Article

Journal of International Students
Volume 16, Issue 11 (2026), pp. 339-356
ISSN: 2162-3104 (Print), 2166-3750 (Online)
jistudents.org
DOI: https://doi.org/10.32674/za5eta42

Empowering Grammar Acquisition through Copilot AI in Higher Education

Ali Al Ghaithi
Sohar University, Oman

Behnam Behforouz
University of Technology and Applied Sciences, Shinas, Oman

Salah Al-Hanshi
Sultan Qaboos University, Oman

Corresponding author: Ali Al Ghaithi, Faculty of Language Studies, Sohar University, Oman. Orcid: https://orcid.org/0000-0002-9653-508X

ABSTRACT: *This research aimed to thoroughly investigate the impact of using Copilot AI within and outside the classroom on the language learning process and to compare the results with those of traditional in-class training without technological access. To collect the required data, 99 college students with intermediate proficiency levels were divided into a control group with traditional face-to-face training, an experiment 1 group with in-class training and using Copilot AI for grammar training, and an experiment 2 group with in-class training and using Copilot AI outside of the class, with the same number of students in each class. Researcher developed tests were designed to compare scores across groups. The findings initially revealed that all groups progressed in the posttest; however, the group in experiment 1, which used Copilot AI in class, performed significantly better than the two other groups did, followed by the group in experiment 2, and the control group. However, the delayed posttest showed that the experimental group 1 outperformed the other two groups, indicating greater potential to remember grammatical structures. In contrast, the other two groups showed no improvement in the delayed posttest. The results of this study are beneficial for teachers, students, and institutions.*

Keywords: Artificial Intelligence, Copilot AI, Grammar, Higher Education

Received: Sept 26, 2025 | **Revised**: April 18, 2026 | **Accepted:** May 10, 2026

How to Cite (APA): Al Ghiathi, A., Behforouz, B., Al-Hanshi, S. (2026). Epowering grammar acquisition through Copilot AI in higher education. *Journal of International Students*, *16*(11), 339-356. https://doi.org/10.32674/za5eta42

INTRODUCTION

Grammar, which varies in structure across languages, is essential for language acquisition (Normawati & Nugrahaeni, 2024). Widodo (2006) underscores that other linguistic abilities, including listening, speaking, reading, and writing, are dependent on grammatical proficiency, emphasizing its essential role in language learning. In addition, Ellis (2006) emphasizes the importance of teaching grammar to enhance language learning, as grammar is a crucial component of effective communication.

EFL learners face several challenges in comprehending grammar. One significant problem is that students lack the basic structural rules they should have learned at the beginning of their education, which makes it difficult for them to continue studying (Nawaz et al., 2022). Additionally, many students see grammar lessons as laborious and tedious, which negatively affects their motivation and engagement (Muhsin, 2015). This is because languages have different structures, making it more difficult for students to learn English since their home languages typically do not follow English syntax (Takala, 2016). Yule (2010) contends that several students recall grammatical rules without developing a comprehensive understanding, which leads to their inability to apply these rules proficiently in writing and speaking. Agregado (2023) reported that grammar instruction faces significant resource constraints. For example, insufficient teaching and learning materials make it very challenging to teach grammar effectively, as teachers often lack the resources to meet the diverse needs of all students in modern classrooms. Kaçani and Mangelli (2013) reported that grammar instruction is more challenging because teachers have to address students' spoken communication issues more frequently than their written communication issues. A common problem is that languages have different structures.

To overcome such challenges in teaching grammar, several studies have demonstrated that AI aids students more effectively (Bannister et al., 2024; Katsarou et al., 2025; Lalira et al., 2024). Stevens (2025) stated that AL promotes the learning of grammar by automatically correcting related mistakes, assessing students in real time, and explicating grammar rules in a contextualized way. Furthermore, AI-supported grammar learning was significantly better than traditional methods of grammar skill development were. Elmotri et al. (2025) reported that the use of AI-based automated written corrective feedback can improve the accuracy of grammar and sentence structures. Sharmithashini and Hashim (2025) reported that AI grammar checkers are more effective at achieving grammatical correctness through real-time corrective feedback. Setiawan and

Alkhowarizmi (2025) reported that combining AI feedback with teacher and peer feedback yields the best possible writing outcomes.

Among these AI-enhanced learning advancements, chatbots are well known applications that have seen significant popularity, particularly for providing personalized, interactive learning experiences that enable students to interact with peers in real time through text or conversation, offering immediate responses tailored to their individual needs (Winkler & Soellner, 2018). Studies have shown that chatbots improve learners' performance across various language skills by providing opportunities for practice and genuine interaction. They develop reading comprehension and writing accuracy by providing instant feedback, modeling dialogs, and offering a variety of interactive exercises (Ittefaq et al., 2025; Pamilu et al., 2021; Safitri et al., 2021). Additionally, several studies have reported that chatbots significantly improve students' grammatical ability (Ahn, 2022; Al Ghaithi & Behforouz, 2024; Kharis et al., 2022).

