2021; Lin & Lin, 2019; Tu et al., 2020). Hao et al. (2021) found that using technology to help students learn L2 vocabulary yielded significantly better results than traditional schooling without technology. Among recent technological developments, Artificial Intelligence (AI) is a technological innovation aimed at advancing procedures and techniques for high-level machine reasoning, utilizing low-level derived features without explicit human oversight. Artificial intelligence has become a valuable tool because it can mimic human thought processes, tailor instruction to each student's individual needs, and provide immediate feedback (Healey, 2020). Artificial intelligence (AI) has made it easier to learn new words on mobile devices by enabling computers to perform tasks similar to those of sophisticated people, such as decision-making, judgment, and concluding (Hwang et al., 2020). Traditional digital learning settings facilitate the sharing of knowledge in multiple ways. On the other hand, artificial intelligence offers students personalized learning paths and resources tailored to their individual performance (Diwan et al., 2023). For example, image-to-text generative artificial intelligence is a type of visual feedback system that aids people in learning new words (Jia et al., 2022; Shadiev et al., 2020). Mohd Amin et al. (2025) argued that artificial intelligence acts as a digital scaffold by providing adaptive feedback, personalized support, and contextualized vocabulary assistance, thereby improving learners' understanding and memory in real-life learning settings.

Among all these technological advancements in AI, chatbots and virtual assistants have gained significant importance in recent years (Clarizia et al., 2018). The goal of educational chatbots has expanded significantly to include creating educational materials, such as those used by a teacher (Han & Lee, 2022; Pérez et al., 2020). A significant amount of research has been conducted on chatbot technology, indicating that this approach is superior to others in answering students` questions (Sinha et al., 2020), evaluating students' performance abilities and offering administrative services (Durall & Kapros, 2020; Rohrig & Heß, 2019), helping students to learn together (Schmulian & Coetzee, 2019), allowing communication in more than one way (Haristiani et al., 2019), giving scaffolding feedback in real time (Gonda et al., 2018), making learning more personalized (Oke & Fernandes, 2020; Verleger & Pembridge, 2019), making sure that it can be scaled up, making it more interactive (Dekker et al., 2020), and effectively promoting the creation and sharing of knowledge (Verleger & Pembridge, 2019).

Yang et al. (2022) demonstrate that chatbots can identify grammatical errors independently during interactive exercise sessions. They also provide a dynamic discussion environment that helps students improve their dialogue skills, vocabulary, and understanding of language structures.

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

Behforouz and Al Ghaithi (2024) found that chatbots assist language learners in developing language skills and are beneficial within the learning process, especially for English reading and language skills. Zhou et al. (2023) found that current NLP improvements enhance chatbots' feedback quality for both grammar and word preference, demonstrating how contemporary chatbot creation might help people learn languages more effectively. Chen et al. (2020) referred to chatbots as personalized instructional assistants or learning companions that help students feel more confident and engaged.

Currently, there is limited empirical research on AI chatbots in EFL contexts (Kim et al., 2020), and several issues about their functional use in language acquisition remain unresolved (Yanguas, 2010). Moreover, the efficacy of these tactics varies with proficiency level. Chatbots are most effective for experienced foreign language learners and are unsuitable for beginners who require foundational knowledge (Fryer & Carpenter, 2006). Maraoengsit et al. (2019), Liu et al. (2020), and Yin et al. (2021) asserted that to gain a comprehensive understanding of the roles and significance of chatbots in educational contexts, it is essential to conduct further research across diverse environments to assess their impact on learning. Education in Oman, particularly at the beginning, focuses on developing general skills and sub-skills, with a strong emphasis on vocabulary development. However, despite technological advancements in language education, empirical research on the impact of AL tools on vocabulary development in Oman remains limited. Therefore, this research examined the use of Copilot, an interactive AI tool, to assist Omani English as a Foreign Language (EFL) students in enhancing their vocabulary skills. Given the importance of vocabulary acquisition in Oman's language learning curriculum, the findings of this study support the implementation of interactive AI tools to improve vocabulary learning and enhance learners' outcomes.

Therefore, the following research question is addressed in this study:

1. Does the use of Copilot as an extra facilitator of language learning improve the vocabulary knowledge of Omani pre-intermediate EFL learners?

## 2. Literature Review

Khalil et al. (2025) examined how AI chatbots can assist students in learning English grammar and vocabulary. The mixed-methods research study divided 60 ESL students into two groups over six weeks. The experimental group consisted of 30 students who used AI chatbots, while the control group comprised 30 students who learned in the usual manner. The experimental group showed significant improvements in grammar and vocabulary, whereas the control group did not change substantially. The study found that students were highly engaged because they received immediate, customized attention and feedback. Nevertheless, the study was restricted by unregulated external factors, such as learner knowledge and motivation levels, which were not accounted for in the assessment.

Similarly, Hutauruk et al. (2024) investigated the impact of AI chatbots in helping college students learn more English words. Using a qualitative research approach, 30 first-semester students from the English department were selected for the study. Students were asked about their use of AI chatbots in vocabulary learning, and the results showed that AI chatbots significantly aided their learning of new words. Students also found the chatbot to be enjoyable and easy to use. The study indicated that students were highly engaged and interested, and that participants believed the chatbot helped them obtain immediate feedback and learn independently on mobile devices. The study only examined students with similar language skills, which limited its generalizability to other populations. Future research should focus on students with a broader range of language skills.

