

Chatbot-Driven Writing Practice: A Boon or a Bust for EFL Learners?

Práctica de escritura impulsada por chatbots: ¿una bendición o una decepción para los estudiantes de EFL?

Eylül Karabulut²

Faculty of Education, Ondokuz Mayıs University, Türkiye

Prof. Dr. Emrah Ekmekçi¹

Faculty of Education, Ondokuz Mayıs University, Türkiye

Received: June 6, 2025 / Accepted: December 10, 2025

¹ emrah.ekmekci@omu.edu.tr

² eylulkarabulut1567@gmail.com

Abstract

While previous studies focused more on the impact of purpose-built AI-powered chatbots on students' L2 writing skills, a gap exists regarding the effects of general-purpose AI-powered chatbots on L2 writing skills. This study aims to address this gap by analyzing the impact of widespread use of AI chatbots on L2 writing skills. A quasi-experimental design was adopted, and a sequential explanatory mixed-methods approach was utilized involving 36 students (19 experimental, 17 control). Quantitative data were gathered from learners' writing tasks, while semi-structured interviews were used to collect qualitative data. Over five weeks, the experimental group used four chatbots for writing tasks, while the control group received teacher-led instruction. The results showed that the experimental group did not exhibit a statistically significant change in writing skills compared to the control group. Nevertheless, students in the experimental group held positive perceptions toward chatbots. Overall, the study offers valuable insights into integrating emerging technologies into English language teaching, providing critical guidance for stakeholders and practitioners on the responsible use of these tools and the pedagogical structuring of general-purpose AI chatbots.

Keywords: General-purpose AI-powered chatbots, artificial intelligence, English language learners, perceptions toward chatbots, writing skills

Resumen

Aunque estudios previos se han centrado principalmente en el impacto de chatbots de inteligencia artificial diseñados específicamente para fines educativos en el desarrollo de las habilidades de escritura en L2, persiste una brecha en la literatura respecto a los efectos de los chatbots de IA de uso general en dichas habilidades. Este estudio busca atender esta brecha mediante el análisis del impacto que tiene el uso extendido de chatbots de propósito general en la competencia escritural en L2. Se empleó un diseño cuasiexperimental y un enfoque mixto secuencial explicativo, con la participación de 36 estudiantes (19 en el grupo experimental y 17 en el grupo de control). Los datos cuantitativos provinieron de las tareas de escritura de los estudiantes, mientras que los datos cualitativos se obtuvieron mediante entrevistas semiestructuradas. A lo largo de cinco semanas, el grupo experimental utilizó cuatro chatbots para la realización de tareas de escritura, en tanto que el grupo de control recibió instrucción directa por parte del docente. Los resultados indican que el grupo experimental no presentó cambios estadísticamente significativos en sus habilidades de escritura en comparación con el grupo de control. Sin embargo, los estudiantes del grupo experimental manifestaron percepciones favorables hacia el uso de chatbots. En conjunto, el estudio ofrece aportes relevantes para la integración de tecnologías emergentes en la enseñanza del inglés, y proporciona orientaciones fundamentales para actores institucionales y profesionales sobre el uso responsable de estas herramientas y sobre las consideraciones pedagógicas necesarias para incorporar chatbots de IA de propósito general.

Palabras clave: chatbots de IA de uso general, inteligencia artificial, aprendientes de inglés, percepciones sobre los chatbots, habilidades de escritura

Resumo

Embora estudos anteriores tenham se concentrado principalmente no impacto de chatbots de inteligência artificial desenvolvidos especificamente para fins educacionais no desenvolvimento das habilidades de escrita em L2, ainda persiste uma lacuna na literatura quanto aos efeitos de chatbots de IA de uso geral nessas habilidades. Este estudo busca preencher essa lacuna por meio da análise do impacto do uso ampliado de chatbots de propósito geral na competência escrita em L2. Empregou-se um delineamento quase-experimental e uma abordagem mista sequencial explicativa, com a participação de 36 estudantes (19 no grupo experimental e 17 no grupo de controle). Os dados quantitativos foram obtidos a partir das tarefas de escrita dos estudantes, enquanto os dados qualitativos foram coletados por meio de entrevistas semiestruturadas. Ao longo de cinco semanas, o grupo experimental utilizou quatro chatbots para a realização das tarefas de escrita, ao passo que o grupo de controle recebeu instrução direta do professor. Os resultados indicam que o grupo experimental não apresentou mudanças estatisticamente significativas em suas habilidades de escrita em comparação com o grupo de controle. No entanto, os estudantes do grupo experimental manifestaram percepções favoráveis sobre o uso de chatbots. Em conjunto, o estudo oferece contribuições relevantes para a integração de tecnologias emergentes no ensino de inglês e fornece orientações fundamentais para gestores e profissionais quanto ao uso responsável dessas ferramentas e às considerações pedagógicas necessárias para incorporar chatbots de IA de propósito geral.