Despite these benefits, studies on the potential of AI-based chatbots to improve the grammatical competence of EFL learners are rare. Although such tools have been verified to result in better speaking, listening, vocabulary usage, and reading skills (Pamilu et al., 2021; Safitri et al., 2021) as well as in grammar comprehension success (Kim, 2019), the current literature lacks work that focuses mainly on its application for learning grammar. Additionally, Copilot AI has been largely unexplored and understudied when combined with WhatsApp, a mobile-based platform for enhancing grammar learning. This research aims to investigate the effect of Copilot AI on the grammatical proficiency of Omani EFL students through WhatsApp, thereby enriching the understanding of AI-driven mobile-assisted language learning in grammar education. Therefore, the following research question is addressed in this study:

> Does Copilot AI improve the grammatical knowledge of Omani intermediate EFL learners?

LITERATURE REVIEW

The Impact of Artificial Intelligence on Grammar Learning

Hadi et al. (2026) investigated the impact of adaptive AI-based learning on the mastery of English grammar structure among 60 undergraduate EFL learners in Indonesia. There were two groups involved: the experimental group, which benefited from learning through adaptive AI platforms, and the control group, which was taught using traditional methods. Pretests, posttests, questionnaires, and interviews were used as data collection tools. The results revealed that the experimental group performed significantly better than the control group did. The analysis of multivariate techniques is a major strength; however, the study's sample size is a crucial limitation, as it impedes generalizability.

Chen (2026) studied the effects of an AI-embedded, interactive, self-regulated system on the grammar, vocabulary, fluency, and pronunciation of 120 EFL learners. The research involved two groups: a 60-member experimental group that interacted with the AI-embedded system and a 60-member control group that

received only conventional instruction for 6 weeks. The research tools included pre- and posttests and a user survey, which served as the data collection instruments. The experimental group that learned through the AI-embedded software significantly surpassed the control group in terms of skills, including grammar, which was 17.4 points greater for the AI group than for the control group (6.6 points greater). The study's major asset is its wide range of participants, while its main limitation is that it was conducted at only one institution.

Focusing specifically on grammar learning, Ali et al. (2025) examined the impact of artificial intelligence on English grammatical competency among Saudi ESL learners. One hundred participants were divided into two groups of 50 students each: a control group that received traditional education and an experimental group that used ChatGPT alongside standard teaching techniques. The study revealed that the treatment group's grammar scores improved significantly compared with those of the control group, and the effect size was also substantial, indicating that ChatGPT is a valuable supplementary resource for learning grammar. The strengths include careful experimental design, appropriate control groups, and statistical validation. A weakness of the study is that it examines only male Saudi students.

Ng et al. (2025) explored the application of Gemini AI in facilitating the grammar and engagement of 75 Omani preintermediate EFL students. The study was conducted with three groups of 25; the first was the control group taught in a traditional way, and the other two were Experimental A, which used Gemini in class, and Experimental B, which used Gemini outside of class for 7 weeks. The data collection tools included researcher-made grammar tests and an engagement questionnaire. Experimental A was the group that significantly outperformed both groups in terms of grammar and engagement according to the findings. On the one hand, the delayed posttest design is a strong point of the study; on the other hand, the small sample size and single context diminish the generalizability of the results.

Expanding this study to a broader scope, Stevens (2025) conducted a larger-scale examination of how AI can be used to teach English grammar. This was accomplished through a mixed-methods study involving 3,676 students. The students were divided into two groups: the experimental group (1,870 students) utilized AI tools such as Grammarly and ChatGPT, while the control group (1,806 students) received regular instruction. The findings showed that the experimental group performed better in terms of grammar than the control group did. A large sample size and a rigorous mixed-methods approach are two key strengths of the study. However, the poor state of infrastructure, including unreliable internet, insufficient devices, and power outages, was a significant shortcoming.

Using a different methodological approach, Mohammed and Ja'ashan (2024) looked at AI-assisted contextual dialogs for learning English grammar in Saudi Arabia. A group of 110 students used the Talkpal AI application for 6 weeks of contextualized grammar practice. The results showed that students' grammar improved significantly from the pretest to the posttest, with a 16.8% increase. The questionnaire responses indicated that students were generally happier with AI

than with traditional methods across all the measures. The strengths include a well-thought-out mixed-methods design and usefulness in the real world. The limitations of this study include its small sample size and the absence of an examination of long-term retention.

Expanding the focus beyond grammar alone, Chen (2025) investigated the effectiveness of an AI-integrated interactive English learning system in enhancing language proficiency, thereby extending the emphasis beyond grammar. Sixty participants were randomly divided into two groups: an experimental group that used the AI-integrated system and a control group that received traditional instruction. The AI system uses speech recognition, natural language processing, and machine learning to deliver personalized content recommendations, instant feedback, and conversation practices. The experimental group had much greater improvements across all skill domains: grammar, vocabulary, fluency, and pronunciation. This study features a strong experimental design and a comprehensive skill assessment; however, the small sample size and short intervention period are among its shortcomings.

