



A mixed-methods study on the use of chatgpt in the pre-writing stage: EFL learners' utilization patterns, affective engagement, and writing performance

Long Quoc Nguyen¹ · Ha Van Le¹ · Phuc Thinh Nguyen¹

Received: 16 July 2024 / Accepted: 4 December 2024 / Published online: 18 December 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Research on the use of ChatGPT in education has proliferated in the past years. However, there is limited understanding of how English as a Foreign Language (EFL) learners engage emotionally with and utilize ChatGPT during the pre-writing stage. Furthermore, how learners' affective engagement (emotional responses) with this tool in the planning phase may be correlated with their subsequent writing remains unclear. This study aims to address these gaps by adopting a mixed-methods approach involving 56 first-year university students who completed two argumentative writing tasks, one with and one without ChatGPT support. Data were collected through multiple sources encompassing written texts, questionnaires, ChatGPT logs, note-taking sheets, and semi-structured interviews. Results showed that learners used a broader range of strategies in the prewriting stage and produced better text quality in the subsequent stage in the ChatGPT condition than in the non-ChatGPT mode. Moreover, while learners exhibited positive affective engagement in both conditions, those with ChatGPT assistance demonstrated considerably higher levels of emotional responses. Additionally, a moderate positive correlation was observed between learners' affective engagement with ChatGPT in the pre-writing phase and their overall writing performance. Pedagogically, it is recommended to integrate ChatGPT, with proper training, in the pre-writing stage to enhance both emotional positivity and text quality. Nonetheless, these benefits appear to be short-term, and further research is needed to explore the long-term impact of ChatGPT on the planning and other stages of EFL writing.

Keywords ChatGPT · EFL learners · Pre-writing stage · Utilization patterns · Affective engagement · Writing performance

Extended author information available on the last page of the article

1 Introduction

The advent of Artificial Intelligence (AI) technologies has significantly transformed various domains in education, including second language (L2) learning (e.g., Barrot, 2023; Steiss et al., 2024). Among these tools, ChatGPT (Chat Generative Pre-trained Transformer), a generative AI developed by OpenAI, has emerged as a robust platform that could support writing skills, particularly in English as a Foreign Language (EFL) contexts (e.g., Law, 2024; Tseng & Lin, 2024). Indeed, ChatGPT can engage learners in human-like conversations, provide instant feedback, and assist in various stages of the writing process (e.g., Yuan et al., 2024; Zou & Huang, 2023). This tool is also beneficial in facilitating idea generation, improving text coherence, and offering personalized feedback, which are crucial components for L2 writing development (Adams & Chuah, 2022; Guo et al., 2022).

Despite these affordances, there have been concerns that the use of ChatGPT could lead to dishonesty or even plagiarism in L2 writing (Yan, 2023; Zou & Huang, 2023). To address this issue, teachers and educators can instruct students to use ChatGPT as a feedback instrument for their written output (post-writing stage) because this tool has been shown to provide sufficient feedback quality (Steiss et al., 2024). An alternative is to utilize ChatGPT in the pre-writing stage, one crucial stage in the writing process (Trang & Barrot, 2023), in which students seek essential input (i.e., ideas, vocabulary, grammar, organization) for their drafts. Nonetheless, empirical research on the use of ChatGPT in the pre-writing stage and its impact on learners' subsequent text quality remains scarce, which necessitates further investigations.

Affective engagement, which encompasses learners' emotional responses and attitudes, plays a crucial role in the writing process (Ellis, 2010; Zhang & Hyland, 2018). High levels of affective engagement can lead to reduced anxiety and a more enjoyable learning experience, which are essential for effective language acquisition and writing improvement (Dai & Wang, 2024; Dewaele et al., 2018). Despite the importance of this dimension, there is a paucity of research on how learners engaged emotionally with ChatGPT during writing tasks. A few studies (e.g., Yan, 2023; Zou & Huang, 2023) attempted to explore L2 learners' perceptions of the role of ChatGPT in writing, which revealed both positivity (i.e., high efficiency) and negativity (i.e., plagiarism). Insightful as they are, these studies adopted the qualitative approach and mainly focused on the perceived usefulness of the tool. Quantitative data, especially when compared to that of the non-ChatGPT condition and when combined with qualitative results, would provide a more comprehensive understanding of L2 learners' affective engagement with ChatGPT in writing tasks.

This study aims to address the aforementioned gaps by examining how EFL learners utilize ChatGPT in the pre-writing stage and whether the use of this AI tool has any significant impact on their writing performance. It also seeks to explore learners' affective engagement with ChatGPT and the relationship between this dimension and their text quality.

2 Literature review

2.1 Theoretical framework for the use of ChatGPT in L2 writing

The integration of ChatGPT into the pre-writing stage of L2 learning can be framed within Sociocultural Theory (SCT) (Vygotsky, 1978), which emphasizes the role of social interaction and mediating tools in knowledge development. SCT includes three primary components: activity, mediating tools, and the Zone of Proximal Development (ZPD) (Storch, 2018). In the context of L2 writing, the pre-writing stage represents the activity component, wherein learners engage in generating and organizing ideas. ChatGPT functions as a mediating tool, facilitating this activity by providing scaffolded assistance within the learners' ZPD. Acting as a more competent user of the target language, ChatGPT offers immediate feedback and tailored support, enabling learners to perform tasks they might not manage independently. This process mirrors the benefits of collaborative writing, where co-constructing a text with a partner has been shown to enhance L2 writing development significantly (Lantolf, 2000; Storch, 2019). Furthermore, by interacting with ChatGPT, learners can also use language to reflect on language, termed as *languaging* (Swain, 2006), which could enhance learning.

Overall, the application of ChatGPT in L2 writing aligns with SCT by serving as a mediational tool that supports learners within their ZPD, fosters collaborative-like interactions, and promotes languaging. These elements collectively contribute to more effective and enriched L2 writing development.

2.2 Learners' use of ChatGPT in L2 writing

There is a growing body of research on the use of ChatGPT in writing, but the strategies that students undertake to utilize this tool are less documented. Among these limited studies, Wang (2024) found that learners used ChatGPT for brainstorming, organizing ideas, and addressing both global (e.g., argument, structure, coherence) and local (e.g., syntax, diction, grammar) writing issues. Similarly, Rojas (2024) investigated ChatGPT's application in a scientific writing assignment. The results revealed that students used ChatGPT for idea generation, topic verification, proofreading, and editing. Students reported ease in verifying information generated by ChatGPT despite its lack of access to primary sources. The tool was particularly useful for grammar and punctuation refinement, sentence structure diversification, and vocabulary expansion (Rojas, 2024).

Nonetheless, these studies solely focused on the use of ChatGPT in first language (L1) writing. Research specifically examining L2 learners' utilization of ChatGPT is still scarce. Koltovskaia et al. (2024) attempted to investigate the engagement of L2 graduate students with ChatGPT during the revision phase of academic writing tasks. Their findings indicated that students predominantly utilized the tool to tackle lower-order concerns, including grammar and vocabulary, rather than for high-level conceptual revisions. However, Koltovskaia et al. (2024) only probed into

the post-writing stage (revision); the role of ChatGPT in the pre-writing phase has been largely overlooked, despite the importance of this stage in the writing process (Hayes & Flower, 1980). Thus, further research is needed.

