Article

Journal of Educational Computing Research
2025, Vol. 63(2) 406–430

DOI: 10.1177/07356331241312363
journals.sagepub.com/home/jec
Sage

An AI Chatbot for EFL Writing: Students' Usage Tendencies, Writing Performance, and Perceptions

Thi-Ngoc-Anh Duong[1] and Hsiu-Ling Chen[1]

Abstract

Writing plays a crucial role in the development of English as a Foreign Language (EFL) learners' language; however, it remains a challenging skill for them to acquire. This study investigated how EFL students at a high school in Northern Vietnam engaged with the Writing Assistant Bot (WAB), an artificial intelligence (AI) chatbot designed to support their writing practice at home. Focusing on students' usage patterns, writing performance, and perceptions, the research included 47 participants, categorized into higher and lower proficiency levels. The mixed-method approach was employed, with chat logs, timed-writing tests, questionnaires, and semi-structured interviews. The findings indicated differences in chatbot usage between the two proficiency levels at various writing stages, despite similarities in their focus on specific writing aspects. Lower-level learners predominantly utilized the chatbot during the Planning stage to generate vocabulary and brainstorm ideas, while higher-level learners mainly used it during the Translating stage to elaborate on ideas and refine their language for diverse and coherent expression. The chatbot significantly enhanced writing performance across all aspects: content, organization, vocabulary, language use, and mechanics, for both levels. Students perceived it as a useful and easy-to-use tool.

[1]Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan

Corresponding Author:
Hsiu-Ling Chen, National Taiwan University of Science and Technology, No.43, Keelung Rd., Sec.4, Taipei 10607, Taiwan.
Email: shirley@mail.ntust.edu.tw

Keywords
AI chatbot, English as a foreign language writing, usage, writing performance, perception

Introduction

Writing is undeniably a pivotal facet in the growth of EFL learners, standing out as one of the key productive skills (Haiyan & Rilong, 2016). Despite its importance, writing remains the most challenging skill among these learners (Şenel, 2018). Students often struggle with vocabulary selection (Alisha et al., 2019), grammar (Nguyen et al., 2021), and idea development due to limited background knowledge and resources (Amalia et al., 2021). Additionally, with limited class time and large class sizes, many teachers struggle to provide personalized feedback (Ali & Ramana, 2018). Tran (2007) observed that teachers often focus on writing analysis in class, leaving practice for homework. Consequently, students have limited opportunities for direct interaction with instructors to address their individual needs and problems during their practice sessions, thereby hindering their full comprehension and engagement in writing. Thus, it is necessary to facilitate students' writing practice process with personalized assistance.

In recent years, the integration of AI chatbots into education has gained considerable interest as a potential solution to language learning problems (Chen et al., 2020). A chatbot is a computer program aiming to mimic human conversation through AI to generate responses (Hsu et al., 2023). According to Huang et al. (2022), in language learning contexts, chatbots typically exhibit three features. First, they excel in generating and expanding ideas (Guo et al., 2022; Kohnke et al., 2023), acting as a repository of extensive language knowledge. Second, chatbots function as tireless assistants. They promptly engage with learners, providing consistent and continuous support throughout their language learning journey (Jeon, 2021; Kuhail et al., 2023). Third, chatbots can provide 24/7 support, allowing learners to practice language skills anytime and anywhere at their own pace (Belda-Medina & Calvo-Ferrer, 2022). Overall, it has been indicated that AI chatbots have the potential to assist students' writing process.

The most prominent chatbot so far is ChatGPT, an advanced AI language model that excels in engaging and meaningful conversations. Developed by Open AI, ChatGPT is able to provide insightful responses, answer questions, and even offer creative suggestions on a wide range of topics (Fuchs, 2023). As an AI-driven conversational partner, ChatGPT opens up new possibilities for enhancing communication and problem-solving across different domains, including EFL writing practice, thereby calling for a new wave of research on its potential (Boudouaia et al., 2024; Su et al., 2023; Tseng & Lin, 2024).

While research on AI chatbots is growing, studies examining the influence of AI chatbots such as ChatGPT on EFL writing remain limited, particularly in high school settings with a focus on different proficiency levels. To address this gap, for this study,

we developed a chatbot based on the ChatGPT language model and evaluated its usage by high school students at various proficiency levels as well as its effectiveness as an AI-powered writing support tool on students' writing performance and perceptions. The study aimed to contribute to the expanding body of knowledge concerning AI-driven educational tools, and to provide practical insights for stakeholders interested in integrating AI technologies into EFL writing.

Literature Review

Writing and EFL Writing Problems

According to Flower and Hayes (1981), writing is a series of distinct cognitive processes that writers coordinate and structure while they are composing. They further introduced a fundamental framework for delving into the cognitive dimensions of writing. According to the model, writing includes three main processes: Planning, Translating and Reviewing (Figure 1). In the Planning process, writers generate ideas, plan the structure of the document, and establish goals for the writing task. The Translating process requires writers to convert their planned ideas into written language. The Reviewing process involves two main sub-processes: evaluation and revision, in which writers can intentionally review their work.

Nunan (1989) further asserted that writing is an intricate metacognitive process that demands the simultaneous management of various variables. In a similar vein, Jacobs (1981) pointed out five fundamental components in writing, namely: content,

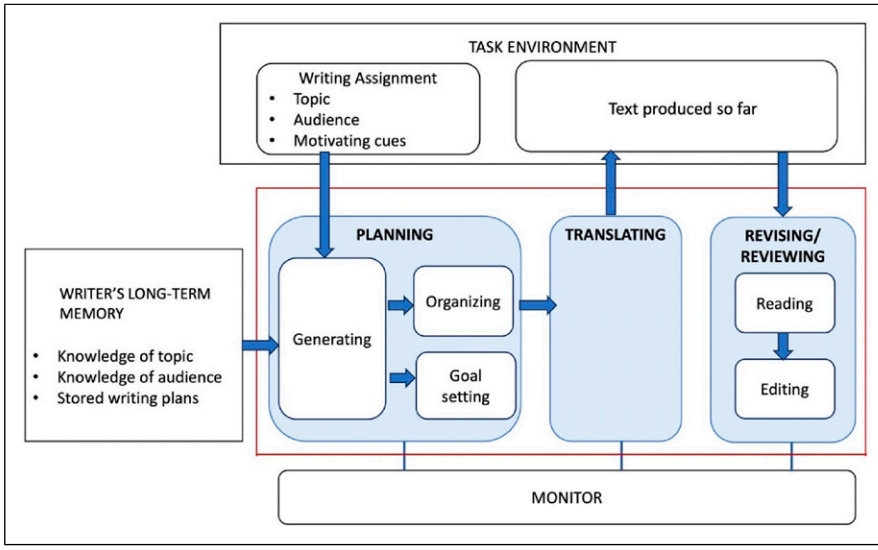

Figure 1. Structure of the writing model (Adapted from Flower & Hayes, 1981).

organization, vocabulary, language use, and mechanics. In this context, content refers to the relevance, accuracy, and development of the arguments. Organization is the arrangement and structure of the written work. Vocabulary indicates the selection and usage of appropriate words and phrases. Language use refers to the choice of grammar and syntax. Finally, mechanics covers the technical aspects of writing, such as spelling, punctuation, capitalization, and formatting.