Palavras-chave: chatbots de IA de uso geral, inteligência artificial, aprendizes de inglês, percepções sobre chatbots, habilidades de escrita

Introduction

The use of emerging technologies, such as artificial intelligence (AI) tools, has helped to create a more engaging, motivating learning environment and enhanced the process of learning language skills such as vocabulary, grammar, and writing (Shadiev et al., 2023). This fact has also facilitated the widespread use of such technologies across various fields¹, including language teaching and learning (Shadiev et al., 2023). They also play an essential role in enhancing the effectiveness of language learning, such as increasing motivation and enriching learning experiences (Sumakul et al., 2021). Despite the positive impacts on language education, some studies raise concerns about the reliability and accuracy of general-purpose AI-powered chatbots and privacy issues (Alhazmi & Muftah, 2025). This also raises some uncertainties about the improvements achieved by general-purpose AI-powered chatbots. That is, some experts question whether the effectiveness relies on learner-related or contextual factors (Safar & Anggraheni, 2024). Therefore, integrating general-purpose AI-powered chatbots into education can be done responsibly and cautiously. Besides, it requires context-specific research that highlights the pedagogical implications of such utilizations. It is also important to highlight their use in more depth by focusing on the four basic language skills: listening, speaking, reading, and writing.

Emerging technologies such as chatbots have now supported these four pillars of a language. These tools can be intelligent assistants and/or conversational agents that provide individualized answers to human prompts using natural language processing (NLP) and machine learning. They have recently been favored in language teaching due to their various functions for teaching and learning the four language skills. For example, a chatbot can provide instant feedback by recognizing missing points and mistakes using a large data set. Despite the widespread use of these tools in language learning, speaking and writing skills have been the most frequently studied in current research (Gayed et al., 2022). Although some studies have examined the effects of AI-powered chatbots on writing skills (Mahapatra, 2024; Barrot, 2023; Song & Song, 2023), research has tended to focus on the impact on speaking skills rather than on writing skills. Considering this, to our knowledge, few studies (Mahapatra, 2024) have analyzed the possible effects of general-purpose AI-powered chatbots on the writing skills of preparatory class students learning English. To fill this gap and contribute to the related literature, we investigated the impact of general-purpose AI-powered chatbots on the writing skills of preparatory class students, including their vocabulary use, text cohesion and coherence, and students' perceptions of using these tools throughout writing instruction. Therefore, the current study addressed the following research question:

1. How do general-purpose AI-powered chatbots affect the development of writing skills among English preparatory class students?

2. What are students' perceptions of the implementation of general-purpose AI-powered chatbots in writing instruction?

Literature Review

In a foreign language class, developing writing skills can be seen as both vital and challenging. However, the definition of writing and what constitutes that process are not monolithic; in other words, there may be various points regarding that concept, and the process might vary across different contexts in L2 learning, as well as including the interplay between cognitive and emotional processes (Vasylets & Marin, 2021). Considering learners' perspectives, recent research indicates that the L2 writing process can be complex, requiring the retrieval of specific vocabulary and grammar items during the planning and revising stages, which also entails cognitive exertion (Güvendir & Uzun, 2023). However, it remains a crucial skill for language learners to achieve. Considering this, the emergence of CALL has led to widespread interest in the use of technology in language classes, such as writing assistants and Web 2.0 tools. In the context of teaching writing, these tools are also frequently used because they can yield positive results. For example, Barrot (2023) shows that using digital tools can enhance the learners' writing performance. In addition, improvements in natural language processing and machine learning have led to the widespread adoption of intelligent technologies in L2 writing skills. Following that, general-purpose AI-powered tools have been integrated into EFL writing classes, and studies demonstrate that their use can improve language proficiency, including grammar and vocabulary (Sumakul et al., 2021). Besides, these tools might also provide instant, detailed feedback on students' output (Escalante et al., 2023).