Building on these findings, Zhang and Huang (2024) studied how chatbots that employ large language models (LLMs) changed the way people learnt vocabulary in a second language. This mixed-

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

methods quasi-experimental study randomly assigned 52 foreign-language learners to experimental and control groups over an 8-week period. The experimental group was given access to chatbots as learning supporters, while the control group was given access to alternative digital resources. The findings showed that the experimental group did far better than the control group on both immediate and delayed vocabulary tests. The study discovered that LLM-based chatbots not only helped second-language learners acquire new words and pick up new ones by accident, but also that students were quite interested in chatting with chatbots frequently. The study was limited, however, as further research is needed to examine students' levels of interest and pleasure, changes in motivation, and how these factors affect incidental and self-directed learning.

In a related study, Silitonga et al. (2024) conducted an experimental investigation to assess the effectiveness of AI chatbots in supporting students' learning of English for Specific Purposes (ESP) vocabulary. The research included 40 undergraduate students from the Science Department, divided into two groups: an experimental group and a control group. The experimental group employed Dialogflow chatbot technology for 16 weeks, while the control group underwent regular instruction. Results showed that the experimental group's vocabulary learning was statistically significantly better than that of the control group. The study's results show that AI chatbots provide engaging environments for deep conversation, which significantly help second-language learners improve their vocabulary. It also found that chatbots should be used in ESP lessons. However, the study was limited, as further research is needed in EFL and ESL settings with a larger sample size.

Correspondingly, Al Ghaithi et al. (2023) investigated how effectively WhatsApp bots aided Omani EFL students in learning new English words. The study comprised 150 students from three distinct proficiency levels: elementary, pre-intermediate, and intermediate. Each level consisted of 25 students in both the experimental and control groups. The students in the experimental group received extra vocabulary practice and feedback through the use of an interactive chatbot developed by the researchers. The findings showed that the experimental groups' ability to learn and remember new words increased a lot. Students have mixed feelings about using chatbots to learn a language. The study was limited, however, because the bot taught only one way. Interactive bots that allow students to communicate with one another could be better.

In another significant study, Waziana et al. (2024) investigated how Indonesian EFL students perceived the impact of AI chatbots on their grammar and vocabulary when writing. The study employed a mixed-methods approach and included 100 undergraduate students from five different schools. The experimental group was given access to chatbots to enhance their writing, with a focus on vocabulary and grammar, by receiving feedback from the AI tools. The findings revealed that the general quality of writing, vocabulary, and sentence structure variety all improved. Some of the primary benefits included improved vocabulary, learning in context, maintaining a consistent voice, increased grammatical accuracy, reduced repetition, and clearer, more coherent writing by students. The study's small sample size from a specific area, on the other hand, limited the usefulness of the conclusions for other groups of people.

## 3. Methodology

This research study employs a quantitative research approach, utilizing a quasi-experimental design to investigate the impact of Copilot AI on vocabulary development. This section includes detailed information about the study population, the instruments used to collect data, and the statistical analysis used to compare students' results.

### *3.1. Participants*

The sample population of this study consisted of 90 Omani students with pre-intermediate English proficiency, as determined by the university's placement test. They were between 18 and 21

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

years old and spoke Arabic as their primary language. These students were selected from a higher education institution in Oman and the Foundation Department. Every student in Oman had to complete the General Foundation Program, which took one to two academic years, as a prerequisite for moving on to their specialities. The GFP program offers students a range of courses, including English, Math, and Computing Skills, with English as the predominant medium of instruction.

Using random sampling, participants were divided into three groups, each comprising 30 students with a combination of males and females. Experimental Group A used Copilot on WhatsApp in the classroom, while Experimental Group B used Copilot on WhatsApp to practice vocabulary outside of class. Moreover, the control group that only learnt language in person, the old-fashioned way, without any help from AI. Since there were no clear instructions or permission to monitor students outside the classroom, and students were not readily available for monitoring in an experimental, controlled setting, there was no control over the use of technological instruments, such as generative AI tools, among the students. However, students in the experimental groups could be monitored by checking the history of their Copilot accounts. Additionally, due to the college's policy and class structure, there was no control over students' age and gender.

### *3.2. Instrumentations*

#### *3.2.1. Researcher-Made Vocabulary Tests*

The following tools were used to obtain the necessary information: three tests developed by researchers to assess whether Copilot affected Omani students' ability to use English vocabulary efficiently. They were supposed to deliver these tests as a pretest, a posttest, and a delayed posttest. The total score for each test was 20, which means 1 point for each correct response; in this case, a negative response will not incur any penalty marks. The first part consisted of 10 multiple-choice questions with four options, followed by another 10 fill-in-the-blank questions.

Before initiating the main study, 30 Omani English language learners from the same educational institutions and with the same level of English proficiency took these tests for piloting. Participants had to finish the test in 60 minutes. The following table shows the results of the reliability measurement.