2.3 Learners' affective engagement with ChatGPT in L2 writing

Fredricks et al. (2004) define affective engagement as a state characterized by a strong sense of curiosity, eagerness, and emotional investment in the process of acquiring knowledge. This aspect of engagement is vital as it has an impact on students' inclination to actively participate and continue with learning activities. Research indicates that experiencing pleasant emotions and engagement is associated with enhanced educational achievements, increased motivation, and more perseverance in academic activities (Shute et al., 2015; Skinner et al., 2016). Pekrun's (2006) control-value theory emphasizes the importance of good emotions in educational environments since they improve cognitive resources and ultimately result in improved academic achievement.

Contemporary research has attempted to investigate the influence of AI tools on the learners' emotions and perspectives in L2 writing. In a study, Ali et al. (2023) also reported that these advanced technologies could motivate and enhance learners' writing skills. Similarly, Punar Özçelik and Yangın Ekşi (2024) found that students demonstrated heightened engagement and proactive involvement in writing assignments when utilizing ChatGPT. Such results were corroborated by other scholarship on learners' affective engagement with this AI tool (e.g., Barrot, 2023; Yan, 2023). Overall, it has been consistently documented that ChatGPT could positively promote learners' emotional responses.

Nevertheless, most of these studies focus on learners' use of ChatGPT in the during- and post-writing stages. Research concentrating particularly on the pre-writing stage is still scarce. Exploring this aspect is crucial as it would shed light on how ChatGPT affects L2 learners' affective engagement in various stages of writing.

2.4 The impact of ChatGPT on L2 learners' text quality

Research has underscored the efficacy of ChatGPT in enhancing diverse facets of writing. Zou and Huang (2023) observed notable enhancements in the writing competence of L2 students, namely in the areas of generating ideas, maintaining coherence, and organizing their texts. In their study, Su et al. (2023) found that learners who used ChatGPT had improved grammatical accuracy and a greater variety of vocabulary. ChatGPT's personalized advice and edits are responsible for these benefits, as they improve writing abilities more effectively than conventional techniques. Barrot (2023) corroborates these findings, observing enhanced involvement in writing assignments and superior outcomes in terms of quality while utilizing ChatGPT. In the same vein, Punar Özçelik and Yangın Ekşi (2024) documented that ChatGPT enhanced formal writing components such as syntax, punctuation, and sentence organization, hence aiding inexperienced authors. A more recent study by Ghafouri

et al. (2024) also showed that ChatGPT not only facilitated teachers' instruction but also promoted L2 learners' writing skills.

Although research has consistently reported the positive impact of ChatGPT on L2 learners' writing, there are still notable limitations. The majority of studies primarily emphasize enhancements in technical writing while allocating less consideration to the essential creative writing elements that are vital for language learners (Barrot, 2023). Furthermore, there are concerns about plagiarism, meaning that inappropriate use of ChatGPT may result in superficial learning (Zou & Huang, 2023). Thus, exploring the impact of utilizing this AI tool in the pre-writing stage can only help overcome this issue, which is still underresearched. Additionally, there is a lack of research specifically focused on the relationship between L2 learners' affective engagement with ChatGPT in the pre-writing stage and their subsequent text quality (Lo et al., 2024). Addressing these gaps would provide a more holistic understanding of the use of ChatGPT in L2 writing.

3 The present study

The present study seeks to address the aforementioned gaps by examining the use of ChatGPT in the context of EFL writing instruction. In particular, it investigates the utilization patterns, affective engagement, and overall impact of ChatGPT on the writing performance of EFL learners. The initial objective of the present research is to clarify the manner in which EFL students employ ChatGPT during the pre-writing phase. The second one is to evaluate the degree of affective engagement that EFL learners exhibit with ChatGPT in the planning stage. The third objective of the study is to assess the influence of the use of ChatGPT during the pre-writing stage on EFL learners' text quality. Finally, this study aims to examine the relationship between the affective engagement of EFL learners with ChatGPT during the pre-writing stage and their subsequent writing performance. The research questions guiding this study include:

1. How do EFL learners utilize ChatGPT during the pre-writing stage?
2. To what extent do EFL learners engage emotionally with ChatGPT during the pre-writing stage?
3. To what extent does the use of ChatGPT during the pre-writing stage affect EFL learners' writing performance?
4. What is the correlation between EFL learners' affective engagement with ChatGPT during the pre-writing stage and their writing performance?

4 Methodology

4.1 Participants and context of the study

The number of the participants was determined using the “*a priori* power analysis”, conducted through in the F-test family function in G*Power version 3.1 (Faul et al., 2009). The parameters ($\alpha=.05$, power value = .95, $f=.25$ medium effect size,

within-groups ANOVA experimental design with two conditions and two writing tasks) indicated that the required sample was 54. This method has been commonly used in prior research on technology (e.g., Chan & Saqip, 2021) and education (e.g., Nguyen & Le, 2022, 2023) to estimate appropriate sample sizes.

Fifty-six first-year Vietnamese EFL students (42 males and 14 females, aged 18–21) from a private university participated in the study. They came from two classes (referred to as Class A and Class B, both instructed by the same teacher) of an English Preparation course, a mandatory program aimed at enhancing academic English writing and reading skills equivalent to the B2 level (according to the CEFR). The students had learned English for about ten years and did not have any overseas training prior to the research. Their English proficiency levels, assessed via the 60-question Oxford Quick Placement Test (OQPT) (Allan, 2004), ranged from B1 to B2 (OQPT score range: 25–48). All of the participants had prior experience in using ChatGPT to assist their learning. They joined the study on a voluntary basis and agreed to sign a written consent form.

4.2 Research instruments

4.2.1 Writing tasks and scoring rubric

The present study included two writing tasks, each requiring participants to compose an argumentative paragraph of at least 120 words to articulate their agreement or disagreement with a specified idea. While each task had a different focus (Writing 1 on promoting public transports and cycling – Writing 2 on restricting individual car and motorbike ownership), they pertained to the same topic (pollution) (see Appendix A) to assure comparability between the two conditions (see Sect. 4.3). These tasks were designed to simulate a timed test condition, with a duration of 25 min per task, mirroring the writing requirements of the students' English Preparation course.

A scoring rubric adopted from Nguyen et al., (2024, p. 17) was employed to evaluate the quality of the students' paragraphs. This rubric assesses written texts across four main criteria: Content, Vocabulary, Grammar, and Organization. Each criterion is scored on a scale from 0 to 5, with detailed descriptions provided for each scoring level. Consequently, the total possible score for each paragraph ranges from 0 to 20.

4.2.2 Affective engagement questionnaire

In this study, an affective engagement questionnaire was utilized to gather data on students' emotional responses and perceived value of the writing tasks, both with and without ChatGPT support. The questionnaire comprises eight items, divided into two sections: items 1–3 assess emotional responses (adapted from Fan & Xu, 2020), and items 4–8 evaluate the perceived value of the writing tasks (based on Kang, 2023). Responses were measured using a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree). To ensure content validity, the questionnaire was reviewed by two experienced researchers with Ph.D. degrees in Applied Linguistics. The validated questionnaire was then administered to participants immediately following each writing task.