To address the vital problem of retention, Behforouz and Al Ghaithi (2024) investigated the effectiveness of artificial intelligence in enhancing grammatical competence among EFL learners, with a particular focus on long-term learning outcomes. Sixty Omani intermediate EFL learners were randomly divided into three groups of twenty students each: a control group that received traditional face-to-face instruction, experimental group 1 that received in-class instruction supplemented by ChatGPT feedback, and experimental group 2 that engaged in entirely online instruction with ChatGPT as the primary feedback provider. The findings showed that experimental group 1 performed significantly better on both the posttest and the delayed posttest. This finding indicates that their grammatical knowledge improved and that they stayed with them for a long time. The strengths include a rigorous experimental design with a three-group comparison and a delayed testing phase for retention assessment; nevertheless, the shortcomings are limited to the investigation of just two grammatical structures.

Methodology

This study is designed on the basis of a quasiexperimental approach. This section provides additional information about the study population, the instruments used to collect data, the procedures for implementing the treatment, and the data analysis, which relies on quantitative statistical analysis using SPSS 27.0.

Participants

To collect the required data, 99 students from one of Oman's higher education institutions were selected. On the basis of the university’s placement test, the students were at an intermediate level of English proficiency. The students' ages ranged from 19 to 21 years, and they spoke Arabic as their native language. These students, a mix of males and females, were studying at the Foundation Department, where they had to complete one-year preparatory courses, such as

IT, Math, and English, before moving on to their specialization in the higher education department.

The learners were divided into three groups, each with 33 students. These groups included a control group that followed in-class traditional face-to-face instruction, Experiment 1, which received in-class training and utilized Copilot AI within the class settings, and Experiment 2, which received in-class training but utilized Copilot AI outside of class. Notably, owing to existing policies, students' use of AI tools was not monitored across all groups. Additionally, the study did not focus on sex as a variable.

Instruments

Grammar Test Package

Three researcher-made tests were used—pretests, posttests, and delayed posttests—to assess the effect of Copilot AI on the English grammatical skills of Omani students. These tests focused on grammatical constructs that include the present simple, present continuous, present perfect, past simple, and future simple. Each assessment consisted of three parts, totaling 30 items and 45 minutes to complete. The first part included a 10-fill-in-the-blank activity, while the second part had 10 questions with multiple-choice options, and the last part included another 10 questions for correcting the grammatical mistakes within the sentences. The maximum possible score was 30 points, with each item being given 1 point. There was no penalty for the wrong answers.

Before the study commenced, these instruments were piloted with 35 Omani EFL students from the same institution, all of whom possessed equivalent proficiency levels. Test-takers were given 45 minutes to complete the test. Pilot test analysis demonstrated a Cronbach's alpha reliability coefficient of 0.887, indicating high instrument reliability.

Procedure

This investigation was implemented during the fall of 2025, the first academic semester, with participants signing a consent form stating that their participation in the study was voluntary. The implementation phase spanned 7 weeks, encompassing a 5-week preassessment phase, a treatment period, a postassessment phase, and a delayed postassessment phase.

Before treatment, all participants completed a standardized grammar pretest focused on the target structures. Both experimental groups subsequently participated in comprehensive training sessions on the use of copula AI for grammar learning and practice.

The intervention followed a structured five-week curriculum addressing fundamental English tense forms: week 1 covered simple tense formation rules, usage contexts, and structural variations; week 2 addressed present continuous tense with progressive formation, temporal markers, and contextual applications; week 3 examined present perfect tense, including auxiliary verb usage, past

participle formation, and experiential contexts; week 4 focused on past simple tense incorporating regular and irregular verb patterns, temporal expressions, and narrative functions; and week 5 concluded with future simple tense covering modal auxiliaries, predictive functions, and planned actions.

In experimental group 1, the classroom treatment for each 60-minute session was delivered in a fixed three-part arrangement throughout the entire week. The first 20 minutes were allotted for the instructor to present the target grammatical structures by employing the PPP (presentation, practice, and production) approach. The next 20 minutes included the students using Copilot AI to obtain more detailed explanations, perform additional exercises, and receive correct, immediate feedback about the specific grammatical point of the lesson in both Arabic and English. They can also generate examples using the tool and check their own understanding of the target form. The last 20 minutes were used for peer dialog and teacher feedback to double-check that the students had learned the grammatical structures properly. The whole week was thus spent teaching all forms of the target structure, including the positive, negative, and interrogative forms. Every week, only one specific tense was taught.