**Table 1.** The Reliability Index For the Pretest, Posttest, and the Delayed Posttest

| Cronbach's Alpha | N of Items |
|---|---|
| .765 | 3 |

Based on the research, Cronbach's Alpha reliability for the pilot tests was 0.765 for all the tests. This shows that the tests were quite reliable. To validate the questions, three Omani graduates with doctorates in Applied Linguistics and over ten years of research and teaching experience reviewed the tests' appropriateness and suitability in cultural and contextual terms.

### *3.3. Procedures*

The study took place in the fall of 2025, throughout the third semester at one of the universities in Oman. Participants were informed that their participation is voluntary and will not affect their continuous assessment throughout the semester. The study had three groups: the first treatment group used Copilot on WhatsApp to learn vocabulary in the educational setting, in addition to in-class training; the second treatment group used Copilot on WhatsApp to learn vocabulary outside of the classroom, in addition to in-class training; and the control group used traditional methods to learn vocabulary. The procedure took seven weeks, which included pretesting, a four-week treatment session, post-testing, and a postponed post-testing session.

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

Initially, a vocabulary pretest was administered to all students a week prior to the start of the program. After that, participants in both treatment groups were shown how to use Copilot on WhatsApp to do vocabulary exercises, learn new words, and practice them. Microsoft Teams included lesson files, discussion boards for peer engagement, and polling tools to facilitate quick feedback on any issues with the Copilot on WhatsApp. The researcher also ensured that all students could communicate with and engage the Copilot on WhatsApp. All teachers were required to attend a two-day training program covering procedures for implementing and monitoring, as well as strategies for engaging students. Teachers also attended the weekly calibration sessions to address any issues that arose in practice.

Students in Experimental Group A (the Copilot in the class) were divided into six smaller groups, each with five students. During the 60-minute session, in the first 30 minutes, the instructor transmitted 10 words on Microsoft Teams, and the students copied them. After that, they started using Copilot on WhatsApp to understand what each phrase meant, come up with examples, and create tasks such as multiple-choice questions and fill-in-the-blank questions for the 10 words to practice them. The following 30 minutes were allocated to the clarification stage. The instructor asked and answered questions about the 10 words, providing further explanations as needed to ensure that all students understood their meanings. After that, students were instructed to use WhatsApp's Copilot tool to create additional practice activities for the 10 words studied in class, and the teacher provided help if needed. This process was always followed throughout weeks 1–4.

The learners in the second treatment group (Outside-classroom, Copilot on WhatsApp) had regular 20-minute classroom lessons. After that, the instructor gave the students a 25-minute exercise on the 10 words, followed immediately by a 15-minute discussion to clarify their answers. The instructor gave the written homework via Microsoft Teams. Students should copy the homework and then paste it into the Copilot on WhatsApp. Learners were informed to use the prompts they learnt during the training session to get further clarification on their answers. Additionally, students were instructed to spend 20 to 30 minutes using Copilot to create practice tasks, such as multiple-choice, fill-in-the-blank, and true/false questions, for each of the 10 terms. They were also encouraged to ask questions directly to obtain answers and explanations. Additionally, the teacher informed students that their conversations would be reviewed using conversation history assessments to ensure they were using Copilot on WhatsApp outside the classroom. This process stayed the same from Week 1 until Week 4.

The control group learnt the vocabulary through the traditional way, without any help from AI. The classes lasted 60 minutes and followed a set pattern: 25 minutes of direct instruction, which included vocabulary clarification with examples on the whiteboard and in students' textbooks; 20 minutes of monitored practice using gap-fill and transformation activities from the directed textbook; and 15 minutes of semi-controlled exercises using pair work and short writing tasks. Students were given homework tasks and informed that the homework would be discussed in the next class. This strategy employed the Presentation-Practice-Production approach, which involved assigning regular assignments to help students practice more effectively. The researcher gave the vocabulary posttest to all three groups in Week 6. At the end of Week 7, a delayed posttest was administered to assess how well the students retained the terminology.

## 4. Results

This section covers all investigations and statistics derived from the collected numerical data, aiming to answer the research question of measuring the impact of using Copilot AI as a learning facilitator in the vocabulary learning process. Before conducting a thorough analysis of the data, it was essential to assess the data's normality to remove outliers and select the appropriate parametric

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

or nonparametric tests. Therefore, the Shapiro-Wilk test of normality was performed on all sets of tests, and the results are presented in Table 2.

**Table 2.** The Results of the Normality Test in All Sets of Tests For the Three Groups

| | groups | Shapiro-Wilk | | |
|---|---|---|---|---|
| | | Statistic | df | Sig. |
| pretest | control | .898 | 30 | .007 |
| | experimental A | .911 | 30 | .015 |
| | experimental B | .918 | 30 | .024 |
| posttest | control | .833 | 30 | .000 |
| | experimental A | .903 | 30 | .010 |
| | experimental B | .883 | 30 | .003 |
| Delayed posttest | control | .830 | 30 | .000 |
| | experimental A | .909 | 30 | .014 |
| | experimental B | .925 | 30 | .037 |

Table 2 shows that the data in all three sets do not follow a normal distribution. All the significant values are less than 0.05. The control group, experimental group A, and experimental group B all had significant values of 0.007, 0.015, and 0.024, respectively, for the pretest. The numbers in the posttest are 0.000, 0.010, and 0.003; in the delayed posttest, they are 0.000, 0.014, and 0.037. These results confirm that the scores in each group did not follow a normal distribution at any point during the evaluation. Thus, to compare students' performance, a nonparametric test is more suitable. Initially, students' performance within their groups was analyzed using a Wilcoxon Signed-Rank Test, and the results are presented in Table 3.