4.2.3 ChatGPT logs and note-taking sheets

The logs from ChatGPT and note-taking sheets served as primary data for content analysis, providing insights into how students interacted with and without ChatGPT. To obtain the logs, the participants were required to save the ChatGPT page, download it to their laptop, and send it to the third author's email. Regarding the note-taking sheets, the students were given blank papers in which they could write down anything they wanted during the preparation stage. The sheets were then collected by the teacher upon completion.

4.2.4 Semi-structured interviews

Semi-structured interviews were employed to capture participants' attitudes and experiences during the writing tasks. The interviews comprised three primary questions (Appendix B), focusing on students' opinions regarding the use of ChatGPT during the pre-writing stage, their perceptions of its impact on subsequent writing, and their preferences between the writing tasks. Participants were also asked to elaborate on the reasons for their responses. The interview questions were reviewed by two experts in the field to ensure content validity. Interviews were conducted face-to-face in Vietnamese to prevent any potential misunderstandings and were audio-recorded with the participants' consent. This approach ensured the collection of detailed and reliable qualitative data on students' experiences and attitudes.

4.3 Design

This study employed a within-subjects design with two distinct writing conditions. In the first condition, the students were given 15 min to use ChatGPT to prepare for a writing task (Writing 1). After this preparation period, they had 25 min to compose a paragraph. In the second condition, the participants also had 15 min to prepare for a writing task (Writing 2) but without the support of ChatGPT. Instead, they were provided with blank paper to note down their brainstorming ideas. Following this, they wrote a paragraph within 25 min in response to the given prompt.

To control for the effect of task order, the two conditions were counterbalanced. Specifically, while Class A undertook the ChatGPT condition, Class B completed the non-ChatGPT one. One week later, the order was reversed, allowing all participants to experience both writing tasks (with and without ChatGPT support). Notably, ChatGPT 4.0 was utilized to ensure students had access to the latest features of this AI tool. As all participants were already familiar with using ChatGPT, no additional training session was provided. This approach allowed researchers to explore how learners independently utilized ChatGPT in the L2 pre-writing stage.

4.4 Procedure

This study was conducted within a four-week timeframe. In week one, all participants were informed of the research purpose, provided their consent, and took the Quick Oxford Placement Test. In week two, Class A went through the ChatGPT condition whereas Class B experienced the non-ChatGPT one. After that, both classes were administered the affective engagement questionnaire. One week later, the task order was reversed: while Class A undertook the non-ChatGPT condition, Class B followed the ChatGPT one. Next, the two classes took the same affective engagement questionnaire. Of note, paragraph writing was conducted on paper, and no materials or discussions were allowed to control for the possible impact of

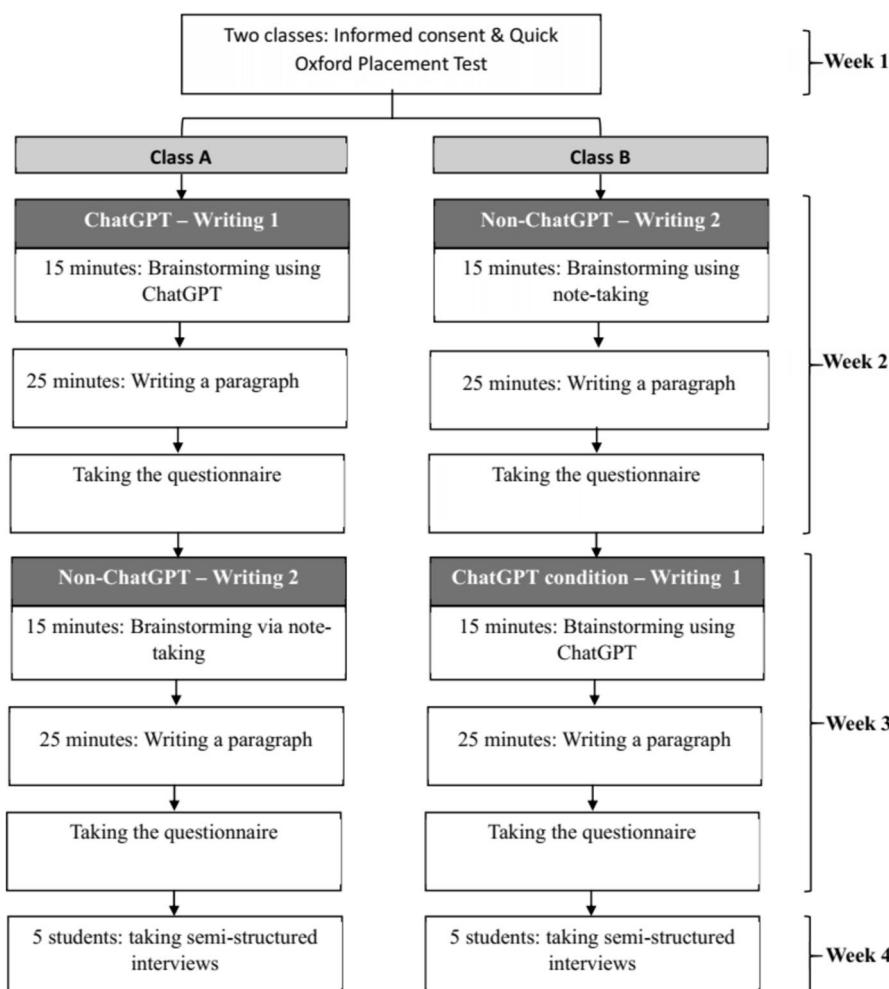


Fig. 1 The procedure of the study

external factors. In week four, five random students from each class (a total of 10 participants) were invited to join the individual semi-structured interviews. The whole procedure is summarized in Fig. 1.

4.5 Data coding and analysis

The data included 112 questionnaires, 56 ChatGPT logs, 56 note-taking sheets, 112 written texts, and 10 interview audio recordings. To answer the first RQ (learners' use), ChatGPT logs and note-taking sheets were analyzed inductively using a theme-based approach (Creswell & Creswell, 2017). First, we read through students' prompts (logs) and notes (sheets), highlighted segments related to learners' behavioral focus, grouped similar segments into codes, and assigned common codes into categories. This process was repeated, and modifications were made until final codes and categories were generated. Next, the frequency of mentions and number of participants were calculated and reported for each finalized category.

In response to the second RQ about learners' affective engagement, the questionnaires were inputted into SPSS 27.0 for reliability and consistency examination. Following this, descriptive analysis with Mean (M) and Standard Deviation (SD) was reported to provide an overview of learners' emotions and perceived usefulness of the writing tasks. To compare the questionnaires between the two conditions (ChatGPT vs. non-ChatGPT), a paired-samples t-test (using the sum scores) was adopted with Cohen's d values representing effect sizes (.20 as small, .50 as medium, and .80 as large) (Cohen, 2013).