In this case, the potential effects of general-purpose AI-powered chatbots are also important, as they may offer promising results in writing performance. Although they have been launched recently, several studies have examined their impact on writing skills, including guidance for learners and instant feedback (Song & Song, 2023). Other studies pointed out that general-purpose AI-powered chatbots can guide students in structuring their text and brainstorming (Mahapatra, 2024). Similarly, Harunasari (2023) used a chatbot to generate ideas and request feedback on grammar structures and suitable vocabulary items. The results showed improvement in the learners' writing performance via pre-test and post-test.

However, despite the positive effects of general-purpose AI-powered chatbots on writing skills, some drawbacks might occur. For example, Yan (2023) states that utilizing these tools might raise problems regarding higher-order thinking skills. Since these tools provide an easier way to access knowledge, learners might not need to analyze and critically consider the information, leading to more superficial thinking.

Nevertheless, using general-purpose AI-powered chatbots might yield positive outcomes regarding writing instruction. Having discussed these effects, pointing out the learners' insights into the use of these tools in writing classes is necessary. Studies that analyze this aspect mostly show that students hold favorable opinions about using general-purpose AI-powered chatbots for writing instruction (Duong & Chen, 2025).

Methodology

Research Design

The purpose of the current study was to analyze the effects of general-purpose AI-powered chatbots on students' writing performance in an English preparatory class and on their perceptions of using these tools in their writing classes. A quasi-experimental design was employed to profoundly investigate the impact of these tools on writing performance. The reason for adopting this design is its suitability for investigating cause-and-effect relationships in a natural classroom setting and its ability to compare results across groups by controlling extraneous variables through pre-and post-tests. Given the current study, adopting this design was more suitable for the participants, as they had already been assigned to a particular group by the institution and could not change their groups mid-semester. In addition, the study was more quantitative and data-oriented, and qualitative data were used to further explain it. Therefore, a sequential explanatory mixed-methods design was preferred to gather both quantitative and qualitative data (Creswell & Creswell, 2023). After the intervention, the qualitative data were collected from the experimental group through semi-structured interviews.

Setting and Participants

36 students studying medicine and English language teaching were recruited for the study, and they enrolled in an English preparatory class at Ondokuz Mayıs University in Türkiye. They were divided into two groups: control and experimental. While 19 students belonged to the experimental group, 17 belonged to the control group. In the School of Foreign Languages, the students enrolling in certain faculties must first take a general language proficiency exam. If they fail to meet the university-determined minimum score, they must enroll in the institute's preparatory English program. The participants were assigned to their groups based on the results of the placement test, which were analyzed using the Common European Framework of Reference for Languages (CEFR). The students were divided into groups, and in each group, the writing courses were taught by different instructors using the same materials and course syllabus. These courses were provided once a week and took half an hour.

Since the students had already been assigned to specific groups, the study groups were decided without random sampling. For qualitative data collection, 13 volunteer students from the experimental group were randomly selected after the intervention.

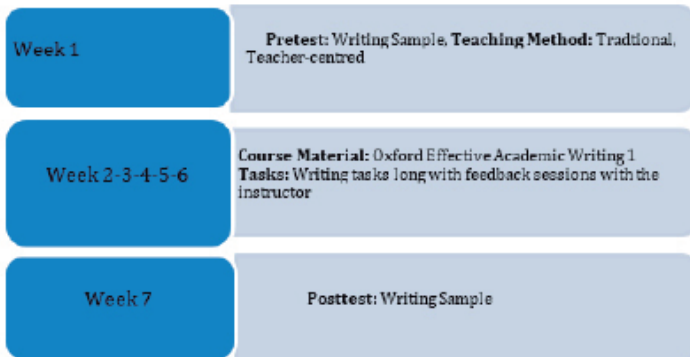
Data Collection Tools

For the quantitative data, a writing assessment scale was used to investigate learners' writing performance before and after the treatment. The instructors at Ondokuz Mayıs University developed an analytic writing assessment rubric, including five sections: relevance and adequacy of content, organization, grammar, lexical range, mechanics, and spelling. Qualitative data were collected to investigate their thoughts on using chatbots while writing text. To this end, a semi-structured interview comprising five open-ended questions was conducted. The purpose of this kind of interview was to elicit further answers that might not be expected by the researcher (Braun & Clarke, 2013). The interview's questions were prepared by consulting professionals in the field. They covered their thoughts on their learning performance, the challenges they faced, and the benefits they experienced throughout this process. The interview also aimed to collect data on changes in students' writing habits after using AI-powered chatbots.