**Table 3.** The Within-Groups Analysis of Vocabulary Knowledge in All Sets of Tests

| | | post- pre | Delayed post-pretest | Delayed post - posttest |
|---|---|---|---|---|
| Control | Z | -4.731 | -3.519 | -4.123 |
| | Asymp. Sig. (2-tailed) | .000 | .000 | .000 |
| Experiment A | Z | -4.821 | -4.805 | -3.502 |
| | Asymp. Sig. (2-tailed) | .000 | .000 | .000 |
| Experiment B | Z | -4.812 | -4.814 | -1.834 |
| | Asymp. Sig. (2-tailed) | .000 | .000 | .000 |

Table 3 shows significant differences within each group over time. The control group had a big jump in scores from pretest to posttest ($Z = -4.731$, $p < .001$) and from pretest to delayed posttest ($Z = -3.519$, $p < .001$). However, there was a significant drop from posttest to delayed posttest ($Z = -4.123$, $p < .001$), which means they did not remember the information well. All the comparisons in

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

experimental group A were statistically significant: posttest vs. pretest (Z = -4.821), delayed posttest vs. pretest (Z = -4.805), and delayed posttest vs. posttest (Z = -3.502), all with p <.001. This means that the learning progress is strong and long-lasting. In experimental group B, the differences between the posttest and the pretest (Z = -4.812) and the delayed posttest and the pretest (Z = -4.814) were both statistically significant (p < .001), which means that the students learned and remembered well. The difference between the delayed posttest and the posttest showed a more minor reduction (Z = -1.834, p < .001), indicating that group B retained most of their gains with minimal loss. All the groups made progress, but Group A in the experimental group made the most noticeable and long-lasting gains in learning.

To compare students' performance, a Kruskal-Wallis test was conducted, and the results are presented in Table 4.

**Table 4.** The Results of Comparing Students' Performance in All Sets of Tests in All Groups

| | pretest | posttest | Delayed posttest |
|---|---|---|---|
| Kruskal-Wallis H | .342 | 64.456 | 71.126 |
| df | 2 | 2 | 2 |
| Asymp. Sig. | .843 | .000 | .000 |

Table 4 shows that there was no significant difference between the three groups in the pretest (H = 0.342, p = .843), indicating that all groups had approximately the same amount of vocabulary knowledge to start with. There were considerable differences in the posttest (H = 64.456, p < .001) and the delayed posttest (H = 71.126, p < .001). This means that the instructional interventions had different effects on how well students learned and remembered language. The statistics indicate that the type of treatment administered to each group affected their results following the intervention. To better understand this difference, a multiple-comparison post hoc test was conducted, and the results are presented in Table 5.

**Table 5.** The Results of Multiple Comparisons Among the Three Groups in All Tests

| tests | (I) groups | (J) groups | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| pretest | control | experimental A | -.0333 | .26135 | 1.000 | -.6713 | .6046 |
| | | experimental B | -.1000 | .26135 | 1.000 | -.7380 | .5380 |
| | experimental A | control | .0333 | .26135 | 1.000 | -.6046 | .6713 |
| | | experimental B | -.0667 | .26135 | 1.000 | -.7046 | .5713 |
| | experimental B | control | .1000 | .26135 | 1.000 | -.5380 | .7380 |
| | | experimental A | .0667 | .26135 | 1.000 | -.5713 | .7046 |
| posttest | control | experimental A | -3.9333 | .27571 | .000 | -4.6064 | -3.2603 |
| | | experimental B | -2.3333 | .27571 | .000 | -3.0064 | -1.6603 |

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

| tests | (I) groups | (J) groups | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| delayed posttest | experimental A | control | 3.9333 | .27571 | .000 | 3.2603 | 4.6064 |
| | | experimental B | 1.6000* | .27571 | .000 | .9270 | 2.2730 |
| | experimental B | control | 2.3333 | .27571 | .000 | 1.6603 | 3.0064 |
| | | experimental A | -1.6000 | .27571 | .000 | -2.2730 | -.9270 |
| | control | experimental A | -6.9333 | .31707 | .000 | -7.7074 | -6.1593 |
| | | experimental B | -4.8000 | .31707 | .000 | -5.5740 | -4.0260 |
| | experimental A | control | 6.9333 | .31707 | .000 | 6.1593 | 7.7074 |
| | | experimental B | 2.1333 | .31707 | .000 | 1.3593 | 2.9074 |
| | experimental B | control | 4.8000 | .31707 | .000 | 4.0260 | 5.5740 |
| | | experimental A | -2.1333 | .31707 | .000 | -2.9074 | -1.3593 |

Table 5 shows that all group comparisons in the pretest had p-values of 1.000, indicating no differences in vocabulary knowledge at the start. In the posttest, all pairwise comparisons were statistically significant ($p < .001$). Experimental A was significantly better than both the control group (mean difference = 3.93) and experimental B (mean difference = 1.60). Experimental B was also better than the control group (mean difference = 2.33). The same trend continued in the delayed posttest: experimental A consistently performed better than experimental B (mean difference = 2.13) and the control group (mean difference = 6.93), while experimental B still had a significant advantage over the control group (mean difference = 4.80). The results show that experimental A consistently had larger learning gains and better retention over time than experimental B, but the control group did not perform as well.