To answer the third RQ (learners' writing performance), written texts were graded using the mentioned rubric with four criteria (content, vocabulary, grammar, and organization). The minimum score for each criterion was .0, and the maximum was 5.0, meaning that the total score ranged from .0 to 20.0 for each participant. As the scores were not normally distributed (Shapiro-Wilk tests, all p values <0.001), Median (Mdn), and Interquartile Range (IQR) values were reported. Linear Mixed-Effects Model (LMM) was performed in the R environment via the *lmerTest* package (Kuznetsova et al., 2017) to examine the effect of the use of ChatGPT in the pre-writing stage on learners' overall writing performance (total scores) because this test could handle data of non-normality and minimize Type I errors (Linck & Cummings, 2015). Particularly, while task conditions (ChatGPT vs. non-ChatGPT) were treated as fixed effects (independent variables), the total scores of learners' written texts were considered as a dependent variable. Participants were regarded as random effects, and the fitted model was Score ~ TaskCondition + (1|Participant). Effect sizes were reported using the conditional R^2 values generated via the *MuMin* package (Barton, 2020), interpreted as small (.20 as small, .50 as large) (Plonsky & Ghanbar, 2018). Variance Inflation Factor values did not show any serious multicollinearity between the independent variables ($VIF < 2.0$) (Berk, 2004). Furthermore, non-parametric tests, particularly Wilcoxon signed-rank tests, were used to compare the writing scores regarding the four criteria between the two conditions, with significance values adjusted according to the Bonferroni correction method. The effect sizes were calculated based on Field's (2018) guideline (Z/\sqrt{N}), with r values

demonstrating effect sizes (.20 as small, .40 as medium, and .60 as large) (Plonsky & Oswald, 2014).

Regarding the last RQ, Spearman correlation tests (*rho*) were employed to examine the relationship between learners' total writing scores and their affective engagement (total sum scores of the questionnaires) in both conditions. The r_s values represented the strengths of the correlations (.10 as weak, .30 as medium, and .50 as strong) (Larson-Hall, 2016).

The semi-structured interviews were also analyzed using the mentioned theme-based approach. The third researcher listened to each recording carefully and transcribed them verbatim before sending the transcripts to the participants for accuracy confirmation (member checking, Creswell & Creswell, 2017). Subsequently, we read through the transcribed documents, highlighted segments regarding learners' affective engagement (RQ1) and beliefs of the impact of ChatGPT on their writing (RQ3), assigned codes, and grouped similar codes into categories. This was an iterative process with repeated modifications until the final codes and categories were confirmed. The frequency of mentions was counted, and participants' typical excerpts were translated into English to triangulate the quantitative data. Pseudonyms were used to protect the interviewees' identities.

With respect to inter-rater reliability, we worked together to discuss the coding plan for the data and analyze four ChatGPT logs, four note-taking sheets, four written texts, and one interview. Following this, the first and the second researchers independently coded 10 other ChatGPT logs, 10 other note-taking sheets, 20 other written texts, and two interviews. The agreement rates between the two coders for the data were 92%, 90%, 88%, and 92%, respectively, which indicated a strong consensus (McHugh, 2012). The first researcher continued to code the rest of the data. Notably, for writing performance grading, the third researcher randomly selected students' texts from the data pool for the first and second researchers to code. In other words, the coders were blinded to the task conditions, which ensured the reliability and validity of the obtained scores (Dunning, 2012).

5 Results

RQ1. How do EFL learners utilize ChatGPT in the pre-writing stage?

EFL learners' interactions with ChatGPT were captured via the prompts they used during the pre-writing stage, which were classified into eight primary categories. Table 1 shows that learners interacted with ChatGPT in various ways, but mainly by asking the tool to produce the whole compositions ($n=55$; 98.2%), provide ideas ($n=52$; 92.9%), translate to their L1 ($n=50$; 89.3%), explain meanings of unfamiliar words ($n=41$; 73.2%), shorten the texts ($n=30$; 53.6%), and ask for more appropriate organization of paragraphs ($n=28$; 50%). Notably, the participants tended to require ChatGPT to write the compositions more than twice ($134/55=2.44$), whereas they focused very little on grammar ($n=8$; 14.3%).

In the non-ChatGPT condition, analysis of note-taking sheets demonstrates that learners' behaviors were quite limited during the pre-writing stage (Table 2),

Table 1 Learners' focus in the ChatGPT condition during the pre-writing stage

Prompt	Number of mentions	Number of participants	Sample
Write the whole text	134	55	“write a paragraph for the following topic [type the whole writing prompt]”
Ask for content	124	52	“another problem different from exhaust fumes”
Translate	112	50	“translate to Vietnamese”
Ask for vocabulary	107	41	“what does ‘fumes’ mean?”
Ask for shortening	57	30	“make it shorter, about 150 words”
Ask for organization	54	28	“write a paragraph, not an essay”
Ask for grammar	15	8	“use simple sentences”
Ask for others	46	26	“how many words are there?”

Table 2 Learners' focus in the non-ChatGPT condition during the pre-writing stage

Note	Number of mentions	Number of participants	Sample
Vocabulary	394	56	“private vehicle”, “improve quality”
Content	124	56	“agree”, “limit the use of cars”
Organization	44	17	“introduction ..., body, ...conclusion...”, “for example”
Grammar	1	1	“simple present”

which were assigned into four categories. Specifically, they mainly concentrated on vocabulary ($n=56$; 100%) and content ($n=56$; 100%) while paying little attention to organization ($n=17$; 30.4%) and grammar ($n=1$; 1.8%). Of note, although all participants focused on lexical items (394 counts), it was far higher than that on content-related features (124 counts).

In brief, learners' behaviors were more diverse in the ChatGPT condition than in the non-ChatGPT condition (Fig. 2). Moreover, content and vocabulary were the two criteria that attracted the participants' attention in both writing tasks.

RQ2. To what extent do EFL learners engage emotionally with ChatGPT?

Cronbach's alpha coefficients for the ChatGPT and non-ChatGPT questionnaires were .879 and .932, respectively, with all items' total-correlation values greater than .30. These figures indicated that the questionnaires for the two writing conditions were reliable and internally consistent (Pallant, 2020).

Table 3 illustrates that learners held positive attitudes towards the writing tasks, either with ChatGPT ($M_{General}=4.37$, $SD_{General}=.51$) or without ChatGPT ($M_{General}=3.82$, $SD_{General}=.92$). Particularly, positive emotions were high in both conditions (M values for items 1–3 greater than 4.10). Furthermore, in the ChatGPT condition, content ($M=4.50$) and vocabulary ($M=4.30$) were reported to be the criteria with the highest improvements, while organization ($M=4.14$) and grammar