Procedure

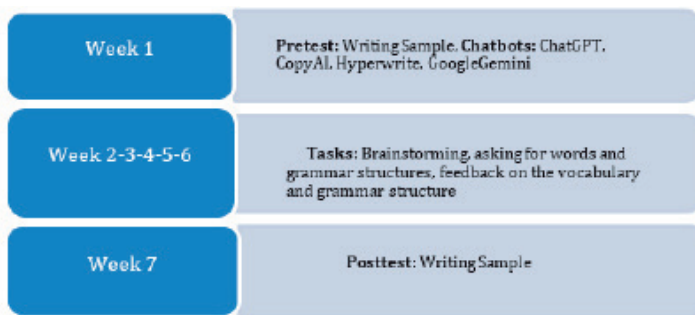
The researcher designed a writing module, and the procedure lasted seven weeks, including pre-test and post-test sessions, while the intervention sessions lasted five weeks. The module was offered to the experimental group, while the control group students received teacher-centered instruction. The procedure for both groups are displayed in Figures 3.1 and 3.2 as follows.

Fig. 1. Writing Instruction in the Control Group



As shown in Fig. 3.1, in the first week, both groups were assigned writing tasks as a pre-test. In the second week, while the control group students received writing tasks and got feedback from their instructor on their use of words and grammar structures, the experimental group students were introduced to the first chatbot (i.e., ChatGPT) as illustrated in Fig. 3.2.

Fig. 2. Writing Instruction in the Experimental Group



The selected chatbots were ChatGPT, Google Gemini, Hyperwrite, and CopyAI, and each was used differently each week. Following this brief introduction to writing prompts and technical features (e.g., subscriptions, website interfaces), they were provided with writing topics from their main coursebook. The tool was used to brainstorm and gather feedback on vocabulary and grammar. To exemplify, for the topic related to their favorite celebrity, they asked questions to the AI-powered chatbots as “I want to write about person A (the celebrity’s name), can you provide an outline for my text?” or “I have used the X word, can you suggest an alternative?” “Is the word or the sentence structure suitable for this context?” Although each writing class usually lasts 30 minutes, the intervention sessions lasted 80 minutes in total because of the additional 10 minutes. In the final week, both groups received a post-test as a writing task. They were not allowed to use any digital tools, such as chatbots, dictionaries, and translators, and results were based on the students’ core abilities. These writing sessions were carried out under the instructors’ supervision. In total, 13 volunteer students attended the semi-structured interview.

Analysis of the Data

IBM SPSS 26 was employed to investigate the quantitative data. Kappa analysis and Weighted Cohen’s Kappa were utilized to ensure the inter-rater reliability

of the test scorers. However, ChatGPT Plus was used to analyze Weighted Cohen's Kappa, ensuring the study's validity. The Shapiro-Wilk Test was used to analyze the normality of the data. The mean and standard deviation for normally distributed variables were reflected in the descriptive statistics of numerical variables. Skewness and kurtosis coefficients were used to assess the conformity of the numerical variables with the normal distribution. For the control group, the Skewness statistics were -0.157 in the pre-test and 0.083 in the post-test, while they were -0.406 (pre-test) and -0.329 (post-test) for the experimental group. Considering the Kurtosis coefficients, they were -0.300 (pre-test) and -0.285 (post-test) for the control group, and the experimental group's statistics were -1.110 (pre-test) and -0.630 (post-test). To assess statistical significance between these two groups, independent samples t-tests and paired samples t-tests were used, and $p < 0.05$ was considered statistically significant. For qualitative data analysis, thematic analysis was preferred, which allowed for the generation of themes and patterns (Braun & Clarke, 2013). The transcribed data were divided into codes and themes, and to ensure reliability and validity, another ELT professional revised the codes and the themes by comparing the first findings.

Results

Regarding the first research question about how general-purpose AI-powered chatbots affect the development of writing skills among English preparatory class students, it was found that experimental group students showed no statistically significant difference in their writing performance after using AI-powered chatbots in their writing classes. The related statistical findings including inter-rater reliability and normality analysis are presented in Tables 4.1, 4.2, 4.3, 4.4, and 4.5. In addition, Tables 4.6, 4.7, 4.8, and 4.9 present the descriptive statistics and overall writing score for both the control and experimental groups, broken down by each category in the rubric. Table 4.1 below illustrates the inter-rater reliability of the instructors who scored the writing tasks for both the control and experimental groups. To assess scorers' agreement on the writing tasks, the Kappa test was used. The Kappa test used for the pre-test indicates that inter-rater reliability was ensured in both the experimental and control groups.