## 5. Discussion

The only objective of this study was to measure the impact of using Copilot as a learning facilitator through WhatsApp on the vocabulary knowledge of Omani EFL learners with pre-intermediate English proficiency. Three groups of students were divided equally into a control group, experimental A (using Copilot within the classroom), and experimental B (using Copilot outside the classroom). The study's findings revealed that experimental group A performed significantly better than the other two groups in both the posttest and the delayed posttest, followed by experimental group B. Additionally, the within-groups analysis of learners` pretests, posttests, and delayed posttests revealed the following results. In the control group, due to the in-class training, students' scores increased from the pretest to the posttest; however, there was a sharp decrease in the delayed posttest, indicating that the retention of vocabulary over time was not adequately supported by in-class instructions alone. In experimental A, the scores increased from the pretest to the posttest and continued to increase to the delayed posttest, indicating a positive and significant impact of in-class training, combined with the use of Copilot, on vocabulary learning. Finally, it was found that experimental group B also showed signs of progress from the pretest to the delayed posttest; however, their scores decreased nonsignificantly by just a few points on the delayed posttest, which

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

could be attributed to the use of Copilot outside the educational setting and in-class training. Therefore, it can be stated that Copilot, as an AI tool, played a significant role in improving students' vocabulary knowledge temporarily and, in some cases, in the long run.

Within experimental group A, using Copilot in an organized classroom setting with teacher supervision and peer participation, both short-term vocabulary learning and long-term retention were significantly enhanced. During their work, the good practices they learned from one another were further enhanced through conversations among colleagues who shared common qualities. The achievements of this group can be attributed to the Dual Coding Theory (Paivio, 1990), as learners benefited from the verbal and visual presentation of vocabulary via Copilot, leading to improved memory encoding. In addition, the findings support the Sociocultural Theory (Vygotsky, 1978), with an emphasis on the Zone of Proximal Development, which states that learners progress significantly through scaffolded learning interactions with their peers and teachers using this AI tool in a collaborative learning setting.

The experimental group, which used Copilot outside the classroom, also showed progress that could be attributed to self-regulated use of Copilot as a vocabulary learning tool. However, the lack of scaffolded learning in the class limited their long-term improvements in vocabulary retention. The results could be linked to the Constructivist Learning Theory, which posits that students can actively construct knowledge by exploring and interacting with learning tools without time constraints. Additionally, the Self-Determination Theory (Deci & Ryan, 1985) is particularly applicable to this group's findings, especially in the autonomy element, where students are highly motivated to control their learning process. However, a lack of structured support may reduce the level of relatedness and competence essential for achieving optimal long-term outcomes.

The control group, which received only in-class traditional training, showed mild progress from the pretest to the posttest, but not in the delayed posttest. It shows that learning vocabulary demands teaching techniques that focus more on repeated or varied interactions in the target language. Nation (2001) stated that vocabulary knowledge needs to be repeated in a meaningful context to remain in the mind for an extended period. This is also supported by the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), which states that engaging, meaningful tasks are essential for vocabulary retention, which was absent in the case of the control group.

The findings of this study align with Romadhon's (2025) research, which examined the effectiveness of the Replika AI chatbot in helping Indonesian polytechnic students learn new words. The study demonstrated that incorporating chatbots into the classroom significantly enhanced students' ability to learn new vocabulary. Abril et al. (2025) also examined how machine learning can be used as a teaching method to help first-year secondary school students in Lago Agrio enhance their English vocabulary. The study's results showed that machine learning may significantly enhance vocabulary learning. The results align with the study by Oktadela et al. (2023), which examined the effectiveness of AI chatbots in helping students strengthen their English vocabulary during community service training sessions. The study found that using AI chatbots might help people learn new words. Qasem et al. (2023) examined the impact of the Dialogflow chatbot on vocabulary learning in online classrooms. They found that using chatbots in English for Specific Purposes (ESP) vocabulary learning had several positive benefits. Hutauruk et al. (2024) investigated the effectiveness of artificial intelligence avatars in helping college students learn new words. The study found that AI avatars help students learn and remember new words, which boosts their confidence and interest in studying English.

## 6. Conclusion

This paper investigates the impact of using Copilot as an additional learning tool to enhance the vocabulary knowledge of Omani pre-intermediate EFL learners. The study's findings revealed the

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

progress of all groups from pretest to posttest. Experimental Group A, which used Copilot in the class setting, demonstrated significantly better performance than the other two groups. Experimental Group B, which used Copilot outside of classes, achieved second place in terms of best performance in both the posttest and the delayed posttest. These findings suggest that using AI tools, such as Copilot, can significantly enhance students' vocabulary knowledge. When used consistently and educationally, it can also aid long-term vocabulary retention. While the control group showed significant progress on the posttest, participants' scores decreased on the delayed posttest, confirming that limited technology use may reduce vocabulary retention among students.