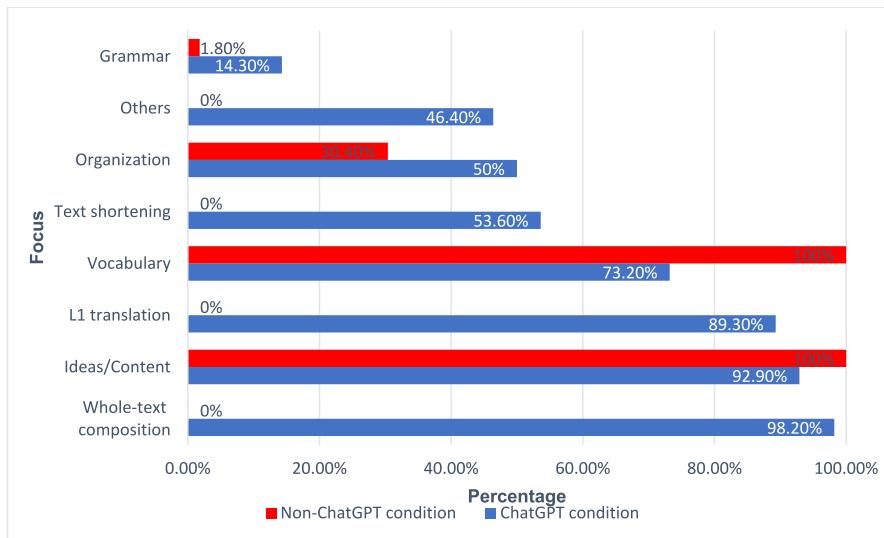


Fig. 2 Comparisons of learners' focus between the ChatGPT and non-ChatGPT conditions

Table 3 Descriptive statistics for the affective engagement questionnaires

Item	ChatGPT		Non-ChatGPT	
	M	SD	M	SD
1. I was willing to participate in this activity	4.71	.59	4.43	1.08
2. I found this activity interesting	4.50	.66	4.14	1.05
3. I was excited to join this activity	4.25	.720	4.20	.77
4. I find this activity helpful in improving my overall writing	4.39	.59	3.82	1.16
5. I find this activity helpful in improving the vocabulary for my paragraph	4.30	.63	3.70	1.16
6. I find this activity helpful in improving the grammar of my paragraph	4.14	.72	3.18	1.21
7. I find this activity helpful in improving the organization of my paragraph	4.14	.90	3.27	1.24
8. I find this activity helpful in improving ideas for my paragraph	4.50	.69	3.79	1.22
General	4.37	.51	3.82	.92

($M=4.14$) were the least ones. In the non-ChatGPT condition, similar patterns were observed, albeit with lower values. Overall, learners' affective engagement with the two writing tasks was positive, with a significant preference for the ChatGPT over the non-ChatGPT condition (Fig. 3) as determined by the paired-samples t-test ($t=2.15$, $p=.036$) at a large effect size ($d=8.90$).

Interview data corroborated these results, with all ten respondents favoring the use of ChatGPT during the writing task. The main reason ($n=9$; 90%) was that this tool could provide learners with essential input, such as ideas and vocabulary needed for their paragraphs. The remarks of two participants demonstrated this point:

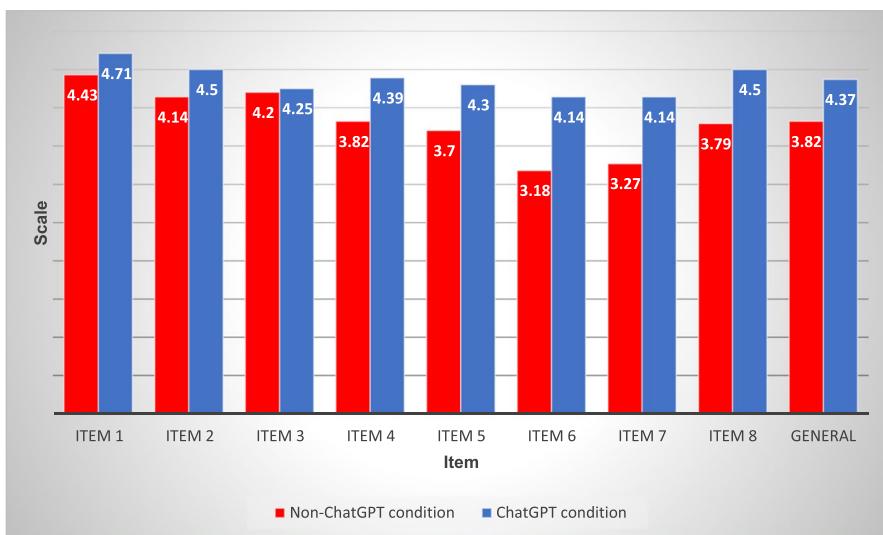


Fig. 3 Comparisons of affective engagement between the ChatGPT and non-ChatGPT conditions

I like writing with the support of ChatGPT because it gives me information and ideas that I cannot think of. (My Thuan)

Well, I prefer to use ChatGPT as it provides me with a lot of great ideas and vocabulary for my writing. (Ngoc Chi)

Another major factor was related to convenience ($n=6$; 60%) since ChatGPT helped students save time and effort. All inquiries were answered quickly in a very detailed manner without the need to read many sources. Two interviewees stated:

I choose ChatGPT also because it helps me save time. I receive answers very fast, and I don't need to go to many websites and combine information. (Le Hoa)

ChatGPT is definitely my choice. All I need is to ask questions. The answers are provided in detail very quickly, saving me a lot of time. (Trong Nghia)

RQ3. To what extent does the use of ChatGPT during the pre-writing stage affect EFL learners' writing performance?

Table 4 presents descriptive statistics for learners' writing scores in both conditions. In general, learners produced better writing with the support of ChatGPT during the pre-writing stage ($Mdn_{Total}=11.0$) than without it ($Mdn_{Total}=10.5$). Notably, content ($Mdn=3.00$) and vocabulary ($Mdn=3.00$) were higher in the ChatGPT condition than in the non-ChatGPT one ($Mdn_{Content}=2.75$; $Mdn_{Vocabulary}=2.50$). Scores regarding grammar and organization remained similar in the two writing tasks.

Linear Mixed-effects Model in Table 5 confirms that learners achieved significantly higher scores in the ChatGPT condition than in the non-ChatGPT one

Table 4 Descriptive statistics for writing performance in two writing tasks

	ChatGPT (<i>N</i> =56)		Non-ChatGPT (<i>N</i> =56)	
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>
Content	3.00	1.40	2.75	.50
Vocabulary	3.00	.50	2.50	.00
Grammar	2.50	.90	2.50	1.00
Organization	3.00	.50	3.00	.50
Total	11.00	2.00	10.50	1.50

Table 5 The impact of ChatGPT use on learners' overall writing performance

Fixed effects	β	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	11.55	.40	28.71	<.001
Task Condition: non-ChatGPT	-.89	.24	-3.72	<.001
Marginal $R^2 = .11$; Conditional $R^2 = .21$				

SE standard error; the non-ChatGPT condition was compared against the ChatGPT one; the significance value was set at .05

Table 6 The impact of ChatGPT use on learners' writing scores regarding four criteria

ChatGPT – non-ChatGPT			
	<i>Z</i>	<i>p</i>	<i>r</i>
Content	-4.18	<.001	.56
Vocabulary	-4.28	<.001	.57
Grammar	-.36	.720	.05
Organization	-.01	.994	.00

Significance value was adjusted at .0125 (= .05/4) for 4 comparisons

($\beta = -.89$, $p < .001$). The observed effect size was small (Conditional $R^2 = .21$), meaning that the model explained about 21% of the total variance.