Table 1. Inter-rater Reliability Results of the Pre-test–Writing Sample

Groups	Sub-Categories	Kappa Statistics	p-value
Control Group	Relevance and Adequacy of Content	0.73	0.046*
	Organization	0.81	0.011*
	Grammar	0.88	0.009*
	Lexical Range	0.78	0.029*
	Mechanics and Spelling	0.79	0.027*
Experimental Group	Relevance and Adequacy of Content	0.85	0.013*
	Organization	0.82	0.017*
	Grammar	0.76	0.035*
	Lexical Range	0.87	0.025*
	Mechanics and Spelling	0.74	0.020*

Similarly, in Table 4.2, the inter-reliability test findings show that the scores provided by the two instructors were reliable, and there is strong agreement between them.

Table 2. Inter-rater Reliability Results of the Posttest – Writing Samples

Groups	Sub-Categories	Kappa Statistics	p-value
Control Group	Relevance and Adequacy of Content	0.72	0.032*
	Organization	0.78	0.001
	Grammar	0.84	0.015*
	Lexical Range	0.90	0.001*
	Mechanics and Spelling	0.76	0.033*
Experimental Group	Relevance and Adequacy of Content	0.88	0.009*
	Organization	0.72	0.032*
	Grammar	0.71	0.023*
	Lexical Range	0.82	0.017*
	Mechanics and Spelling	0.77	0.021*

Table 3 (pre-test) and 4 (post-test) display the results of the analysis using quadratic-weighted Cohen's kappa.

Table 3. Inter-rater Reliability Results of the Pre-test Writing Results Based on Weighted Cohen's Kappa Results (Control & Experimental Groups)

Subscale	Kw (Cont)	95% CI (Cont)	N (Cont)	Kw (Exp)	95% CI (Exp)	N (Exp)
Relevance & Adequacy of Content	.481	[.143, .752]	21	.321	[-.023, .622]	19
Organization	.273	[-.032, .533]	21	.246	[-.038, .515]	19
Grammar	.357	[-.025, .635]	21	.715	[.399, .882]	19
Lexical Range	.184	[-.197, .533]	21	.647	[.341, .820]	19
Mechanics & Spelling	.512	[.167, .831]	21	.653	[.334, .811]	19

Table 4. Inter-rater Reliability Results of the Post-test Writing Results Based on Weighted Cohen's Kappa (Control & Experimental Groups)

Subscale	kw (Con)	95%CI (Cont)	N (Cont)	kw (Exp)	95% CI (Exp)	N (Exp)
Relevance & Adequacy of Content	.454	[.134, .711]	21	.691	[.417, .857]	19
Organization	.361	[.022, .587]	21	.655	[.332, .837]	19
Grammar	.517	[.098, .765]	21	.768	[.493, .901]	19
Mechanics & Spelling	.494	[.173, .707]	21	.726	[.457, .859]	19

The data for the experimental group illustrates that the agreement was moderate to substantial, especially in the Grammar and Lexical Range sections, while in the control group, agreement levels ranged from slight to significant. Overall, the post-test results (Table 4.4) revealed greater consistency among the raters. In Table 4.5, the results of the normality test indicate that the data follow a normal distribution.

Table 5. Shapiro–Wilk Test of Normality Results

Group	Statistic	df	Sig.
Control Group Pre-Test	,959	19	,544
Experimental Group Pre-Test	,905	19	,059
Control Group Post-Test	,956	19	,500
Experimental Group Post-Test	,930	19	,173

On the other hand, Table 6 presents the descriptive statistics for both groups. The pretest results show a similarity in their writing scores collected before using the general-purpose AI-powered chatbots.

Table 6. Descriptive Statistics of the Control Group and Experimental Group

Group	Test	Min	Max	Mean	Std. Dev.	Sk.St.	Kt. St.
Control Group	Pretest	6	15.50	11.11	2.60	-0.157	-0.300
	Posttest	6	19.00	12.66	3.34	0.083	-0.285
Experimental Group	Pretest	5.50	16.00	11.26	3.40	-0.406	-1.110
	Posttest	5.00	17.00	11.26	3.75	-0.329	-0.630

The main findings gathered after the intervention session indicate that teacher-centered instruction had the potential to enhance the control group students' writing performance, as indicated by the pre-test results (11.11) and the post-test results (12.66). Regarding the experimental group students, no significant changes were observed before and after adopting the chatbots. The findings in Table 4.7 indicate a statistically significant change in writing performance in control group.