Writing plays a vital role in language production. However, writing is not an innate skill; rather, it is something humans must learn to acquire. Tuan (2010) emphasized that the more one dedicates oneself to practicing writing, the more proficient one will become. However, EFL students often lack opportunities to practice writing due to the product-based approach whereby most class time is dedicated to writing instruction, leaving the writing practice part as homework (Tran, 2007) without adequate support.

Writing skills are challenging to attain, and this difficulty is notably amplified for EFL learners (Teng et al., 2022). Moreover, the ability of students to identify and overcome writing challenges varies with their English proficiency. Mojica (2010) found that low proficiency learners could mainly report some basic difficulties with vocabulary and grammar. On the other hand, a study by Derakhshan and Karimian Shirejini (2020) involving intermediate or higher-level learners showed that advanced students expressed concerns over a wider range of issues, including grammar, spelling, punctuation, coherence, organization, word choice, and rhetorical structures. Therefore, it is vital to explore the specific writing challenges faced by students at different proficiency levels and how they utilize support to overcome these challenges.

AI Chatbots in EFL Learning

AI Chatbots and Language Learning Enhancement. Social Constructivism, developed by Vygotsky (1978), emphasizes that learning is an active process in which learners construct knowledge through interaction and experience. This theory aligns with the capabilities of chatbots, which facilitate real-time interaction, provide immediate feedback, and offer guidance throughout the language learning process (Guo et al., 2024; Kuhail et al., 2023). Several studies have illustrated the promise of chatbots for enhancing specific language skills. For example, Kim (2018) demonstrated that Elbot, a chatbot, improved Korean college students' listening and reading skills by delivering authentic written and spoken language input. Similarly, chatbots are noted for their ability to provide immediate guidance, personalized feedback, and automated assessments, which can reduce learners' cognitive load (Feng, 2024) and alleviate teachers' workloads (Huang et al., 2024). Beyond individual skills, Chen and Lin (2023) reported significant improvements in grammar, while Xu et al. (2021) identified chatbots as effective reading partners capable of providing adaptive support and thought-provoking feedback. Additionally, Jeon (2021) highlighted how chatbots employing graduated prompts enhanced vocabulary learning and tracked learners' progress. Overall, these studies suggest that chatbots, when thoughtfully implemented, can foster an active learning environment and enhance EFL learners' outcomes.

Despite these promising findings, their application across different language skills and learner contexts remains uneven and underexplored. Most research has prioritized skills such as speaking, reading, and vocabulary acquisition, while writing—a skill that poses unique challenges—has received comparatively little attention. This lack of focus underscores the need for more targeted studies to explore how chatbots can enhance writing proficiency and whether their capabilities align with the nuanced requirements of this skill.

AI Chatbots and Learners' Perceptions. Learner perceptions of chatbots have revealed complex and sometimes contradictory findings. For example, Mahmoud (2022) explored learner satisfaction with chatbots, revealing positive experiences and perceptions of these tools as being beneficial and user-friendly. The study underscored chatbots' significant contribution to enhancing speaking skills and overall English language proficiency, leading to increased satisfaction in speaking classes. Similarly, Kohnke (2023) found that EFL students valued chatbots for their timely guidance and feedback during language practice. However, these positive views varied by learner proficiency. Yang et al. (2022) noted that higher proficiency learners often perceived chatbots as being less effective, while Fryer and Carpenter (2006) found them to be beneficial for informal conversational practice among advanced learners. In contrast, Yin and Satar (2020) observed greater benefits for lower proficiency learners, who reported increased engagement and improved outcomes. Higher proficiency learners, however, expressed dissatisfaction and disengagement, indicating that chatbots may not sufficiently meet their advanced learning needs.

These divergent findings highlight the complex interplay between learner proficiency, chatbot design, and use context. While chatbots have been praised for their adaptability and immediate feedback, their impact appears highly dependent on how they are integrated into learning environments and on learners' characteristics. Moreover, little is known about how learners engage with chatbots in tasks that demand higher-order thinking, such as writing. This gap underscores the need for further research to explore the full potential and limitations of chatbots in fostering language learning.

AI Chatbots in the EFL Writing Domain. Early studies on chatbots in writing instruction highlighted their potential benefits while also revealing limitations. Lin and Chang (2020) introduced a chatbot for a two-week writing tutorial, which improved learners' argumentative skills and peer review feedback. However, the absence of a pretest left learners' initial writing abilities unmeasured. Similarly, Guo et al. (2022) found that chatbots supported idea generation and counterargument integration. Notably, both studies utilized retrieval-based chatbots that provided predefined responses, limiting their adaptability to contextualized student needs. These findings underscore the need for further exploration of generative chatbots, which create responses dynamically based on the conversation context.

The advent of ChatGPT in late 2022 marked a turning point in investigating generative chatbots for writing instruction. Yan (2023) conducted a one-week practicum in a Chinese university, demonstrating that collaborative and reflective practices with ChatGPT significantly enhanced students' writing proficiency. Students appreciated ChatGPT's speed and high-quality feedback. Similarly, Su et al. (2023) integrated ChatGPT into writing classes to address concerns related to language, content, and organization, finding it effective in terms of supporting structural and linguistic aspects of argumentative writing. In a qualitative study, Tseng and Lin (2024) analyzed students' written works and reflections using ChatGPT-3.5, finding that it enhanced writing efficiency, improved cohesion, and substituted peer reviewers by providing critical, objective feedback. However, the study's small sample size and lack of quantitative data limited its generalizability. Addressing this gap, Boudouaia et al. (2024) conducted a quantitative study with Algerian EFL undergraduates, focusing on the effects of ChatGPT-4. Students used ChatGPT primarily for feedback during a structured writing and revising process. The findings indicated significant improvements in task achievement, logical flow, grammatical accuracy, and vocabulary range, suggesting ChatGPT's potential to address diverse writing challenges.

While these studies demonstrate the promise of AI chatbots in addressing writing difficulties, most were conducted in university settings with short interventions lasting one to five weeks. Longer-term implementations are necessary to evaluate sustained performance, and future research should also investigate how students actively construct knowledge in these contexts. This study aimed to fill these gaps by examining high school students' usage patterns, writing performance, and perceptions across different proficiency levels. In line with these objectives, the following research questions were formulated:

1. How do students at different writing levels use the AI chatbot for writing practice?
2. How does the integration of the AI chatbot into EFL writing practice affect the writing performance of students at different writing levels?
3. Is there any difference in the higher- and lower-level EFL students' perceptions of the chatbot?