To better understand the differences in writing scores between the two conditions, a series of Wilcoxon signed-rank tests were run. The results (Table 6) reveal that there were significant discrepancies in content ($Z = -4.18$, $p < .001$) and vocabulary ($Z = -4.28$, $p < .001$) scores, with large effect sizes (r values close to .60). Nonetheless, the disparities in grammar ($Z = -.36$, $p = .720$) and organization ($Z = -.01$, $p = .994$) performance were not considerable, with very small effect sizes (r values smaller than .20). In brief, it could be indicated that the use of ChatGPT significantly improved learners' writing scores, particularly regarding content and vocabulary. For a concrete example of a participant's written texts in the two conditions, see Figs. 4 and 5.

The words in bold (Fig. 4) are the ones incorporated from ChatGPT or note-taking sheets. As can be seen, in the ChatGPT condition, the participant provided three

ChatGPT prompt

Anonymous

You should spend about 30 minutes on this task. Write an opinion paragraph about the following topic: Many people believe that promoting public transport and cycling is the most effective way to address urban air pollution. To what extent do you agree or disagree? Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 150 words.



ChatGPT

Promoting public transport and cycling as solutions to urban air pollution is undoubtedly a commendable approach, and I firmly support this viewpoint. Firstly, public transportation systems such as buses, trains, and subways can significantly reduce the number of individual vehicles on the road, thus curbing emissions from exhaust fumes. Additionally, investing in cycling infrastructure encourages people to opt for eco-friendly modes of transportation, further reducing the reliance on cars and motorcycles.

Written text (ChatGPT condition)

... From my point of view, **promoting public transport** is the best way. First, it **reduces** the amount of **exhaust fumes** released from vehicles, lead to the decrease of **air pollution**. In addition, when the number of public transport increases, the less private ones remain. As a result, there is a less accident occur. **Additionally**, we should **invest** in **eco-friendly** vehicles such as electric **buses and subways** to **encourage people** to use more public transport and **reduce the use of cars and motorbikes**...

Fig. 4 Sample of a participant's written text in the ChatGPT condition

Note-taking sheets

Person said, to fight against, air pollution, limit private vehicle, essential, agree, by:

+ use electric transport or public transport instead of private vehicles → reduce gases, fumes

Written text (non-ChatGPT condition)

...I strongly agree with this statement that **limiting** the number of **private vehicle** engage in traffic activities. First, by restricting individual vehicle, the amount of **gases** from vehicle will decrease and reduce the rate of **air pollution**. Second, instead of individual car and motorbike ownership, use public transport in traffic also **reduce air pollution**...

Fig. 5 Sample of a participant's written text in the non-ChatGPT condition

different ideas (i.e., decrease air pollution, trigger fewer accidents, invest in eco-friendly vehicles) and a wide range of words/ phrases provided by ChatGPT. These features help enhance the quality of the text. In contrast, in the non-ChatGPT condition (Fig. 5), although the participant also integrated some vocabulary and ideas from the notes, the final product was quite limited in content (only mentioning air pollution again and again). Moreover, the lexical use is less accurate and diverse when compared to the ChatGPT version. Taken together, it is evident that ChatGPT could promote EFL learners' writing quality, especially content and vocabulary.

These results were supported by the interview data in which the respondents ($n=10$; 100%) believed that ChatGPT helped them improve their writing, particularly content and vocabulary. The sharing of Nhu My and Tuong Van demonstrated this idea:

I think ChatGPT helped me a lot with my writing. After the preparation stage, I started to write my paragraphs and reused some ideas and words provided by ChatGPT. (Nhu My)

I could think of only one idea in the preparation stage, so I asked ChatGPT to help me. It gave me many interesting ideas for my paragraph. I even asked the tool to provide me with some synonyms to avoid repetition. I integrated these ideas and words into my writing later. (Tuong Van)

RQ4. What is the correlation between EFL learners' affective engagement with ChatGPT during the pre-writing stage and their overall writing performance?

Table 7 illustrates that there was a moderate positive relationship between learners' affective engagement and their overall writing performance in the ChatGPT condition ($r_s=.301, p=.021$). However, no significant correlation was found in the non-ChatGPT condition ($r_s=.160, p=.238$). In other words, learners' emotions and perceived usefulness of ChatGPT in the pre-writing stage could moderately predict their writing outcomes.

6 Discussion

The present study seeks to explore the use of ChatGPT in L2 writing, particularly in the pre-writing stage. It extends previous studies in three main ways. First, it could address the concern that students may rely excessively on ChatGPT, which could lead to plagiarism, by only allowing the participants to use this AI tool in the preparation stage only. Second, it sheds light on the relationship between L2 learners' affective engagement and their writing quality. The results are discussed in detail below.

6.1 EFL learners' use of ChatGPT during the pre-writing stage

One significant finding was that learners adopted a wide array of strategies in the ChatGPT condition. They primarily used the tool to write the whole text and

Table 7 Correlation matrix between learners' affective engagement and writing performance in both conditions

	Total score (ChatGPT)		Total score (Non-ChatGPT)	
	r_s	p	r_s	p
Affective engagement	.301	.021	.16	.238

The significance value was set at .05

provide ideas as well as vocabulary related to the topic. These findings are in line with previous studies reporting that students often looked for ideas and lexical items when interacting with ChatGPT (Rojas, 2024; Wang, 2024). In contrast, the frequency of requesting information related to grammar was far lower. Thus, it could be indicated that learners focused more on meanings when interacting with ChatGPT in the pre-writing stage. There are three possible explanations for this result. First, by asking for learners' opinions and explanations on a given topic, the writing task was inherently meaning-based. To fulfill the requirement, learners needed to seek content and vocabulary, particularly arguments and examples, to support their viewpoint (Kang, 2020), which corroborates the notion that meanings were often prioritized over forms (VanPattern, 2004). Second, the Writing Process Model (Hayes & Flower, 1980) posits that in the planning (pre-writing) stage, learners had the tendency to concentrate on idea generation. This might explain why in the current study, the participants were more likely to pay attention to content and vocabulary (needed to convey ideas). Third, according to the Trade-Off Hypothesis (Skehan, 1998), learners often had to sacrifice some aspects (i.e., grammar, organization) for others (i.e., ideas, lexis), especially in a timed task. As grammar and organization added little to meaning (the main focus as explained earlier), these features could have been overlooked.

Regarding the non-ChatGPT condition, it was found that learners' strategies during the pre-writing stage were quite limited when compared to the ChatGPT condition, mostly about noting down content-related and lexical features. Without any additional support, learners appeared to rely mainly on their own background knowledge and memory to prepare for the writing task (Hayes & Flower, 1980). The lack of exposure to input may have refrained them from performing diverse strategies as observed in the ChatGPT condition. Put differently, ChatGPT could promote learners to perform a wider range of techniques to process the task at hand, specifically the planning for paragraph writing.

It is noteworthy that a significant number of participants requested ChatGPT to shorten and translate its outputs into their L1 multiple times. Such actions suggest that learners may perceive ChatGPT's responses as lengthy or challenging to comprehend. Given ChatGPT's tendency to generate texts in natural language (Ray, 2023), L2 learners, particularly those at lower proficiency levels, as observed in this study, might find these responses difficult to understand. This challenge likely prompts them to seek translated or simplified versions. Therefore, employing appropriate prompts to guide ChatGPT towards producing texts that align with learners' needs or proficiency levels is crucial to mitigate the tendency to rely excessively on L1.