Table 7. Control Group and Experimental Group Students' Writing Scores Total

	Variable	N		ss	sd	t	p
Control Group	Pretest	21	11.11	2.60	20	-2.320	
	Posttest	21	12.66	3.34			
Experimental Group	Pretest	19	11.26	3.40	3.95	0.000	1.000
	Posttest	19	11.26	3.75			

Overall, the results of the statistical analysis of the experimental group students show no statistically significant change in their writing performance after using AI-powered chatbots in their writing classes (pre-test = 11.26, post-test = 11.26). Table 4.8 shows the findings of the investigation conducted to detect changes in the control group across the subcategories of writing skills. It shows that there was a statistically significant change in the “Mechanics and Spelling” section (pre-test=1.88, post-test=2.40). Considering this, teacher-centered writing classes led to improvements in the mechanics and spelling skills of the control group.

Table 8. Pre and Posttest Results Regarding the Sub-Categories of the Control Group

Sub-Categories	Variable	N		ss	sd	t	p
Relevance and Adequacy of Content	Pretest	21	2.80	0.71	20	-0.761	0.456
	Posttest	21	2.97	0.85			
Organization	Pretest	21	2.14	0.57	20	-1.139	0.268
	Posttest	21	2.33	0.67			
Grammar	Pretest	21	2.02	0.69	20	-1.783	0.090
	Posttest	21	2.30	0.84			
Lexical Range	Pretest	21	2.28	0.62	20	-2.000	0.059
	Posttest	21	2.61	0.72			
Mechanics and Spelling	Pretest	21	1.88	0.70	20	-3.859	0.001*
	Posttest	21	2.40	0.75			

However, in Table 9, there was no statistically significant change in the experimental group after using chatbots, despite a slight improvement in the organization (pre-test=2.07, post-test=2.18) and lexical range categories (pre-test=2.31, post-test=2.55).

Table 9. Pre and Posttest Results Regarding the Sub-Categories of the Experimental Group

Sub-Categories	Variable	N		ss	sd	t	p
Relevance and Adequacy of Content	Pretest	19	2.31	0.60	18	0.170	0.867
	Posttest	19	2.28	0.87			
Organization	Pretest	19	2.07	0.55	18	-0.639	0.531
	Posttest	19	2.18	0.76			
Grammar	Pretest	19	2.34	0.95	18	0.818	0.424
	Posttest	19	2.13	0.84			
Lexical Range	Pretest	19	2.31	0.71	18	-1.013	0.324
	Posttest	19	2.55	0.95			
Mechanics and Spelling	Pretest	19	2.18	0.85	18	0.236	0.816
	Posttest	19	2.13	0.76			

The second research question focuses on the students' perceptions after using the general-purpose AI-powered chatbots in their EFL writing classes. The findings are illustrated in Table 10, which presents three main themes: enhancement of lexical capacity, generation of ideas, and awareness of the target language. These themes are also supported by the codes generated from the students' transcripts. Overall, the

results of the interviews show that the students in the experimental group had positive perceptions of general-purpose AI-powered chatbots for developing vocabulary and grammatical knowledge, as well as brainstorming and generating ideas.

Table 10. The Themes and Codes of the Students' Responses to the Interview Questions

Themes	Codes
Enhancement of the Lexical Capacity	Using high-level words
	Spelling
	Using varied advanced vocabulary items
Generation of Ideas	Initiating the ideas for the text
Awareness of the Target Language Structure	Assisting with the use of grammar

Regarding thoughts on developing vocabulary knowledge, most participants held the view that they could find a more advanced and diverse set of vocabulary items through chatbots which can be seen in Table 4.10 "Enhancement of the Lexical Capacity part". For example, participant 9 stated that *"In the past, I always used the same words; now, I use more advanced and suitable vocabulary items."* The participants also noted that AI-powered chatbots helped them become aware of and use appropriate grammar structures, as in the Awareness of the Target Language Structure section in Table 4.10, and assisted them in developing their grammar skills. Student 13 stated that *"They allowed us to use new words and sentence structures and created prototypes about how we should think. We became more experienced in terms of writing. It completed our knowledge gap about words and grammar."*

Regarding the "Generation of Ideas" in Table 4.10, they also stated that these tools serve as guides to help them explore new ideas and start the writing process. Student 8 said, *"These tools have had a positive impact on my writing. They helped us with writing and gave us ideas. For example, when I could not decide on how to start writing, they helped me determine the beginning"*. To sum up, the results of the semi-structured interviews indicate that the students held positive perceptions of integrating general-purpose AI-powered chatbots into their writing classes, as they could provide assistance with the appropriate use of vocabulary items and grammar structures, as well as guidance to initiate the writing process.