Methodology

Research Design

This study employed a quasi-experimental design featuring two groups divided according to their writing proficiency levels to enable outcome comparisons between them. A mixed-method approach was utilized, incorporating quantitative data such as timed writing scores, chat logs, and questionnaires, along with qualitative data from open-ended responses and semi-structured interviews. This approach facilitated the collection and analysis of both types of data, aiming to provide a comprehensive

understanding of the practical situation under investigation (Creswell et al., 2006). By integrating quantitative and qualitative methods, the study aimed to triangulate findings, thereby enhancing the validity and reliability of the results.

Participants

The study recruited 47 11th-grade students through convenience sampling, all 16 years old, from a high school in Hanoi, Vietnam. Although all students had access to mobile phones and computers at home, none had prior experience of using AI chatbots for learning.

Based on IELTS writing standardized pretest scores, the participants were categorized into two distinct levels:

- A higher-level group: 25 students with scores ranging from 6.0-7.0 on the IELTS scale. They were categorized as competent and good users according to IELTS band descriptors.
- A lower-level group: 22 students with scores ranging from 4.0-5.0. According to the IELTS band descriptors, these students were described as modest and limited users.

The participants were instructed by the same teacher during class sessions. Before signing the consent form, participants were briefed on the research's objectives, procedures, benefits, and potential risks.

AI Chatbot Design

To help students practice writing at home, the AI chatbot, called the Writing Assistant Bot (WAB), was built on the advanced GPT-3.5 Turbo language model and was capable of generating human-like responses. Trained on extensive textual information, it internalized patterns, grammar, semantics, and general knowledge. The WAB online platform has three main functions. First, it provided students with a weekly writing topic, accessible upon logging in (Figure 2). Second, students could compose and submit essays using the "Write" function (Figure 3). Of particular importance, the study focused on the third pivotal function of the platform, which involved the utilization of the AI chatbot to aid students throughout their writing process. The chatbot engaged in dynamic conversations with students, answering questions, providing explanations, offering suggestions, and assisting with various writing tasks (Figure 4).

Experimental Procedure

As shown in Figure 5, the study spanned 11 weeks from November 2023 to January 2024. In the first week, students took a pretest for classification into two groups and

Figure 2. Login screen.

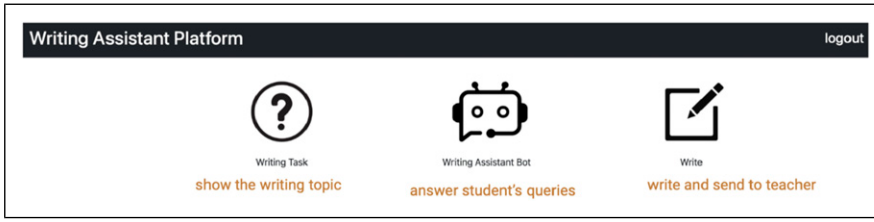

Figure 3. Platform interface.

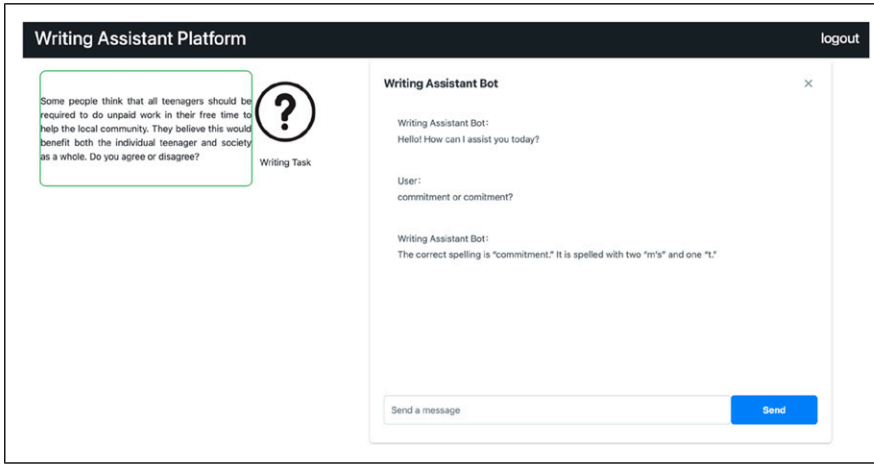

Figure 4. A conversation between a user and the chatbot.

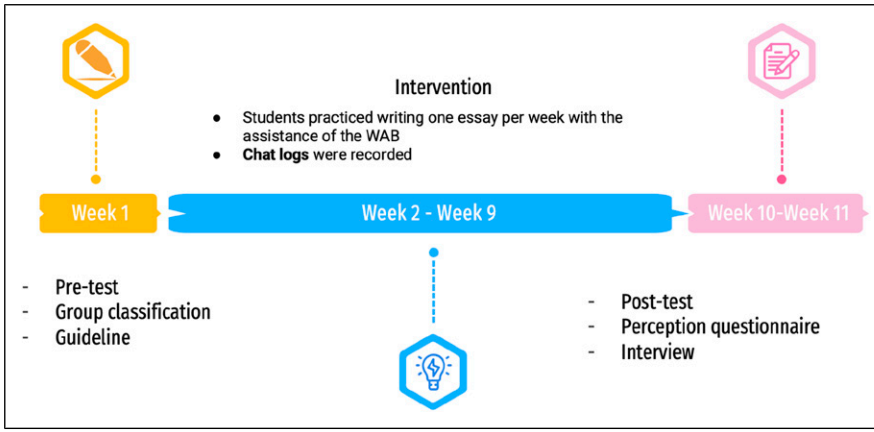

Figure 5. Procedure.

received guidance on using the WAB. From weeks 2–9, they practiced writing essays with chatbot assistance. Each week, students were assigned the task of writing one essay. The interactions with the AI chatbot were documented through the recording of chat logs, which captured the manner in which students made use of the chatbot. In week 10, a posttest and a perception questionnaire were administered. Week 11 featured semi-structured interviews to gain in-depth insights into the participants' learning experiences and perceptions.

Data Collection Tools

Chat Logs. The chat logs were recorded to understand how students utilized the AI chatbot for writing practice, capturing total chatting time and meaningful inquiries related to weekly writing tasks. These inquiries were analyzed using content analysis with two coding schemes. The first, by Flower and Hayes (1981), identified stages of the writing process - Planning, Translating, and Reviewing (see Appendix E Table 1) where students engaged with the chatbot. The second, by Jacobs (1981), categorized writing aspects: Content, Organization, Vocabulary, Language Use, and Mechanics (see Appendix E Table 2). The first author and an expert conducted the analysis, achieving a high inter-rater reliability score of 0.93 using Cohen's Kappa.