In experimental group 2, the students received the same training, materials, and follow-up activities as in experimental group 1 did, but the major difference was that the students in this group used the Copilot AI outside the classroom as an additional resource. The instructor's presentation of the goal's grammatical structure was delivered during the first twenty minutes. Using examples and brief guided practice, the instructor introduced the rule, its purposes, and its various forms. Collaborative classroom exercises designed to effectively reinforce the target structure were a second twenty-minute filler. Among the activities were sentence-making challenges, error-correction tasks, short conversation competitions, mini whiteboard contests, and a pair-based transformation in which students made affirmative statements to negative and interrogative ones. Games based on sentence structures, peer explanations, group discussions, and instructor feedback took up the last 20 minutes of the consolidation of communication. After each lesson, they used Copilot AI at home to review the grammatical points taught in class, receive further explanations in English and Arabic, generate extra exercises, check their answers, and obtain immediate feedback on errors. The group also used the tool to write positive, negative, and interrogative sentences related to the target tense of the week. All the grammatical structure forms were covered for the week.

As in the other two groups, each session lasted 60 minutes of similar training. For the first 20 minutes, the instructor explains the grammatical structures using the PPP (presentation, practice, production) model. The second 20 minutes were used for traditional classroom practice activities such as fill-in-the-blank exercises, sentence transformation tasks, and, for the last time, questions, drills, and short written sentences on the grammatical point of the session. The final 20 minutes of the timetable were set aside for pair work, peer checking, and whole-class feedback under teacher guidance to ensure that all the students were clear about the target structures. The hard thing on their side was that the students in this group received grammar instruction only through the traditional face-to-face

approach, and there was no copilot AI inside or outside the classroom. Throughout the week, the students learned various grammatical structures, including positive, negative, and interrogative forms.

Following the completion of week 5, all participants undertook the grammar posttest, with week 6 involving the delayed administration of the posttest to evaluate the retention of the targeted tense structures across all the experimental conditions.

Data Analysis

In this section, the data collected from all test sets across all groups are statistically analyzed using SPSS 27.0. First, to assess the distribution of the data in the students' scores, a Kolmogorov–Smirnov test for normality was conducted, and the results are presented in Table 1 below.

Table 1: Results of the Reliability Test among the Groups in all Sets of Tests

		Kolmogorov–Smirnov		
	groups	Statistic	df	Sig.
pretest	control group	.160	33	.031
	experiment 1	.182	33	.007
	experiment 2	.190	33	.004
posttest	control group	.148	33	.065
	experiment 1	.155	33	.043
	experiment 2	.179	33	.009
Delayed posttest	control group	.139	33	.107
	experiment 1	.121	33	.200
	experiment 2	.183	33	.007

Table 1 indicates that, for the pretest, the control group ($p = 0.031$), experiment 1 ($p = 0.007$), and experiment 2 ($p = 0.004$) exhibited significant deviations from normality. In the posttest, the control group ($p = .065$) did not significantly differ from normal. However, both experiment 1 ($p = .043$) and experiment 2 ($p = .009$) showed notable deviations. The delayed posttest indicated that the control group ($p = 0.107$) and Experiment 1 ($p = 0.200$) followed a normal distribution, but experiment 2 (statistic = 1.83, $p = 0.007$) showed a significant deviation from normality. Therefore, for all other statistical analyses, nonparametric tests were selected. To continue the analysis of students' performance within their groups, a Wilcoxon signed-rank test was conducted, and the results are presented in Table 2.

Table 2: Results of the Within-Group Analysis of All Sets of Tests among all the Groups

Groups		Pretest-posttest	Posttest-delayed posttest	Pretest-delayed posttest
Control	Z	-1.966	-.458	-1.158
	Asymp. Sig.	.049	.647	.247
experiment 1	Z	-4.792	-1.195	-4.788
	Asymp. Sig.	.000	.232	.000
experiment 2	Z	-4.908	-4.829	-.838
	Asymp. Sig.	.000	.000	.402

As shown in Table 2, the control group significantly improved from the pretest to the posttest ($Z = –1.966$, $p = .049$), but there were no significant differences between the delayed posttest and the posttest ($Z = –0.458$, $p = .647$) or between the delayed posttest and the pretest ($Z = –1.158$, $p = .247$). In experiment 1, significant increases were noted from the pretest to the posttest ($Z = –4.792$, $p < .001$) and from the pretest to the delayed posttest ($Z = –4.788$, $p < .001$); however, the difference between the delayed posttest and posttest was not statistically significant ($Z = –1.195$, $p = .232$). Experiment 2 demonstrated a significant improvement from the pretest to the posttest ($Z = –4.908$, $p < .001$) and a notable decline from the posttest to the delayed posttest ($Z = –4.829$, $p < .001$). However, the difference between the delayed posttest and the pretest was not statistically significant ($Z = –0.838$, $p = .402$). To compare students' results across groups, a Kruskal–Wallis test was conducted. Table 3 below presents the results.