6.2 EFL learners' affective engagement with ChatGPT

Another important finding was that learners were positively engaged with both the writing tasks, but the level of affective engagement in the ChatGPT condition was significantly higher than that in the non-ChatGPT condition. This result was further

supported via the interview data, which showed that all respondents preferred ChatGPT to non-ChatGPT tasks. Such a finding aligns with previous studies (e.g., Ali et al., 2023; Koltovskaya et al., 2024; Zou & Huang, 2023), which reported positive attitudes towards ChatGPT use in L2 writing. The explanation was that learners perceived ChatGPT as a useful tool for their writing. Indeed, research has demonstrated that this platform could provide users with instant responses tailored to their needs and assistance in various stages of the writing process (Yuan et al., 2024; Zou & Huang, 2023). These affordances might have resulted in learners' favorable emotions and attitudes towards the ChatGPT condition.

6.3 The impact of ChatGPT use in the pre-writing stage on learners' writing performance

The use of ChatGPT in the pre-writing stage was found to significantly affect learners' writing performance, particularly in terms of content and vocabulary. This result could be explained in light of the strategies learners employed during their interactions with ChatGPT, as discussed earlier (Sect. 6.1). Learners predominantly sought assistance with content-related and lexical features, which likely led them to focus on and notice information in these aspects. According to Schmidt's (2001) Noticing Hypothesis, learners who notice linguistic features more are likely to experience greater learning gains. This may explain why improvements were observed in ideas and vocabulary – the two key writing criteria. The enhancement in learners' writing performance in the ChatGPT condition is in line with prior research (e.g., Ghafouri et al., 2024; Zou & Huang, 2023), wherein the use of this tool was reported to lead to improvements in writing skills. These results lend support to the Sociocultural Theory (Storch, 2018; Vygotsky, 1978), which argues that interactions with a more competent user as well as a mediational tool, ChatGPT in this case, could foster learning gains.

6.4 The correlation between EFL learners' affective engagement with ChatGPT and overall writing performance

The findings from the Spearman tests indicated a moderate positive correlation between learners' affective engagement with ChatGPT and their overall writing quality. This correlation, however, was not observed in the non-ChatGPT condition. These results echo prior research (e.g., Ali et al., 2023; Fredricks et al., 2004), which has consistently shown that learners' affective engagement is positively linked to academic performance. Ali et al. (2023) specifically reported that AI-based instruction enhances learners' motivation and writing skills, suggesting that heightened engagement during the writing process leads to increased focus on the task and greater attention to essential information (Schmidt, 2001). Additionally, the control-value theory (Pekrun, 2006) underscores the importance of positive emotions in educational contexts, as they enhance cognitive resources and thereby improve academic performance. Thus, the use of ChatGPT in the pre-writing stage can foster positive emotions, which in turn promotes better writing outcomes.

In a broader sense, it can be indicated that AI-tools such as ChatGPT play a pivotal role in language learning. They provide learners with a positive learning environment (Ali et al., 2023; Pınar Özçelik & Yangın Eksi, 2024) by reducing anxiety and stress levels (Saif et al., 2024). This positive attitude has been shown to be closely linked to improved academic performance (Dai & Wang, 2024; Dewaele et al., 2018), making AI a powerful asset in modern language education.

7 Conclusion

The present study aimed to explore EFL learners' use of ChatGPT during the pre-writing stage, how they emotionally engaged with this tool, and the relationship between their affective engagement and writing performance. Via a mixed-methods approach, it was found that learners employed a wide range of strategies when interacting with ChatGPT in the pre-writing stage, mainly focusing on meaning rather than form. Moreover, their affective engagement with this tool was highly positive, and their writing scores (content, vocabulary) were significantly better in the ChatGPT condition than in the non-ChatGPT one. Finally, a moderate positive relationship was established between learners' affective engagement and their overall writing quality. These results imply that how EFL learners utilized and perceived ChatGPT could predict their writing performance.

Several limitations of the current study should be considered. First, it only examined the use of ChatGPT in the pre-writing stage, leaving its impact on L2 learners' writing in the subsequent stages unclear. Further research is warranted to explore the role of ChatGPT throughout the whole writing process: planning, drafting, and revising. Second, the study focused mainly on affective engagement with this AI tool; other dimensions of engagement (i.e., cognitive, behavioral, and social engagement; see Vo, 2024) should be investigated to provide a more comprehensive understanding of learners' engagement. Lastly, the present research primarily addressed the short-term use of ChatGPT within a single task. Longitudinal studies are needed to get insights into how learners' engagement changes over time, especially given its malleability (Zhang & Gao, 2024).

The present study contributes to the literature on ChatGPT use in L2 writing by shedding light on how this tool can affect learners' affective engagement and writing performance, particularly in the planning stage. From a pedagogical perspective, institutional leaders and teachers are recommended to embrace ChatGPT in L2 classrooms rather than diminishing its use because this tool could facilitate writing gains (Ghafouri et al., 2024; Zou & Huang, 2023). Its use should extend beyond the post-writing stage (Steiss et al., 2024) to include the pre-writing phase as well. However, it is essential that students receive proper training on how to use ChatGPT effectively, given the tool's potential to produce inaccurate information (Yan, 2023). Additionally, since learners' affective engagement is linked to their writing performance, teachers should monitor students' interactions with ChatGPT. For example, by reviewing ChatGPT logs or observing usage, teachers can gain insights into students' emotions and actual use of the tool, allowing them to provide timely feedback and adjustments for more effective integration of AI in the L2 classroom.

Appendix A Writing Questions

Task 1:

You should spend about 25 min on this task. Write about the following topic:

Many people believe that promoting public transport and cycling is the most effective way to address urban air pollution

To what extent do you agree or disagree?

Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 120 words

Task 2:

You should spend about 25 min on this task. Write about the following topic:

Many people believe that in order to combat air pollution, restricting individual car and motorbike ownership is necessary

To what extent do you agree or disagree?

Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 120 words

Appendix B Interview Questions

1. How did you feel when interacting with ChatGPT for your paragraph writing preparation? Why?
2. How helpful did you find the use of ChatGPT for your paragraph writing? Why? (*Follow-up: Grammar, Vocabulary, Ideas, Organization?*).
3. Which activities do you prefer: writing with or without the assistance of ChatGPT? Why?

Acknowledgements We wish to thank all the participants who generously shared their time and experiences for the purposes of this study. We also extend our gratitude to our university and colleagues for their support and the facilities provided.

Author Contributions Long Quoc Nguyen: Be responsible for conceptualizing and designing the study, analyzing and interpreting results, manuscript preparation, writing, editing, and proofreading the final version of the manuscript. Ha Van Le: Be responsible for conceptualizing and designing the study, synthesizing and reviewing the literature, collecting data, analyzing and interpreting qualitative results, writing, editing, and proofreading the final version of the manuscript. Phuc Thinh Nguyen: be responsible for conceptualizing and designing the study, collecting data, manuscript preparation, writing, editing, and proofreading the final version of the manuscript. The authors have read and agreed to the published version of the manuscript.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability Data and materials are available if required.