Timed-Writing Tests. To evaluate the writing performance before and after the intervention, two 40-min writing tasks based on IELTS Writing Task 2 were used as the pretest and posttest. Scores were given by the instructor and another experienced teacher using Jacobs's (1981) rubric, which assesses mechanics, language use, vocabulary, organization, and content. This rubric has been widely recognized as standardized criteria for scoring writing. The Cohen's Kappa test showed scores of

0.83 for the pretest and 0.87 for the posttest, indicating high agreement between evaluators.

Perception Questionnaires. A 7-point Likert scale questionnaire, based on Davis's (1989) Technology Acceptance Model, was utilized to explore students' perceptions of the AI chatbot, focusing on perceived usefulness and ease-of-use. "Perceived usefulness" measures the belief in the system's ability to enhance performance, while "perceived ease-of-use" evaluates its effortlessness. The questionnaire included 10 items on usefulness (e.g., "Using the AI chatbot in writing practice can help me accomplish the task more quickly"), six items on ease-of-use (e.g., "I find it easy to get the AI chatbot to do what I want it to do"), and two open-ended questions developed specifically for this study. The questionnaire showed a high level of internal consistency with a Cronbach's alpha of .92.

Interviews. To investigate students' learning experiences with the WAB, semi-structured interviews were conducted. According to Lynch (1996), semi-structured interviews have been advocated for their capacity to improve data comprehension, streamline findings, and render data collection more methodical when contrasted with informal conversational methods. Five students from each group were selected randomly to participate in the interview.

Data Analysis. The quantitative data in this study were analyzed using SPSS 21.0. We employed a non-parametric statistical analysis due to the small sample size. Furthermore, the Shapiro-Wilk test showed that our data had a non-normal distribution. Usage tendencies were examined using the Mann-Whitney U test and descriptive statistics, while timed-writing tests were assessed using the Wilcoxon signed-rank test to determine the effectiveness of utilization within each group. Perception questionnaires were analyzed using the Mann-Whitney U-test to compare differences between the two groups.

For qualitative analysis, the responses were analyzed using Excel and NVivo 14. Responses to open-ended questions were meticulously analyzed and coded to identify themes regarding the participants' perceptions of the AI chatbot. Interviews were recorded, transcribed, translated into English, and then analyzed through repeated readings to identify common patterns and divergences, following the procedures outlined by Miles and Huberman (1994). The results underwent review by a qualitative research expert to enhance the credibility and reliability of the analysis.

Ethical Considerations. This study adhered to strict ethical standards to protect participants' rights and well-being. Informed consent was obtained through a signed form detailing the study's objectives, procedures, benefits, and potential risks (see Appendix D). Participants were informed that the chatbot might occasionally make errors, potentially causing minor frustration or confusion; however, these risks were deemed manageable. Participation was entirely voluntary, and participants could withdraw at

any time without penalty. Anonymity was ensured using pseudonyms, and all data, including tests, questionnaires, chat histories, and recordings, was securely stored in password-protected files accessible only to the research team.

Results

Chat Log Analysis

To assess student usage of the chatbot during the intervention, we first focused on analyzing the total chatting time and the number of meaningful inquiries. As presented in Table 1, the Mann-Whitney U test revealed no significant difference between the two groups in terms of total chatting time (U = 248.50, z = −.565, p = .572, r = .08) or meaningful inquiries (U = 220.00, z = −1.173, p = .241, r = .17). These findings suggest that both groups engaged similarly when interacting with the AI chatbot.

We then conducted content analysis on the meaningful inquiries, with the results presented in Table 2 and Table 3.

Regarding writing stages, lower-level students asked the most questions during the Planning stage, averaging 36.1 inquiries, while higher-level students asked the most during the Translating stage, averaging 25.9 questions. During the Planning stage (see Appendix F Figure 1), most questions were about generating vocabulary (P2), with higher-level students averaging 12.9 questions and lower-level students 26.6. In the Translating stage (see Appendix F Figure 2), questions focused on refining ideas and expressions (T4), with higher-level students averaging 19.4 questions and lower-level students 10.8. In the Reviewing stage (see Appendix F Figure 3), higher-level students posed 11.5 questions on average, compared to 9.9 for lower-level students. Specifically, both groups primarily sought feedback (R6), with higher-level students asking 5.2 questions and lower-level students 4.3.

Both groups showed similar inquiry levels in writing aspects, particularly focusing on Vocabulary and Content. The higher-level group asked 32.5 Vocabulary questions, emphasizing avoiding repetition (V2), while the lower-level group asked 37 questions, focusing on vocabulary translation (V4) (see Appendix F Figure 6). For Content, both groups asked nearly identical numbers of questions: 21.3 for the higher-level group and 21.5 for the lower-level group, with an emphasis on generating ideas (C1) (see Appendix F Figure 8). Additionally, Organization and Language Use received

Table 1. Mann-Whitney U Test Results of Total Chatting Time and Meaningful Inquiries.

Variable	Group	*N*	Mean-Rank	Sum of Ranks	U	z	*r*
Total chatting time	Higher-level	25	22.94	573.50	248.50	−.565	0.08
	Lower-level	22	25.20	554.50			
Meaningful inquiries	Higher-level	25	21.80	545.00	220.00	−1.173	0.17
	Lower-level	22	26.50	583.00			

Table 2. Average Inquiries Regarding Writing Stages in the 8 Week-Intervention.

Stage	Higher-level group	Lower-level group
P1	7.6	8.2
P2	12.9	26.6
P3	0.7	1.4
P4	1.8	1.6
Planning	22.9	36.1
T1	1.0	1.1
T2	5.2	4.5
T3	0.3	0.1
T4	19.4	10.8
Translating	25.9	16.5
R1	1.2	2.2
R2	1.0	0.6
R3	2.0	1.8
R4	1.1	0.7
R5	0.3	0.2
R6	5.2	4.3
R7	0.8	1.0
Reviewing	11.5	9.9

considerable interest, with the higher-level group asking 7.7 Organization questions compared to the lower-level group's 6.2 (see Appendix F Figure 7). In Language Use, the higher-level group asked 5.3 questions, while the lower-level group asked 6.2 (see Appendix F Figure 5). Mechanics received the least focus, with the higher-level group averaging 1.0 question and the lower-level group 1.3 (see Appendix F Figure 4).

Analysis of Timed-Writing Test Scores

To assess differences in writing performance after the intervention, we utilized the Wilcoxon signed-rank test.