Table 3: Results of Comparison of all Sets of Tests among the Groups

	pretest	posttest	delayedposttest
Kruskal–Wallis H	1.698	70.994	62.968
df	2	2	2
Asymp. Sig.	.428	.000	.000

As shown in Table 3, there was no significant difference between the groups in the pretest ($p = 0.428$), confirming that the groups were initially equivalent. At the posttest stage, a statistically significant difference was seen among the groups ($p < .001$), indicating that the instructional strategy produced divergent performance outcomes. At the delayed posttest, the difference between groups remained statistically significant ($p < .001$), indicating that the effects of the teaching approaches persisted over time. To measure the size of these differences, the groups were compared using post hoc pairwise comparisons, and the results are presented in Table 4 below.

Table 4: Results of Post Hoc Pairwise Comparison among the Groups in all Sets of Tests

Dependent Variable	(I) groups	(J) groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	Upper Bound
pretest	control	experiment 1	.485	.426	.773	-.553	1.522
		experiment 2	.364	.426	1.000	-.674	1.401
	experiment 1	control group	-.485	.426	.773	-1.522	.553
		experiment 2	-.121	.426	1.000	-1.159	.916
	experiment 2	control group	-.364	.426	1.000	-1.401	.674
		experiment 1	.121	.426	1.000	-.916	1.159
posttest	control	experiment 1	-7.970	.519	.000	-9.235	-6.705
		experiment 2	-3.576	.519	.000	-4.841	-2.311
	experiment 1	control group	7.970	.519	.000	6.705	9.235
		experiment 2	4.394	.519	.000	3.129	5.659
	experiment 2	control group	3.576	.519	.000	2.311	4.841
		experiment 1	-4.394	.519	.000	-5.659	-3.129
Delayed posttest	control	experiment 1	-7.606	.570	.000	-8.994	-6.218
		experiment 2	.333	.570	1.000	-1.054	1.721
	experiment 1	control group	7.606	.570	.000	6.218	8.994
		experiment 2	7.939	.570	.000	6.552	9.327
	experiment 2	control group	-.333	.570	1.000	-1.721	1.054
		experiment 1	-7.939	.570	.000	-9.327	-6.552

Table 4 shows that the three groups were statistically similar in the pretest. The mean score of the control group exceeded that of experiment 1 by 0.485 points ($p = .773$) and that of experiment 2 by 0.364 points ($p = 1.000$). Experiment 1 showed a difference of only 0.121 points from experiment 2 ($p = 1.000$). All confidence intervals include zero, and the updated p values are significant, indicating baseline equality. At the posttest, both treatment groups outperformed the control group, with the in-class condition showing the greatest advantage. Experiment 1 outperforms the control group by 7.970 points ($p < .001$) and experiment 2 by 4.394 points ($p < .001$). Experiment 2 outperformed the control group by 3.576 points ($p < .001$). Experiment 1 had a significant edge during the delayed posttest. Experiment 1 outperforms the control group by 7.606 points ($p < .001$), and Experiment 2 outperforms the control group by 7.939 points ($p < .001$). In experiment 2, on the other hand, no difference from the control group was detected; the difference was 0.333 points ($p = 1.000$). As a result, the improvements seen in the posttest from extracurricular practice were not evident

in the delayed assessment; however, the improvements made in class were significant and long-lasting.

DISCUSSION

This study aimed to assess the impact of using Copilot AI on students' grammatical knowledge. The study's findings revealed that, while all groups showed progress in the posttest, experiment 1 outperformed the other two groups. Additionally, the posttest results revealed that the experiment 2 group performed better than the control group did. In the delayed posttest, compared with the other two groups, experiment 1 showed significant performance; in contrast, experiment 2 and the control group showed no sign of progress and maintained their scores at levels comparable to those of their posttests. These results can be confidently attributed to the use of Copilot AI as a grammar-learning tool, mostly in class, with some positive effects outside the class.

The results of the study can be attributed to how the groups utilized Copilot AI. The experiment 1 group had the best scores because they could use the tool immediately during lectures, which helped them check their grammar, correct mistakes, and reinforce what they had learned. Not only did this regular, planned practice help them perform better on the posttest, but it also ensured that they would remember what they had learned for the delayed posttest. The group that used Copilot AI outside of class performed better on the posttest because the extra practice helped them learn more. However, without a structured classroom setting, their progress did not last in the long term; thus, their scores remained the same. The control group, which relied solely on traditional training, demonstrated improvement from pretest to posttest because of repeated practice and exposure; however, their performance did not progress in the delayed posttest, indicating that while conventional instruction was somewhat effective, it lacked the reinforcement required for long-term retention. These findings collectively demonstrate that Copilot AI AI provided substantial benefits, with the most notable and lasting effects observed when it was integrated directly into classroom instruction.