Declarations

Ethics approval and consent to participate This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of FPT University, Vietnam. All participants provided their informed consent to participate in this study. Prior to participation, each participant was informed about the study's aims, the procedures involved, potential risks, and the confidentiality of the data collected. Written informed consent was obtained from all participants.

Cosent for publication Consent for publication was obtained from all participants involved in this study. Participants were informed that anonymized quotes and data might be used in publications or presentations resulting from this research.

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Adams, D., & Chuah, K. M. (2022). Artificial intelligence-based tools in research writing: Current trends and future potentials. In Churi, P.P., Joshi, S., Elhoseny, M., & Omrane, A. (Eds.), *Artificial Intelligence in Higher Education: A Practical Approach* (1st ed.). CRC Press. <https://doi.org/10.1201/9781003184157>
- Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation: Teachers and Students' Voices. *Journal of English Studies in Arabia Felix*, 2(1), 41–49. <https://doi.org/10.56540/jesaf.v2i1.51>
- Allan, D. (2004). *Oxford placement test*. Oxford University Press.
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.aww.2023.100745>
- Barton, K. (2020). MuMin: Multi-model inference (version 1.43.17). Retrieved from <https://CRAN.R-project.org/package=MuMin>.
- Berk, R. A. (2004). *Regression analysis: A constructive critique*. Sage.
- Chan, E. Y., & Saqib, N. U. (2021). Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. *Computers in Human Behavior*, 119, 106718.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.
- Dai, K., & Wang, Y. (2024). Enjoyable, anxious, or bored? Investigating Chinese EFL learners' classroom emotions and their engagement in technology-based EMI classrooms. *System*, 123, 103339. <https://doi.org/10.1016/j.system.2024.103339>
- Dewaele, J. M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, 22(6), 676–697. <https://doi.org/10.1177/1362168817692161>
- Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press
- Ellis, R. (2010). EPILOGUE: A Framework for Investigating Oral and Written Corrective Feedback. *Studies in Second Language Acquisition*, 32(2), 335–349. <https://doi.org/10.1017/S027226310990544>
- Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*, 50, 100775. <https://doi.org/10.1016/j.jslw.2020.100775>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. Sage.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241239764>
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 100666. <https://doi.org/10.1016/j.aww.2022.100666>
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 4–30). Erlbaum.

- Kang, E. Y. (2020). Using model texts as a form of feedback in L2 writing. *System*, 89, 102196. <https://doi.org/10.1016/j.system.2019.102196>
- Kang, E. Y. (2023). EFL learners' perceptions and their association with the effectiveness of model texts as a feedback tool. *Innovation in Language Learning and Teaching*, 18(1), 47–58. <https://doi.org/10.1080/17501229.2023.2226144>
- Koltovskia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lantolf, J. P. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford University Press.
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. Routledge.
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Linck, J. A., & Cummings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185–207. <https://doi.org/10.1111/lang.12117>
- Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 105100. <https://doi.org/10.1016/j.comedu.2024.105100>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Nguyen, L. Q., & Le, H. V. (2022). Quizlet as a Learning Tool for Enhancing L2 Learners' Lexical Retention: Should It be Used in Class or at Home? *Human Behavior and Emerging Technologies*, 2022L, 8683671. <https://doi.org/10.1155/2022/8683671>
- Nguyen, L. Q., & Le, H. V. (2023). Enhancing L2 learners' lexical gains via Quizlet learning tool: The role of individual differences. *Education and Information Technologies*, 28(9), 12143–12167. <https://doi.org/10.1007/s10639-023-11673-0>
- Nguyen, L. Q., Nguyen, B. T. T., & Phuong, H. Y. (2024). Exploring the use of model texts as a feedback instrument in expository writing: EFL learners' noticing, incorporations, and text quality. *Assessing Writing*, 62, 100890. <https://doi.org/10.1016/j.asw.2024.100890>
- Pallant, J. (2020). *SPSS Survival Manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). Routledge.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting *R*² values. *The Modern Language Journal*, 102(4), 713–731. <https://doi.org/10.1111/modl.12509>
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. <https://doi.org/10.1111/lang.12079>
- PunarÖzçelik, N., & YangınEkşİ, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant - a case study. *Smart Learning Environments*, 11(1), 10. <https://doi.org/10.1186/s40561-024-00296-8>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Rojas, A. J. (2024). An Investigation into ChatGPT's Application for a Scientific Writing Assignment. *Journal of Chemical Education*, 101(5), 1959–1965. <https://doi.org/10.1021/acs.jchemed.4c00034>
- Saif, N., Khan, S. U., Shaheen, I., ALotaibi, F. A., Alnfai, M. M., & Arif, M. (2024). Chat-GPT; validating Technology Acceptance Model (TAM) in education sector via ubiquitous learning mechanism. *Computers in Human Behavior*, 154, 108097. <https://doi.org/10.1016/j.chb.2023.108097>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
- Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., ... & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, 86, 224–235. <https://doi.org/10.1016/j.comedu.2015.08.001>

- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skinner, E. A., Pitzer, J. R., & Steele, J. S. (2016). Can student engagement serve as a motivational resource for academic coping, persistence, and learning during late elementary and early middle school? *Developmental Psychology, 52*(12), 2099–2117. <https://doi.org/10.1037/dev0000232>
- Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching, 51*(2), 262–277. <https://doi.org/10.1017/S0261444818000034>
- Storch, N. (2019). Collaborative writing. *Language Teaching, 52*(1), 40–59. <https://doi.org/10.1017/S0261444818000320>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Swain, M. (2006). Languaging, agency and collaboration in advanced second language learning. In H. Byrnes (Ed.), *Advanced language learning: The contributions of Halliday and Vygotsky* (pp. 95–108). Continuum.
- Trang, N. H., & Barrot, J. S. (2023). Pre-Writing and Post-Writing Explicit Instruction: Differential Effects on L2 Learners' Writing Accuracy. *RELC Journal*. Advanced online publication. <https://doi.org/10.1177/00336882231184270>
- 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>
- VanPatten, B. (2004). *Processing instruction*. Lawrence Erlbaum.
- Vo, H. (2024). Towards Clarification of the Second Language Learner Engagement Construct: Taking Stock of its Conceptual Overlap and Hierarchical Structure. *Applied Linguistics, amae012*. <https://doi.org/10.1093/applin/amea012>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, C. (2024). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning, 1*–22. Advance online publication. <https://doi.org/10.1007/s10758-024-09744-3>
- 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>
- Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. *Language Learning in Higher Education, 14*(1), 41–56. <https://doi.org/10.1515/cercles-2023-0006>
- Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing, 60*, 100842. <https://doi.org/10.1016/j.asw.2024.100842>
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>
- Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology, 14*, 1259531. <https://doi.org/10.3389/fpsyg.2023.1259531>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Long Quoc Nguyen¹  · Ha Van Le¹  · Phuc Thinh Nguyen¹

 Ha Van Le
vanlh8@fe.edu.vn

Long Quoc Nguyen
quocnl2@fe.edu.vn

Phuc Thinh Nguyen
thinhnp10@fe.edu.vn

¹ Department of English Language, FPT University, Ho Chi Minh City, Vietnam