The results presented in Table 4 indicate significant differences between the pre- and posttest scores for both groups across all assessed aspects. Particularly, posttest scores were notably higher than pretest scores in Mechanics (Lower-level: $z = -3.500$, $p = .000$; Higher-level: $z = -2.236$, $p = .025$), Language Use (Lower-level: $z = -2.959$, $p = .003$; Higher-level: $z = -2.627$, $p = .009$), Vocabulary (Lower-level: $z = -3.505$, $p = .000$; Higher-level: $z = -3.611$, $p = .000$), Organization (Lower-level: $z = -3.075$, $p = .002$; Higher-level: $z = -3.183$, $p = .001$), and Content (Lower-level: $z = -3.685$, $p = .000$; Higher-level: $z = -3.461$, $p = .001$). Additionally,

Table 3. Average Inquiries Regarding Writing Aspects in the 8 Week-Intervention.

Aspect	Higher- level group	Lower-level group
Mechanics	1.0	1.3
L1	2.9	2.8
L2	0.3	0.5
L3	2.1	2.9
Language use	5.3	6.2
V1	1.0	0.3
V2	17.4	7.8
V3	5.4	6.5
V4	6.3	19.8
V5	2.5	2.5
Vocabulary	32.5	37
O1	1.5	1.6
O2	0.5	0.3
O3	5.7	4.3
Organization	7.7	6.2
C1	8.7	10.3
C2	5.3	5.1
C3	7.3	6.1
Content	21.3	21.5

overall writing performance significantly improved for both groups (Lower-level: $z = -3.990$, $p = .000$; Higher-level: $z = -4.162$, $p = .000$). These findings underscore the positive impact of the chatbot on writing performance across different proficiency levels.

Analysis of Perception Questionnaires

The Mann-Whitney U test results showed no significant difference between the two groups regarding Perceived Usefulness, with $U = 228.50$, $z = -.705$, $p = .481$, and $r = .10$. However, for the Perceived Ease-of-Use scale, the higher-level group (Mdn = 6.00) exhibited a significantly higher rate compared to the lower-level group (Mdn = 5.50), with $U = 237.00$, $z = -2.061$, $p = .039$, and $r = .30$, as shown in Table 5. This indicates that while both groups perceived the tool to be equally useful, the higher-level group found it significantly easier to use compared to the lower-level group.

Analysis of Open-Ended Responses

Responses to the two open-ended items were analyzed using content analysis (Table 6). The most favored features across both groups were prompt generation and idea

Table 4. Wilcoxon Signed-Rank Test Results for Writing Performance of the Two Groups.

Criteria	Group	Writing Performance	*N*	Mean	*SD*	Wilcoxon W	Z	*p*
Mechanics	Lower-level	Pretest	22	4.09	.61	638.00	−3.500	.000
		Posttest	22	4.73	.46			
	Higher-level	Pretest	25	4.76	.44	490.00	−2.236	.025
		Posttest	25	4.96	.20			
Language use	Lower-level	Pretest	22	17.68	.89	575.50	−2.959	.003
		Posttest	22	18.82	1.10			
	Higher-level	Pretest	25	20.96	1.17	552.50	−2.627	.009
		Posttest	25	21.60	.96			
Vocabulary	Lower-level	Pretest	22	13.91	.97	619.50	−3.505	.000
		Posttest	22	15.41	.85			
	Higher-level	Pretest	25	16.20	.65	508.50	−3.611	.000
		Posttest	25	17.20	.82			
Organization	Lower-level	Pretest	22	13.73	1.42	555.50	−3.075	.002
		Posttest	22	15.14	1.28			
	Higher-level	Pretest	25	16.16	.99	572.50	−3.183	.001
		Posttest	25	17.24	.93			
Content	Lower-level	Pretest	22	21.59	1.71	644.50	−3.685	.000
		Posttest	22	24.00	1.51			
	Higher-level	Pretest	25	24.76	1.01	483.50	−3.461	.001
		Posttest	25	25.84	1.03			
Overall	Lower-level	Pretest	22	71.00	3.65	681.00	−3.990	.000
		Posttest	22	78.09	3.80			
	Higher-level	Pretest	25	82.84	3.08	447.00	−4.162	.000
		Posttest	25	86.84	2.85			

Table 5. Mann-Whitney U Test Results for Perception Questionnaires Between the Two Groups.

Scale	Group	*N*	Mean-rank	Sum of ranks	U	z	*r*
Perceived usefulness	Higher-level	25	25.32	633.00	228.50	−.705	0.10
	Lower-level	22	22.50	495.00			
Perceived ease-of-use	Higher-level	25	27.84	696.00	237.00	−2.061*	0.30
	Lower-level	22	19.64	432.00			

*$p < .05$.

inspiration, interactive feedback, and vocabulary enhancement. One student noted, "It can answer my questions and suggest ideas to enrich essays." Another said, "It helps me extend my ideas with deeper explanations and examples." They found the chatbot's responses "super helpful and practical." Regarding the interactive feedback, students

Table 6. Content Analysis Results of the Open-Ended Items.

Theme	Sub-theme	Description	Higher-level group	Lower-level group
Favored features of the AI chatbot	Prompt generation and idea inspiration	Stimulate creativity and provide diverse prompts	18	13
	Interactive feedback	Provide real-time, constructive feedback	7	7
	Vocabulary enhancement	Expand user's lexicon with new words and phrases	9	4
Things to improve	Technical issues	Problems such as lagging or logging issues	6	6
	Not updated new information	Not reflecting the latest news or information	3	0
	Content presentation	Lengthy responses with difficult words	0	2
	Chat history retention issues	Not able to save chat history for each writing task	1	1

noted the chatbot's efficiency in providing detailed comments and essay correction compared to waiting for a teacher's feedback. Additionally, students valued the chatbot's comprehensive explanations of new vocabulary, helping them understand and apply words correctly in context. They used the chatbot for translation, finding topic-related vocabulary, and expanding their lexical resources with high-level words and synonyms.

Regarding areas for improvement, the most commonly reported issue across both groups was technical problems, such as lagging or login issues and history retention. Challenges related to lengthy responses with the use of difficult words were primarily reported by the lower-level group. Meanwhile, the content not being regularly updated with the latest news was a more significant concern for the higher-level group.

Analysis of Interviews

At the end of the intervention, five students from the higher-level group coded as H1, H2, H3, H4, and H5, and five students from the lower-level group coded as L1, L2, L3, L4, and L5 were invited to participate in the interview. After analysis, three themes emerged from the interviews, namely: Usage and strategies, Perceived improvement, and Perceptions of the chatbot.

Regarding "Usage and strategies," students employed the chatbot in various ways for assistance, with idea generation, vocabulary enhancement, and idea development being frequently mentioned, aligning with their responses to the open-ended questions. L4 mentioned, "I often use the chatbot when I struggle to generate ideas for writing, especially on unfamiliar topics." H2 added, "The chatbot generates precise and diverse ideas, helping me expand my vocabulary by suggesting synonyms and related words." Both groups used the chatbot to enrich their viewpoints with examples or explanations. H4 also noted, "When I sift through ideas on the chatbot, it is not just one main line of thought; there are supporting points underneath. So, I understand that my essays can use such ideas." This indicates that students not only exploited the rich content, but also learned from the way the chatbot presents them.