Discussion

The overall findings from the quantitative data indicate that teacher-led writing tasks were more effective than those in which students used only general-purpose AI-powered chatbots. Although some studies suggest that general-purpose AI-powered chatbots led to improvements in writing performance (Harunasari, 2023; Yan, 2023), our findings show the opposite, underscoring the importance of the role assigned to teachers. The findings can be interpreted as outcomes of the novelty effect, which refers to using a tool for a while and its effects wearing off over time. Huang et al. (2021) state that this concept means learners may feel less enthusiastic about using a specific tool after using it for a while. This explanation can also be supported by Tsay et al. (2020), who showed that participants tend to lose motivation to use a particular tool once the tool's initial impact diminishes as they become familiar with it. Considering this, the experimental group students were exposed to these tools for five weeks, and they might have begun to lose the motivation and enthusiasm they had at the beginning. Not finding a statistically significant change in the writing performance could also stem from the limited knowledge of writing suitable prompts. The experimental group received a brief introduction to writing prompts that was not very detailed. This can be another reason why these tools failed to improve their writing performance. Despite these findings, there was a slight improvement in the experimental students' vocabulary use with the help of general-purpose AI-powered chatbots. Kim (2018) studies changes in students' vocabulary acquisition using chatbots. A similar result was also found in organizing the text's structure, which refers to the concept of coherence. It can be inferred that general-purpose AI-powered chatbots have the potential to enhance students' ability to write meaningful, logical text. A study conducted by Mahapatra (2024) also shows that using such tools may help develop skills in structuring a text by taking coherence and cohesion into account.

The transcripts show that the participants had positive views on integrating these tools, as they believed the chatbots could provide ways to use a wide range of vocabulary items and grammar structures in context. They could also assist them in generating ideas and brainstorming before starting the writing process. Several studies in the field have similar results, which can support the benefits of these tools as stated by our participants. Sumakul et al. (2021), for example, noted that students using AI-powered chatbots could receive guidance on vocabulary and grammar use. Another study, carried out by Mohamed and Alian (2023), aimed to investigate the use of AI-powered chatbots and students' perceptions of them, and found that students held positive perceptions. Considering their benefits for brainstorming and generating brand-new ideas, Xiao and Zhi (2023) found that the chatbots helped students explore new ideas during their writing process. Although the students in the current study showed little change in their writing skills after using general-purpose AI-powered chatbots, they perceived them as valuable tools for improving their vocabulary and grammar.

Conclusion

The present study was designed to investigate the impact of general-purpose AI-powered chatbots on the English preparatory students' writing skills and their perceptions of using them. The quantitative data did not show a statistically significant change in the writing performance of the students in the experimental group. Nevertheless, there were minor improvements in the lexical range and organization of the ideas sections. On the other hand, the control group students displayed an improvement in their writing skills. Regarding the semi-structured interviews, the students generally have positive perceptions of using these tools in the writing class. Besides, the most obvious finding to emerge from these interviews is that the chatbots were said to improve their vocabulary and grammar and to help them generate ideas. Taken together, these results suggest that general-purpose AI-powered chatbots remain effective at enhancing learners' vocabulary and grammar knowledge. However, the analysis showed no significant change in the experimental group. At the same time, teacher-led instruction plays a vital role in teaching language skills, particularly writing. Considering the students' responses to the interview questions, these tools can create a learning environment where learners can access a wide range of information related to vocabulary and grammar use. These findings also have implications for the understanding that earlier teacher-led instruction still holds an important place in language learning. However, these emerging technologies offer various benefits for learners when used not as the primary teaching means or source in the classroom, but rather as assistants for both the teacher and the learners. The results also indicate that teachers and learners alike need proper, detailed training on how to use these effectively to meet their teaching and learning needs.