The findings of this study can be better understood in the context of current educational theories. The group that used Copilot AI in the classroom received the best grades because the tool provided them with instant feedback and support during lessons. This aligns with Vygotsky's concept of the Zone of Proximal Development, which posits that students learn most effectively when they receive assistance at the precise moment they need it (Vygotsky, 1978). The consistent enhancements shown in the delayed posttest further corroborate Swain's Output Hypothesis, as students were encouraged to produce language, recognize deficiencies, and refine their grammar through active engagement (Swain, 1995). The group that used Copilot AI outside of class also performed better on the posttest, indicating that more independent practices helped them retain information more effectively in the short term. However, without guided engagement or collaborative support, their advancements were not retained in the delayed posttest. This pattern aligns with the conclusions of Lai and Gu (2011),

suggesting that technology used without structured guidance may increase engagement but fails to guarantee long-term retention. The control group, which used only traditional methods, progressed from the pretest to the posttest because of repeated exposure and rehearsal, in keeping with the practice effect described in cognitive psychology (Roediger & Karpicke, 2006). Nonetheless, their stalling on the postponed posttest exemplifies Ebbinghaus's forgetting curve, which posits that information declines over time unless it is reinforced by new or recurrent learning events (Ebbinghaus, 1885).

These findings align with those of a study by Phieanchang (2024), which revealed that AI models have the potential to serve as dynamic, AI-driven tools for enhancing English grammatical competence. It provides students with a prompt-based learning environment, immediate feedback, and customized tasks, all of which may lead to quantifiable improvements in grammar proficiency. Similarly, Song and Song (2023) investigated how an AI-assisted language-learning tool improved Chinese language learners' writing skills and enthusiasm. The findings revealed that the experimental group demonstrated enhanced proficiency in writing grammar. Furthermore, Wei (2023) examined the impact of AI-assisted language-learning tools on EFL learners' self-regulated learning, L2 motivation, and English learning success. The findings demonstrated that, compared with the control group, the treatment group achieved significantly better grammar performance. Dehraoui (2024) investigated the impact of integrating authentic materials and generative artificial intelligence in EFL grammar classrooms to enhance students' grammar learning. The study revealed that students perceived generative artificial intelligence as effective for improving their grammar proficiency and that its integration enhances students' skills. Kucuk (2024) investigated the impact of AI intelligence on EFL students' grammar learning in EFL classes. The findings revealed that the experimental group significantly outperformed the control group in terms of grammar proficiency. Mohammed and Ja'ashan (2024) explored the impact of AI-driven contextual conversations on EFL students' grammar learning, and their findings revealed a significant improvement in students' grammar proficiency.

Conclusion

The main objective of this study was to measure the implementation of Copilot AI within an English-language learning context to determine its impact on the grammatical knowledge of EFL students, and the findings revealed that using AI tools such as Copilot AI assists learners in learning words efficiently and helps them recall words over a longer period. The findings of this study are helpful for teachers, students, and institutions.

The findings of this study are highly relevant to educators, learners, and academic institutions. The results show that compared with traditional methods, using copula AI in the classroom significantly enhances grammar learning. This provides empirical proof for the use of technology-assisted teaching. The application can serve as an additional feedback system, enabling instructors to better address the needs of large classrooms while offering individualized

grammatical support. The results show that students who used Copilot AI demonstrated better grammar and greater confidence, especially when actively participating in class. The delayed posttest results revealed that these students retained their grammatical knowledge for a long time, indicating that practising with AI helps language learning that lasts. Schools are realizing that the money they spend on AI-supported platforms helps learners improve their grammar skills, which may lead to better overall grades. Institutions that utilize Copilot AI are open to new ideas in digital education, enhancing their reputation for innovation in language teaching and keeping them aligned with global education trends.

Although this study offers valuable insights into the use of Copilot AI in learning contexts, it has several limitations. One drawback of this study is the inadequate monitoring of students' engagement with Copilot AI outside the classroom. Although students in the second experimental group were required to submit screenshots and reflection notes, it remains unclear how frequently they utilized the tool outside of class, which could impact the reliability of the results. Another limitation is that the study included both male and female students but did not examine or compare differences in grammar learning outcomes by gender. Another problem with the sample is that it included only intermediate-level learners. Lower- and higher-level students were not included, even though their reactions to Copilot AI might differ. The study was conducted at a single higher education institution, which limits the applicability of the findings to broader educational contexts. The treatment lasted approximately 5 weeks, making it difficult to determine whether Copilot AI is effective for learning grammar over a longer period.