Students also described varied approaches to integrating the chatbot into their writing process. L3 primarily used it during the Planning stage. H2 utilized it in both the Planning and Translating stages, exploring different approaches, combining suggestions with personal ideas, and enhancing vocabulary. Conversely, H4 restricted the usage to Planning and Reviewing to maintain focus during writing, while H5 consulted the chatbot for ideas before writing, referenced its sample essays afterwards to refine vocabulary, and checked for errors to ensure organizational coherence.

H3 expressed concerns about overreliance on the chatbot, noting that it hindered her ability to write independently, "I found it quite difficult to write an essay on my own afterwards, as I had been relying heavily on the chatbot for ideas and vocabulary… This reliance resembled copying too closely, prompting me to reduce my chatbot usage to find a better balance." Meanwhile, L2 shared a different perspective, emphasizing the importance of generating original ideas before consulting the chatbot: "I ensure my ideas are formulated first, and often, they are similar to [those of] the chatbot. However, the chatbot's ideas are more advanced and deeper, indicating a complementary rather than dependent relationship." L1 added, "The chatbot does not provide complete essays; it acts more as a guide, stimulating us to complete our own work and independently develop our writing skills."

In the theme of "Perceived improvement," all interviewees reported advancements in their writing, particularly in terms of content, vocabulary, coherence, and structure. H2 described her experience: "Before using the chatbot, my vocabulary was limited, and I often repeated words. The chatbot expanded my vocabulary and taught me new ways to structure arguments and use words correctly." H4 echoed this sentiment, stating, "Using the chatbot helped me explore different perspectives, improving my thought process and idea generation. Additionally, my essay structure, including how I present supporting ideas, became clearer and less convoluted." L2 added, "Before using the chatbot, my ideas were limited, and my writing lacked flow; paragraphs were disconnected. After using the chatbot, my ideas became broader and clearer in each section."

Under the theme "Perceptions of the chatbot," most students claimed that the chatbot featured a user-friendly interface, was easy to communicate with, and did not require high language proficiency for interaction. H4 shared, "I just need to use simple commands, and the AI can understand and fulfill my requests, making using this

chatbot very straightforward." L5 found the chatbot's responses "very easy to understand because the chatbot breaks it down into different parts, making it quite clear." Additionally, they all agreed that basic English proficiency was sufficient for effective communication. H1 mentioned, "Even with simple words, the chatbot understands." L1 added, "The chatbot can understand multiple languages, removing any language barriers."

The interviewees highlighted that the chatbot was notably helpful, attributing its effectiveness to several key features: accessibility, broad knowledge base with diverse ideas, real-time assistance, and its overall versatility. H3 and L1 remarked on the chatbot's flexibility, noting that it can be used anywhere and assistance can be requested at any time. H2 remarked on its extensive knowledge, describing it as "seemingly unlimited." H5 also pointed out that the chatbot could suggest numerous ideas, allowing him to find ones that closely aligned with his personal thoughts. H4 shared that "Before having the chatbot, I felt almost like I was lacking support from somewhere. In crowded classes, despite teachers' efforts, I had to manage many small steps alone, so having a chatbot makes me more confident as if I have someone to support me step by step…" H3 added the chatbot's advantage in providing faster feedback than teachers could, enhancing the learning experience. Regarding the chatbot's versatility, the interviewees noted that it fulfilled a range of functions, from brainstorming to reviewing. They utilized it for consultation, expanding or refining vocabulary and arguments, as well as for outlining, structuring, and linking their essays. The ability to perform all these tasks within a single chatbot sets it apart from other sources of assistance.

The interviewees also reported usage difficulties and provided suggestions. The first issue was giving commands. H4 said, "At first, I did not know how to give correct commands to get exactly what I needed… Then I learned to specify my requirements for a suitable sample essay." L1 added, "Commands need to be detailed and objective." Three lower-level learners had trouble understanding the chatbot's responses. L2 mentioned, "The chatbot provided so many ideas that I got confused… I felt overwhelmed by the number of ideas and difficult vocabulary." They addressed this by reviewing suggestions and translating unfamiliar words into Vietnamese. Additionally, three interviewees noted lagging response times and occasional login issues. To enhance user experience, H4 suggested the chatbot cite information sources, and L1 wanted the chatbot to retain chat history after logging out.

Discussion

Usage Tendencies of Students at Different Writing Levels

In response to Research Question 1, which investigates how students at different writing proficiency levels engage with the AI chatbot, the findings revealed that both groups actively engaged with the chatbot for writing assistance. Students l leveraged its effective features such as prompt idea generation and development, as supported by Guo et al. (2022). Additionally, EFL learners extensively utilized the chatbot for

vocabulary enhancement, interactive feedback, and structured writing guidance. Analysis of chat logs and qualitative data revealed diverse needs among students at different proficiency levels, primarily focusing on vocabulary enhancement and content development. This aligns with the findings of Alisha et al. (2019) and Amalia et al. (2021) who highlighted challenges in lexical resources and content generation. Interestingly, while previous studies emphasized language use as a significant hurdle for EFL students (Alisha et al., 2019; Nguyen et al., 2021), this study found that language use was not the primary focus of student interactions with the chatbot. The reasons could be that learners in this study had attained a level of comfort with basic language use through traditional methods, viewing the chatbot as an opportunity to address more advanced challenges in vocabulary and content development where they felt less confident and lacked support elsewhere.

It was also noted that even when students used the chatbot for the same aspects, the way they utilized it was different. Lower-level learners used the chatbot to bridge the gap between their current understanding and the language they aspired to master, aiming to grasp meanings of words. On the other hand, higher-level learners sought to refine their language skills, using the chatbot to find synonyms to express themselves more precisely and diversely. This demonstrates the chatbot's dual role not just as a facilitator of basic language learning, but also as an advanced tool for linguistic enhancement and expansion.

Regarding stages of writing, students in the lower-level group predominantly used the chatbot during the Planning stage, followed by the Translating stage. This suggests that lower-level learners lean on the chatbot as a preparatory tool, helping them to organize their thoughts and translate their ideas into the target language. This aligns with Mojica's (2010) study, which found that beginner English learners engaged deeply in thinking before writing, with decreased engagement during writing and the least in post-writing. On the other hand, the higher-level group exhibited a more balanced use across all three stages of writing, with the most use during the Translating stage. This indicates that higher-level learners use the chatbot as a real-time assistant during the writing process, leveraging it for immediate feedback, language refinement, and vocabulary enhancement.