The findings of this study are helpful for teachers, students, and e-learning coordinators in institutions. The results provide teachers with clear guidance on how to utilize AI-driven tools, such as Copilot, to enhance vocabulary instruction in both in-person and online settings. These technologies can effectively enhance traditional teaching methods by providing students with personalized support, timely feedback, and increased engagement opportunities. Copilot's controlled yet flexible features, together with the popularity of platforms like WhatsApp, provide students with the option to practice vocabulary in real-life situations, review on their own, and regularly learn new aspects of the language. This helps students learn right away and remember words for a long time, primarily when they are organized in a classroom setting. The results indicate that affordable and scalable AI systems may help e-learning coordinators improve language learning outcomes. The study demonstrates that Copilot can be easily integrated into existing digital systems without necessitating significant modifications to the underlying technology. Additionally, it works well in both formal (in-class) and informal (out-of-class) settings, making it easy to build hybrid or fully remote language systems. As schools explore new approaches to teaching following the pandemic, the findings of this study provide a strong case for the strategic use of AI-driven technologies in language education policy and practice.

## 7. Limitations

This study offers a wealth of helpful information on how AI-assisted technology, such as Copilot, can aid individuals in acquiring language, but it also has certain limitations. Due to the small sample size, the results may not be applicable to other groups of Omani EFL learners in different institutions with varying English proficiency levels. The second limitation is that the intervention may not have been sufficiently lengthy to fully measure the long-term effects of Copilot on language retention. Another limitation is that no comparisons were made with other AI technologies, which would have provided a more comprehensive picture of how practical the lessons were. Another limitation is that it solely uses quantitative data, which makes it challenging to understand the user experience and the factors that affect it, as it doesn't include qualitative data such as learner feedback, motivation, or engagement. Finally, since there was no control over the age and gender of the students, it could be considered another limitation of the study.

## 8. Suggestions

Based on the results and scope of the current study, several avenues for further research could be explored. In the future, researchers may examine how AI-assisted vocabulary learning systems, such as Copilot, impact a broader range of English proficiency levels, including both beginners and advanced learners, to determine if they offer similar benefits. Additionally, examining how Copilot impacts other language skills, such as listening, speaking, and reading, would provide a clearer understanding of its benefits in helping people learn English. Researchers may investigate how using AI technology in class versus outside of class affects students in different types of schools, such as rural versus urban schools and public versus private schools. Additionally, comparing Copilot to other AI-powered platforms, such as ChatGPT, Grammarly, or Quizlet, can help us determine which technologies work best for specific language learning goals. Using qualitative methods, such as

https://doi.org/10.22521/edupij.2025.20.12 Published online by Universitepark Press

student interviews or focus groups, would help us better understand how students perceive AI in education, its perceived usefulness, and what motivates them to use it. Long-term research is necessary to investigate the long-term impact of AI technologies on vocabulary retention and overall language skills. Moreover, experimental designs that combine AI with blended or flipped classroom models may lead to innovative approaches to teaching that enable students to learn more independently and achieve better outcomes in language learning contexts that utilize technology. And finally, further studies could be conducted in a more controlled environment to measure the effects of age and gender on AI literacy.

## Declarations

**Author Contributions.** All authors have read and approved the final version of the article.
**Conflicts of Interest.** The authors declare no conflict of interest.
**Funding.** Not Applicable
**Ethical Approval.** Prior to the implementation of the study, approval was obtained from the university's authorities.
**Data Availability Statement.** The data could be shared upon a reasonable request.

## References

Abril, B. V. D., Carlosama, K. M. D., & Pachay, J. F. Z. (2025). Exploring the effect of ChatGPT as an educational tool to improve vocabulary acquisition in the English Language. *Latin Science Multidisciplinary Scientific Journal*, *9*(1), 4874–4886. https://doi.org/10.37811/cl_rcm.v9i1.16190

Al Ghaithi, A., Behforouz, B., & Isyaku, H. (2024). The effect of using a WhatsApp bot on English vocabulary learning. *Turkish Online Journal of Distance Education*, *25*(2), 208–227. https://doi.org/10.17718/tojde.1297285

Altyari, A. W. (2017). English vocabulary uptake among Saudi Arabic-speaking students in public schools. *British Journal of English Linguistics*, *5*(1), 10–16.