Future research could expand the scope of this study by employing more rigorous monitoring techniques, such as digital tracking tools, to more precisely evaluate learners' authentic engagement with AI technologies beyond the classroom. Additionally, subsequent studies may examine gender differences to determine whether male and female learners experience similar benefits from Copilot AI or whether tailored instructional approaches are needed. Additional research could examine the integration of Copilot AI with other linguistic skills, such as writing or speaking, or assess its effectiveness in comparison to that of competing AI-driven platforms across various competency levels and cultural contexts. Future research could expand the scope of this study by incorporating students with diverse ability levels, ranging from novices to experienced learners, to examine whether Copilot AI has a uniform or varied impact across different language proficiencies. Furthermore, conducting similar experiments across multiple institutions rather than a single college would improve the generalizability of the findings and provide a more thorough understanding of the functionality of the Copilot AI across various educational contexts.

REFERENCES

Ahn, S. (2022). Effects of chatbots on the grammar competence of Korean EFL college students. *Journal of Digital Convergence, 20*(3), 53–61. https://doi.org/10.14400/JDC.2022.20.3.053

Al Ghaithi, A., & Behforouz, B. (2024). The use of an interactive chatbot in grammar learning. *Journal of Educators Online*, *21*(4). https://doi.org/10.9743/JEO.2024.21.4.12

Ali, Y. A., Ali, M. M., Alharbi, W., Hamid, S., & Abbas, S. (2025). Influence of ChatGPT on English grammar among Saudi ESL learners. *Arab World English Journal (AWEJ)*, 358–371. https://dx.doi.org/10.24093/awej/AI.21

Bannister, P., Alcalde Peñalver, E., & Santamaría Urbieta, A. (2024). International Students and Generative Artificial Intelligence: A Cross-Cultural Exploration of HE Academic Integrity Policy. *Journal of International Students*, *14*(3), 149-170. https://doi.org/10.32674/jis.v14i3.6277

Behforouz, B., & Ghaithi, A. A. (2024). Grammar gains: Transforming EFL learning with ChatGPT. *Educational Process International Journal*, *13*(4), 25–41. https://doi.org/10.22521/edupij.2024.134.2

Chen, X. (2026). Enhancing interactive English learning through embedded artificial intelligence technology. *International Journal of High Speed Electronics and Systems*, *35*(04), 2540727.

Dehraoui, C. (2024). *Exploring the impact of integrating authentic materials and ChatGPT in teaching grammar: A case of second-year students in the department of English at Ain Temouchent University* [Publised Master's thesis]. University of Ain Temouchent.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. Teachers College, Columbia University.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(1), 83–107. https://doi.org/10.2307/40264512

Elmotri, B., Harizi, R., & Boujlida, A. (2025). The impact of AI-generated feedback explicitness (generic vs. specific) on EFL students' use of automated written corrective feedback. *Arab World English Journal*, *16*(1), 384–402. https://dx.doi.org/10.24093/awej/vol16no1.24

Hadi, M. W., Suryadi, H., & Suparlan, S. (2026). The Effectiveness of Adaptive Artificial Intelligence (AI)-Based Learning in Enhancing EFL Learners' English Structure Mastery. *JOLLT Journal of Languages and Language Teaching, 14*(1), 223-235.

Ittefaq, M., Zain, A., Arif, R., Ahmad, T., Khan, L., & Seo, H. (2025). Factors influencing international students' adoption of generative artificial intelligence: The mediating role of perceived values and attitudes. *Journal of International Students*, *15*(7), 127-154. https://doi.org/10.32674/fnwdpn48

Kaçani, L., & Mangelli, S. (2013). Albanian teachers' perceptions about difficulties in teaching and learning grammar of EFL. *Journal of Educational and Social Research*, *3*(3), 149–156.