Effects of the AI Chatbot on Writing Performance

Addressing Research Question 2, which explores the impact of the AI chatbot on students' writing performance, the results underscore the AI chatbot's effectiveness in terms of enhancing students' writing performance, with improvements observed across all writing aspects, supporting the study of Boudouaia et al. (2024). Functioning as a real-time tutor, the chatbot offered constant assistance throughout various stages of the writing process, thereby facilitating students' writing practice and enhancing their writing skills. This is consistent with the concept of the Zone of Proximal Development (Vygotsky, 1978) which suggests that learners perform tasks better under the guidance of a more knowledgeable other. Chatbots can act as the "more knowledgeable other" in

this case by providing students with guidance and support throughout the writing process.

As reported in the timed-writing test scores and interviews, students at both levels demonstrated more relevant and diverse ideas with better supporting arguments. Additionally, lower-level learners used a wider and more accurate range of vocabulary, while higher-level learners demonstrated more sophisticated lexical resources, which aligns with their focus in chatbot usage. The results also showed fewer grammar mistakes and more complex sentence structures. Furthermore, students could organize their essays more logically and cohesively. Spelling and punctuation errors also decreased significantly. This wide range of enhancements facilitated by the chatbot indicates that it is a versatile tool, capable of addressing diverse learning needs and advancing students' writing abilities in a holistic manner.

In this study, lower-proficiency students, consistent with Yin and Satar's (2020) findings, significantly improved their writing with chatbot support. By offering real-time assistance with prompts and vocabulary suggestions, the chatbot helped low-proficiency learners overcome language barriers. Interestingly, unlike Yin and Satar's results, higher-level students in our study also benefited from the chatbot. This discrepancy may have stemmed from differences in chatbot design: while their study used retrieval-based and generative chatbots deemed inadequate by higher-level students, ours used a generative chatbot enhanced with Natural Language Processing technologies. This allowed for a more responsive learning experience tailored to each learner's needs, including advanced requirements. These findings suggest that the effectiveness of chatbots in terms of enhancing writing abilities may depend on the specific system designs tailored to the target audience's proficiency levels.

User Experience and Perceptions of the AI Chatbot

Regarding Research Question 3, which examines students' experiences and perceptions of the AI chatbot, the findings highlight that students in both the higher- and lower-level groups expressed positive perceptions of the chatbot's usefulness and ease of use. They appreciated its accessibility, versatility, and ability to provide tailored assistance across various aspects and stages of writing, which facilitated their improvement. The chatbot's broad knowledge base and responsive, bilingual capabilities were highlighted as beneficial features, consistent with the findings of Huang et al. (2022).

The chatbot's simple and intuitive interface was also praised. Students found it easy to interact with and understand the chatbot's responses, although some lower-level students initially struggled with giving commands and understanding complex responses. Qualitative data indicated that lower-level learners faced challenges in rephrasing questions for clarity and comprehending the chatbot's replies, potentially influencing their slightly lower ratings of flexibility and simplicity compared to their higher-level peers. This shows the need to use simpler language when responding to

lower-level learners, and to offer step-by-step guidance to help them feel less overwhelmed.

Conclusion, Implications and Limitations

The study highlights the chatbot's versatility as an assistance tool, capable of addressing the varied needs of learners at different proficiency levels. For lower-level learners, it is used to overcome language barriers and assist in planning and generating essays. For higher-level learners, it serves as a resource for language refinement and writing improvement. The chatbot significantly enhances students' writing performance, and students recognize it as a useful and user-friendly learning aid.

The study results have several implications for educators, students and AI developers. For educators, chatbot usage provides educators insights into common writing challenges. Higher-level learners need the most support during the Translating stage, while lower-level learners mostly need help in the Planning phase. Both groups focused on content development and vocabulary, suggesting the need for more targeted classroom instruction. Educators should also consider incorporating chatbots as supplementary tools for language practice. In this study, some students experienced difficulty formulating inquiries; therefore, teachers should provide explicit guidance on crafting clear and effective prompts to maximize the chatbot's utility. By teaching students to frame questions and instructions with specific keywords and relevant details, teachers can help ensure that the chatbot provides more accurate and precise responses. For students, chatbots can facilitate writing practice and help overcome obstacles. Students can employ the chatbot for specific writing tasks, such as essay planning, translating, and revising. For instance, during the Planning stage, users can brainstorm ideas and outline their thoughts with the chatbot's help. In the Translating stage, the chatbot can assist in refining their language and connecting ideas. During the Reviewing stage, users can leverage the chatbot to identify grammatical errors or revise their arguments. However, to avoid overreliance, they should be used strategically and critically, supplementing personal critical evaluation of suggestions. In this study, students were provided with clear guidance and prompt examples to maximize the chatbot's benefits. Interviews revealed that students used the chatbot selectively, relying on it for specific tasks rather than throughout the entire writing process. For instance, some sought ideas and vocabulary in the Planning stage, others independently drafted outlines and brainstormed before seeking additional input, and some used the chatbot solely for final feedback. Students were also cautioned that the chatbot's responses could contain errors, emphasizing the importance of critically evaluating its suggestions and cross-checking information for accuracy. To ensure effective use of the chatbot, students should be encouraged to critically evaluate its suggestions, incorporate their own ideas, and view it as a supportive tool rather than as a substitute for personal effort and critical thinking. For AI developers, feedback highlights the need for enhanced chatbot adaptability. Responses should be more relevant and tailored to align with the learner's proficiency and needs. This could involve developing

algorithms that better assess the user's language level and adjust the complexity of feedback accordingly. For example, beginners might benefit from simplified explanations and basic vocabulary suggestions, potentially delivered in bilingual responses, while advanced learners could receive more nuanced feedback and sophisticated language recommendations. Additionally, qualitative data identified key areas for improvement, such as chat history saving, information sourcing and frequent updates. Many chatbots, including Gemini and ChatGPT, already allow users to save and view chat history, enabling learners to track their progress. To further enhance trustworthiness and user experience, chatbots should be updated with the latest information and include advanced features such as real-time validation of information and source traceability. These improvements would increase reliability and provide learners with easier access to accurate and credible information.

This study has some limitations. First, the sample size and context of this study should be considered. With 47 participants from a single high school in Northern Vietnam, the findings offer valuable insights into a specific group of learners. However, a larger and more diverse sample could provide a broader understanding of students' usage tendencies, writing performance, and perceptions of AI chatbots across various educational contexts. Second, the absence of a control group presents a limitation in isolating the exact effects of the WAB. Although the inclusion of rich qualitative data added depth to the analysis, incorporating a control group in future research would help validate these findings and distinguish the chatbot's impact from other factors, such as natural writing development or additional teacher or peer support. Third, the study primarily focused on the short-term effects of the WAB over an 11-week period. While this duration is longer than that of many similar studies, extending the intervention in future research could uncover insights into the long-term sustainability of observed improvements and whether students maintain engagement with the chatbot over time. Finally, the research exclusively examined the WAB and did not compare it to other widely used writing support tools, such as grammar checkers or online tutoring platforms. A comparative analysis could provide valuable insights into the WAB's relative effectiveness, user satisfaction, and unique contributions to EFL writing support, informing future development efforts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is supported by National Science and Technology Council, Taiwan (NSTC 113-2410-H-011-003-MY3) and The "Empower Vocational Education Research Center" of National Taiwan University of Science and Technology (NTUST) from the

Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

ORCID iD

Hsiu-Ling Chen https://orcid.org/0000-0002-8951-7043

Supplemental Material

Supplemental material for this article is available online.