Behforouz, B., & Ghaithi, A. A. (2024). Investigating the effect of an interactive educational chatbot on reading comprehension skills. *International Journal of Engineering Pedagogy (iJEP)*, *14*(4), Article 4. https://doi.org/10.3991/ijep.v14i4.48461

Chen, H.-L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, *58*(6), 1161–1189. https://doi.org/10.1177/0735633120929622

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161–182. https://doi.org/10.1007/s10796-022-10291-4

Chen, Z., Jia, J., & Li, W. (2021). Learning curriculum vocabulary through mobile learning: Impact on vocabulary gains and automaticity. *International Journal of Mobile Learning and Organisation*, *15*(2), 149–163. https://doi.org/10.1504/IJMLO.2021.114518

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An education support system for students. *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29-31, 2018, Proceedings 10*, 291–302. https://doi.org/10.1007/978-3-030-01689-0_23

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer. https://doi.org/10.1007/978-1-4899-2271-7

Dekker, I., De Jong, E. M., Schippers, M. C., De Bruijn-Smolders, M., Alexiou, A., & Giesbers, B. (2020). Optimizing students' mental health and academic performance: AI-enhanced life crafting. *Frontiers in Psychology*, *11, 1063.* https://doi.org/10.3389/fpsyg.2020.01063

Durall, E., & Kapros, E. (2020). Co-design for a competency Self-assessment chatbot and survey in science education. *Learning and Collaboration Technologies. Human and Technology Ecosystems*, 13–24.https://doi.org/10.1007/978-3-030-50506-6_2

Dwivedi, Y. k., Kshetri, N., Hughes, L., Slade, E. louise, Jeyaraj, A., Kar, A. kumar, Baabdullah, A. m., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. ahmad, Al-Busaidi, A. s., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management*, *71*. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning and Technology*, *10*(3), 8–14. https://doi.org/10.64152/10125/44068

Gonda, D. E., Luo, J., Wong, Y. L., & Lei, C. U. (2018). Evaluation of developing educational chatbots based on the seven principles for good teaching. *IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 446–453. https://doi.org/10.1109/TALE.2018.8615175

Gu, P. Y. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *Tesl-Ej*, *7*(2), 1–25.

Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers and Education*. Advanced Online Publication. https://doi.org/10.1016/j.compedu.2021.104395

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

Haristiani, N., Danuwijaya, A. A., Mumu Muhammad, R., & Sarila, H. (2019). Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology, 14*(6), 3158–3173.

Healey, J. (2020). *Artificial Intelligence*. The Spinney Press.

Hutauruk, B. S., Purba, R., Sihombing, S., & Nainggolan, M. (2024). The effectiveness of artificial intelligence by chatbot in enhancing the students' vocabulary. *JETAL: Journal of English Teaching & Applied Linguistic*, *6*(1), 13–19. https://doi.org/10.36655/jetal.v6i1.1610

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jia, F., Sun, D., Ma, Q., & Looi, C. K. (2022). Developing an AI-based learning system for L2 learners' authentic and ubiquitous learning in English language. *Sustainability*, *14*(23), 15527. https://doi.org/10.3390/su142315527

Khalil, M., Amreen, S., & Ather, F. (2025). Exploring the role of AI chatbots in enhancing English grammar and vocabulary acquisition: a study of learning dynamics, user engagement, and linguistic development. *Journal of Applied Linguistics and TESOL (JALT)*, *8*(1), 801–812.

Khan, I. A. (2011). Learning difficulties in English: Diagnosis and pedagogy in Saudi Arabia. *Educational Research*, *2*(7), 1248–1257.

Kim, N. Y. (2020). *Chatbots and language learning: Effects of the use of AI chatbots for EFL learning*. Chisinau, Moldova: Eliva Press.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Liu, Q., Huang, J., Wu, L., Zhu, K., & Ba, S. (2019). CBET: design and evaluation of a domain-specific chatbot for mobile learning. *Universal Access in the Information Society, 19,* 655-673. https://doi.org/10.1007/s10209-019-00666-x

Mendoza, S., Hernández-León, M., Sánchez-Adame, L. M., Rodríguez, J., Decouchant, D., & Meneses-Viveros, A. (2020). Supporting student-teacher interaction through a chatbot. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Human and Technology Ecosystems. HCII 2020. Lecture Notes in Computer Science* (pp. 93-107), vol. 12206. Springer, Cham. https://doi.org/10.1007/978-3-030-50506-6_8

Mohd Amin, M. R., Ismail, I., & Sivakumaran, V. M. (2025). Revolutionizing education with artificial intelligence (AI)? challenges, and implications for open and distance learning (ODL). *Social Sciences & Humanities Open*, *11*, 101308. https://doi.org/10.1016/j.ssaho.2025.101308

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge university press Cambridge. https://doi.org/10.1017/cbo9781139524759

Oke, A., & Fernandes, F. A. P. (2020). Innovations in teaching and learning: Exploring the perceptions of the education sector on the 4th industrial revolution (4IR). *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(2), 31. https://doi.org/10.3390/joitmc6020031

Oktadela, R., Elida, Y., & Ismail, S. (2023). Improving English vocabulary through artificial intelligence (AI) chatbot application. *Journal of English Language and Education*, *8*(2), Article 2. https://doi.org/10.31004/jele.v8i2.308

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Perez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. https://doi.org/10.1002/cae.22326

Qasem, F., Ghaleb, M., Mahdi, H. S., Al Khateeb, A., & Al Fadda, H. (2023). Dialog chatbot as an interactive online tool in enhancing ESP vocabulary learning. *Saudi Journal of Language Studies*, *3*(2), 76–86. https://doi.org/10.1108/SJLS-10-2022-0072

Röhrig, C., & Heß, D. (2019). OmniMan: A Mobile Assistive Robot for Intralogistics Applications. *Engineering Letters*, *27*(4), 1-8.