References

- Alhazmi, A., & Muftah, M. (2025). Evaluating ChatGPT's reliability in second language acquisition (SLA): Insights on language skills and technology's role. *Electronic Journal of e-Learning*, 23(1), 96-112. <https://doi.org/10.34190/ejel.23.1.3762>
- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Journal of Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Routledge.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed method approaches* (6th ed.). SAGE Publications.
- Duong, T. N .A., & Chen, H. L. (2025). An AI chatbot for EFL writing: Students' usage tendencies, writing performance, and perceptions. *Journal of Educational Computing Research*, 63(2), 406-430. <https://doi.org/10.1177/073563312413123632>
- Escalante, J., Pack, A. & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Education in Technology Higher Education*, (20)57. <https://doi.org/10.1186/s41239-023-00425-2>
- Güvendir, U. & Uzun, K. (2023). L2 writing anxiety, working memory, and task complexity in L2 written performance. *Journal of Second Language Writing*, 60. <https://doi.org/10.1016/j.jslw.2023.101016>
- Gayed, J. M., Carlon, M. K., Oriola, A. M. & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computer and Education: Artificial Intelligence*, 3, Article 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Harunasari, S. (2023). Examining the effectiveness of AI-integrated approach in EFL writing: A Case of ChatGPT. *International Journal of Progressive Sciences and Technologies*, 39(2), 357-368. [10.52155/ijpsat.v39.2.5516](https://doi.org/10.52155/ijpsat.v39.2.5516).
- Huang, W., Hew, K. F. & Fryer, L. (2021). Are chatbots for learning really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237-257. <https://doi.org/10.1111/jcal.12610>
- Kim, N. Y. (2018). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence*, 16(2), 1-7. <https://doi.org/10.14400/JDC.2018.16.2.001>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a

- mixed methods intervention study. *Smart Learning Environments*, 11(9). <https://doi.org/10.1186/s40561-024-00295-9>
- Mohamed, S. S. & Alian, E. M.I. (2023). Students' attitudes toward using chatbots in EFL learning. *Arab World English Journal*, 14(3), 15-27. <https://doi.org/10.24093/awej/vol14no3.2>
- Safar, M., & Anggraheni, D. (2024). Language learning through AI chatbots: Effectiveness and cognitive load analysis. *Journal of Social Science Utilizing Technology*, 2(3), 430-445. <https://doi.org/10.70177/jssut.v2i3.1346>
- Shadiev, R., Wen, Y., & Uosaki, N. (2023). Future language learning with emerging technologies. *Journal of Computers in Education*, 10, 463-467. <https://doi.org/10.1007/s40692-023-00285-9>
- 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 Psychology*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Sumakul, D., Hamied, F. A. & Sukyadi, D. (2021). Students' perceptions of the use of AI in a writing class. In *Proceedings of the 67th TEFLIN International Virtual Conference & the 9th ICOELT 2021* (pp. 52-57). Atlantis Press. <https://doi.org/10.2991/assehr.k.220201.009>
- Tsay, C. H. H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2020). Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning*, 36(2), 128-146. <https://doi.org/10.1111/jcal.12385>
- Vasylets, O., & Marin, J. (2021). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing*, 52, Article 100786. <https://doi.org/10.1016/j.jslw.2020.100786>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages* (8)3, 212. <https://doi.org/10.3390/languages803021>
- Yan, D. (2023). Impact of ChatGPT on learners in an L2 writing practicum: An exploratory investigation. *Educational Information Technology*, 28, 13943-13967. <https://doi.org/10.1007/s10639-023-11742-4>

Authors

Eylül Karabulut works as an English language teacher in Türkiye. Besides, she holds an MA degree in the field of English language teaching. Her research interests encapsulate English language teaching, intercultural communication, technology integration into language classes.

ORCID: <https://orcid.org/0009-0006-4353-9569>

Emrah Ekmekçi currently works as a Prof. Dr. in the English Language Teacher Education Program, Faculty of Education at Ondokuz Mayıs University. He holds an MA from the English Language and Literature Department of Suleyman Demirel University and a PhD from the English Language Teacher Education Program at Gazi University. His research interests include Foreign Language Teacher Education, Technology-Enhanced Language Learning and Teaching, Flipped Learning, and Second Language Writing. He has many publications at national and international level on foreign language learning and teaching. He is the chief-editor of Journal of Language Research (JLR), which is an international refereed journal.

ORCID: <https://orcid.org/0000-0001-5585-8512>

This study is part of an MA thesis titled “Analysing the effects of AI-powered chatbots on the writing skills of English preparatory class students: An experimental study” written by Eylül Karabulut and supervised by Dr. Emrah Ekmekçi.

The authors declare that artificial intelligence tools were used only for a single purpose during the preparation of this manuscript: computing the Weighted Cohen’s Kappa coefficient for inter-rater reliability, carried out with ChatGPT Plus. All AI-assisted analytical outputs were manually reviewed, verified, and validated by the authors, and the use of AI is explicitly acknowledged in the Analysis of the Data section. No AI tools were used to generate or revise text, nor to create or modify images or figures. The manuscript fully complies with GiST’s ethical and editorial guidelines.