References

Ali, S. S., & Ramana, V. L. (2018). Academic writing challenges at Universities in Saudi Arabia and solutions. *International Journal of English Language and Humanities*, *4*(10), 291–298.

Alisha, F., Safitri, N., Santoso, I., & Siliwangi, I. (2019). Students' difficulties in writing EFL. *Professional Journal of English Education*, *2*(1), 20–25.

Amalia, H., Abdullah, F., & Fatimah, A. S. (2021). Teaching writing to junior high school students: A focus on challenges and solutions. *Journal of Language and Linguistic Studies*, *17*(S2), 794–810. https://doi.org/10.17263/jlls.904066

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a Foreign Language writing learning. *Journal of Educational Computing Research*, *62*(6), 1509–1537. https://doi.org/10.1177/07356331241247465

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, M. R. A., & Lin, Y. H. (2023). AI chatbot-supported grammar tense learning on EFL students learning effectiveness. In *INTED2023 proceedings* (pp. 1207–1216). IATED.

Creswell, J. W., Shope, R., Plano Clark, V. L., & Green, D. O. (2006). How interpretive qualitative research extends mixed methods research. *Research in the Schools*, *13*(1), 1–11.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Derakhshan, A., & Karimian Shirejini, R. (2020). An investigation of the Iranian EFL learners' perceptions towards the most common writing problems. *Sage Open*, *10*(2), Article 215824402091952. https://doi.org/10.1177/2158244020919523

Feng, L. (2024). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. *Journal of Educational Computing Research*, *62*(8), 1961–1994. https://doi.org/10.1177/07356331241268349

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition & Communication*, *32*(4), 365. https://doi.org/10.2307/356600

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. Language learning and technology. *Language, Learning and Technology*, *10*(3), 8–14. https://doi.org/10.125/44068

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is chat GPT a blessing or a curse? *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1166682

Guo, K., Li, Y., Li, Y., & Chu, S. K. W. (2024). Understanding EFL students' chatbot-assisted argumentative writing: An activity theory perspective. *Education and Information Technologies*, *29*(1), 1–20. https://doi.org/10.1007/s10639-023-12230-5

Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, *54*(2), Article 100666. https://doi.org/10.1016/j.asw.2022.100666

Haiyan, M., & Rilong, L. (2016). Classroom EFL writing: The alignment-oriented approach. *English Language Teaching*, *9*(4), 76–82. https://doi.org/10.5539/elt.v9n4p76

Hsu, M. H., Chen, P. S., & Yu, C. S. (2023). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, *31*(7), 4297–4308. https://doi.org/10.1080/10494820.2021.1960864

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they 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

Huang, W., Jia, C., Hew, K. F., & Guo, J. (2024). Using chatbots to support EFL listening decoding skills in a fully online environment. *Language, Learning and Technology*, *28*(2), 62–90. https://hdl.handle.net/10125/73572

Jacobs, H. L. (1981). *Testing ESL composition: A practical approach. English composition program*. Newbury House Publishers, Inc.

Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, *36*(7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272

Kim, N. (2018). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence*, *16*(8), 19–26. https://doi.org/10.1016/j.archger.2018.01.002

Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation*, *17*(1/2), 214–226. https://doi.org/10.1504/ijmlo.2023.10053355

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

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, *23*(1), 78–92. https://www.jstor.org/stable/26915408

Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge University Press.

Mahmoud, R. H. (2022). Implementing AI-based conversational chatbots in EFL speaking classes: An evolutionary perspective. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-1911791/v1

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Sage Publications.

Mojica, L. A. (2010). An investigation on self-reported writing problems and actual writing deficiencies of EFL learners in the beginners' level. *TESOL Journal*, *2*(1), 24–38. https://tesol-international-journal.com/wp-content/uploads/2013/11/A3V2_TESOL.pdf

Nguyen, D. K, Nguyen, T. H. U, & Phan, T. T. V (2021). Common errors in writing of EFL sophomores in a context of the Mekong delta. *Eurasian Journal of Applied Linguistics*, *7*(2), 46–57.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.

Şenel, E. (2018). The integration of creative writing into academic writing skills in EFL classes. *International Journal of Languages' Education and Teaching*, *6*(2), 115–120. https://doi.org/10.18298/ijlet.2869

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, Article 100752. https://doi.org/10.1016/j.asw.2023.100752

Teng, M. F., Wang, C., & Zhang, L. J. (2022). Assessing self-regulatory writing strategies and their predictive effects on young EFL learners' writing performance. *Assessing Writing*, *51*(100573), Article 100573. https://doi.org/10.1016/j.asw.2021.100573

Tran, L. T. (2007). Learners' motivation and identity in the Vietnamese EFL writing classroom. *English Teaching: Practice and Critique*, *6*(1), 151–163. https://files.eric.ed.gov/fulltext/EJ832183.pdf

Tseng, Y. C., & Lin, Y. H. (2024). Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, *22*(2), 78–97. https://doi.org/10.34190/ejel.21.5.3329

Tuan, L. T. (2010). Enhancing EFL learners' writing skill via journal writing. *English Language Teaching*, *3*(3), 81–88. https://doi.org/10.5539/elt.v3n3p81

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*(1), Article 104059. https://doi.org/10.1016/j.compedu.2020.104059

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28*(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

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

Yin, Q., & Satar, M. (2020). English as a Foreign Language learner interactions with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching*, *7*(2), 390–410. http://files.eric.ed.gov/fulltext/EJ1250583.pdf

Author Biographies

Thi-Ngoc-Anh Duong is a Master's student in the Graduate Institute of Digital Learning and Education at the National Taiwan University of Science and Technology. Her research interests include computer supported language learning and human-computer interaction.

Hsiu-Ling Chen is a Professor in the Graduate Institute of Digital Learning and Education at the National Taiwan University of Science and Technology. She earned her Ph.D. from the University of Texas at Austin. Her main research field is ICT integration in education. Her other interests include flipped classrooms, computer-supported collaborative learning (CSCL), digital storytelling, and higher-order cognitive skills (critical thinking, creative thinking, problem-solving, and metacognition).