Romadhon, R. (2025). Replika AI chatbot as a tool for enhancing ESP business vocabulary acquisition: A study on polytechnic students. *SALEE: Study of Applied Linguistics and English Education*, *6*(1), 183–201. https://doi.org/10.35961/salee.v6i1.1671

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmulian, A., & Coetzee, S. A. (2019). The development of Messenger bots for teaching and learning, and the accounting students' experience of using them. *British Journal of Educational Technology*, *50*(5), 2751-2777. https://doi.org/10.1111/bjet.12723

Shadiev, R., Wu, T.-T., & Huang, Y.-M. (2020). Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. *ReCALL*, *32*(2), 195–212. https://doi.org/10.1017/S0958344020000038

Silitonga, L. M., Wiyaka, W., & Prastikawati, E. F. (2024). Boosting students' ESP vocabulary by utilizing ai chatbot. *ETERNAL (English Teaching Journal)*, *15*(2), 275–283. https://doi.org/10.26877/eternal.v15i2.605

Sinha, S., Basak, S., Dey, Y., & Mondal, A. (2020). An Educational chatbot for answering queries. *Emerging Technology in Modelling and Graphics Proceedings of IEM Graph 2018.* https://doi.org/10.1007/978-981-13-7403-6_7

Susanto, A. (2016). How English learner succeeded in difficult circumstances. *UNSIKA Education Journal*, *4*(2). https://journal.unsika.ac.id/judika/article/view/383

Susanto, A. (2017). The teaching of vocabulary: A perspective. *Word Journal*, *1*(2), 183. https://doi.org/10.22216/jk.v1i2.2136

Tellier, M. (2008). The effect of gestures on second language memorization by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Tsai, Y.-L., & Tsai, C.-C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education*, *125*, 345–357. https://doi.org/10.1016/j.compedu.2018.06.020

Tu, Y., Zou, D., & Zhang, R. (2020). A comprehensive framework for designing and evaluating vocabulary learning apps from multiple perspectives. *International Journal of Mobile Learning and Organisation*, *14*(3), 370. https://doi.org/10.1504/IJMLO.2020.108199

Verleger, M., & Pembridge, J. (2019). A pilot study integrating an AI-driven chatbot in an introductory programming course. *IEEE Frontiers in Education Conference (FIE)*, 1-4. https://doi.org/10.1109/FIE.2018.8659282

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). *Harvard University Press.* https://doi.org/10.2307/j.ctvjf9vz4

Waziana, W., Andewi, W., Hastomo, T., & Hasbi, M. (2024). Students' perceptions about the impact of AI chatbots on their vocabulary and grammar in EFL writing. *Register Journal*, *17*(2), 352–382. http://dx.doi.org/10.18326/rgt.v17i2.352-382

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 327–343. https://doi.org/10.1017/S0958344022000039

Yanguas, I. (2010). Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology, 14*(3), 72-93. https://doi.org/10.64152/10125/44227

Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research, 59*(1), 154-177. https://doi.org/10.1177/0735633120952067

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

## About the Contributor(s)

**Miew Luan Ng,** serves as Associate Professor at the Faculty of Education and Liberal Arts, INTI International University, and held the position as the Deputy Director of Center of Sustainable Business Innovation and Corporate Responsibility. She is also the Assistant General Secretary of the Malaysian Social Sciences Association. Her research spans newspaper discourse analysis, media education, political communication, language and power, and issues affecting minority and disabled communities. Affiliation(s): INTI International University, Malaysia[1]; Shinawatra University, Thailand[2].
Email: miewluan.ng@newinti.edu.my
ORCID: https://orcid.org/0000-0003-0949-5858

**Ali Al Ghaithi,** is an English Lecturer at Foundation Department of Sohar University in Oman. Currently, he is a Ph.D. candidate focusing on Applied Linguistics. Ali got his master's degree from University of Putra Malaysia. Ali is interested in research studies that mainly implement Artificial Intelligence in teaching and learning processes. Affiliation(s): Sohar University, Oman.
Email: aghaithi@su.edu.om
ORCID: https://orcid.org/0000-0002-9653-508X

**Behnam Behforouz,** is an English Lecturer in the Preparatory Studies Center at University of Technology and Applied Sciences, Shinas, Oman. His main areas of interest are TESOL, Applied Linguistics, Language Education, and Educational Technologies. Affiliation(s): University of Technoloy and Applied Sciences, Shinas, Oman[1]; Sohar University, Oman[2].
Email: behnam.behforouz@utas.edu.om
ORCID: https://orcid.org/0000-0002-0078-2757

***Publisher's Note:*** *The opinions, statements, and data presented in all publications are solely those of the individual author(s) and contributors and do not reflect the views of Universitepark, EDUPIJ, and/or the editor(s). Universitepark, the Journal, and/or the editor(s) accept no responsibility for any harm or damage to persons or property arising from the use of ideas, methods, instructions, or products mentioned in the content.*