https://doi.org/10.5901/jesr.2013.v4n3p149
Katsarou, D. V., Mantsos, E., Papadopoulou, S., Sofologi, M., Efthymiou, E., Vasileiou, I., Megari, K., Theodoratou, M., & Kougioumtzis, G. A. (2025). Exploring AI technology in grammar performance testing for children with learning disabilities. *Education Sciences*, *15*(3), 351. https://doi.org/10.3390/educsci15030351
Kharis, M., Schön, S., Hidayat, E., Ardiansyah, R., & Ebner, M. (2022). Mobile gramabot: Development of a chatbot app for interactive german grammar learning. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(14), 52–63. https://doi.org/10.3991/ijet.v17i14.31323
Kim, N.-Y. (2019). A Study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence*, *17*(8), 37. https://doi.org/10.14400/JDC.2019.17.8.037
Kucuk, T. (2024). ChatGPT integrated grammar teaching and learning in EFL classes: A Study on Tishk international university students in Erbil, Iraq. *Arab World English Journal (AWEJ)*. https://papers.ssrn.com/abstract=4814669
Lai, C., & Gu, M. (2011). Self-regulated out-of-class language learning with technology. *Computer Assisted Language Learning*, *24*(4), 317–335. https://doi.org/10.1080/09588221.2011.568417
Lalira, J. E., Pangemanan, Y. A. T., Scipio, J. E., Lumi, S., Merentek, T. C., & Tumuju, V. N. (2024). Evaluating the impact of ai tools on grammar mastery: a comparative study of learning outcomes. *Voices of English Language Education Society*, *8*(3). https://doi.org/10.29408/veles.v8i3.27856
Mohammed, G. M. S., & Ja'ashan, M. M. N. H. (2024). Exploring the effect of ai-driven contextual conversations on EFL grammar learning at university level in Saudi Arabia. *Journal of Ecohumanism*, *3*(8), 11909–11924. https://doi.org/10.62754/joe.v3i8.5790
Ng, M. L., Behforouz, B., & Al Ghaithi, A. (2025). Grammar and engagement in focus: Evaluating Gemini AI's impact on an educational environment. *Computers and Education Open,9*. https://doi.org/10.1016/j.caeo.2025.100302
Muhsin, M. A. (2015). *The correlation between students' grammar knowledge and writing ability*. Muhammadiyah University of Makassar.
Nawaz, S., Ahmed, S. R., Aqeel, M., & Qureshi, H. H. (2022). Challenges of teaching grammar at elementary level: a qualitative study of ESL teachers' perceptions. *European Journal of English Language, Linguistics and Literature*, *2*(2), 1–6.
Normawati, A., & Nugrahaeni, D. A. (2024). Grammar teaching and learning in English language class: Students' view. *English Education: Journal of English Teaching and Research*, *9*(1), 41–48. https://doi.org/10.29407/jetar.v9i1.21536
O. Agregado, C. J. (2023). Trends and challenges in teaching grammar in the 21st century learning perspective: Their influence on learner's performance. *International Journal of Science and Management Studies*

(IJSMS), *6*(5), 151–190. https://doi.org/10.51386/25815946/ijsms-v6i5p111

Pamilu, A., Subarkah, A. I., & Anwar, M. F. Z. (2021). The use of WhatsApp group discussion to improve grammar competence in MAN 1 Banyumas. In *Proceeding of 1st International Conference on Enhancing English Language Teaching through the Use of Technology in the Industry 4.0 Era*, 99–111. https://doi.org/10.24090/celti.2021.270

Phieanchang, T. (2024). Using ChatGPT to improve English grammar skills for EFL learners. *Journal of Thonburi Rajabhat University*, *18*(1), 223–240.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Safitri, F., Hardini, T. I., Setiadi, R., & Mutiarsih, Y. (2021). *Usability measurement: Chatbot as a pedagogical support for learning French grammar*. 272–276. https://doi.org/10.2991/assehr.k.211119.042

Setiawan, F., & Alkhowarizmi, A. (2025). Exploring an artificial intelligence as automated feedback program in EFL writing. *ETERNAL: English Teaching Journal*, *16*(1), 202–224. https://doi.org/10.26877/eternal.v16i1.1206

Sharmithashini, M., & Hashim, H. (2025). Sustaining ESL writing development with AI-driven automated feedback systems: A systematic review (2006–2025). *International Journal of Research and Innovation in Social Science*, *6*(8), 531–550. https://doi.org/10.2139/ssrn.5065151

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1–14. https://doi.org/10.3389/fpsyg.2023.1260843

Stevens, J. P. (2025). Artificial intelligence in the teaching of English grammar. *International Journal of Research and Innovation in Applied Science*, *10*(6), 878–896.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford University Press.

Takala, A. (2016). *grammar teaching methods in EFL lessons: Factors to consider when making instructional decisions* [Published Masters thesis, University of Jyväskylä]. https://jyx.jyu.fi/bitstreams/1a4dd798-c55f-416e-9cf7-8b1fc949f2ba/download

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1–14. https://doi.org/10.3389/fpsyg.2023.1261955

Widodo, H. P. (2006). Approaches and procedures for teaching grammar. *English Teaching*, *5*(1), 122–141.

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings*, *2018*(1). https://doi.org/10.5465/AMBPP.2018.15903abstract

Yule, G. (2010). *The study of language* (4th edition). Cambridge University Press.

Author bios

ALI AL GHAITHI, PhD candidate, is an English Lecturer in the Faculty of Language Studies at Sohar University in Oman focusing on applied linguistics. Ali received his master's degree from the University of Putra Malaysia. Ali is interested in research studies that mainly implement artificial intelligence in teaching and learning processes. Email: AGhaithi@su.edu.om

BEHNAM BEHFOROUZ, PhD, is an English Lecturer at the Preparatory Studies Center at the University of Technology and Applied Sciences, Shinas, Oman. He has been teaching English at various Omani universities since 2015. His main areas of interest include TESOL and educational technologies. Email: Behnam.Behforouz@utas.edu.om

SALAH AL-HANSHI, PhD candidate, is an Omani instructional designer, lecturer and author. With expertise in EduTech, curriculum design, and vocabulary acquisition, he has developed e-learning platforms and authored educational resources, combining technology and pedagogy to enhance English language teaching and learning. Email: s.alhanshi@squ.edu.om