Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process

Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process

David James Woo [a, *], Deliang Wang [b], Kai Guo [b], and Hengky Susanto [c]

[a] Precious Blood Secondary School, Hong Kong, China

[b] The University of Hong Kong, Hong Kong, China

[c] Education University of Hong Kong, Hong Kong, China

[*] Corresponding author

- Postal address: Precious Blood Secondary School, 338 San Ha Street, Chai Wan, Hong Kong, China

- Email address: wootang01@hotmail.com

- Phone: +852 2570 4172

Statements and Declarations

Biographical Note

David James Woo is a secondary school teacher. His research interests are in artificial intelligence, natural language processing, digital literacy, and educational technology

innovations. Email address: wootang01@hotmail.com ORCID: https://orcid.org/0000-0003-4417-3686

Deliang Wang is a Ph.D. student in the Faculty of Education at The University of Hong Kong. His research directions include artificial intelligence in education, explainable artificial intelligence, and educational data mining. His recent publications have appeared in international peer-reviewed journals and conferences, such as *Education and Information Technologies*, *International Journal of Educational Research*, *International Journal of Artificial Intelligence in Education*, *International Conference on Artificial Intelligence*, and *International Conference on Educational Data Mining*. Email address: wdeliang@connect.hku.hk ORCID: https://orcid.org/0009-0008-6488-0234

Kai Guo is a Ph.D. candidate in the Faculty of Education at the University of Hong Kong. His research interests include technology-enhanced language learning, computer-supported collaborative learning, artificial intelligence in education, and gamification in education. His recent publications have appeared in international peer-reviewed journals such as *Computers & Education*, *Education and Information Technologies*, *Interactive Learning Environments*, *Journal of Educational Computing Research*, *TESOL Quarterly*, and *Assessing Writing*. Email address: kaiguo@connect.hku.hk ORCID: https://orcid.org/0000-0001-9699-7527

Hengky Susanto received his BS, MS and PhD degree in computer science from the University of Massachusetts system. He was a post-doctoral research fellow at University of Massachusetts Lowell and Hong Kong University of Science and Technology. He was also senior researcher at Huawei Future Network Theory Lab. Currently, he is a principal researcher in a startup mode research laboratory and a lecturer at Education University of Hong Kong. His research interests include applied AI (computer vision and NLP) to solve

complex social problems, smart city, and computer networking (e.g., datacenter network, congestion control, etc.).

Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process

Abstract

This mixed methods study explores EFL students' experiences and perceptions as they learn to write a composition with ChatGPT's support in a classroom instructional context. Students' perceptions are explored in terms of their motivation to learn about ChatGPT, cognitive load and satisfaction with the learning process. In a workshop format, twenty-one Hong Kong secondary school students were introduced to ChatGPT, learned prompt engineering skills, and attempted a 500-word English language writing task with ChatGPT's support. Data collected included a pre-workshop motivation questionnaire, think-aloud protocols during the writing task, and a post-workshop questionnaire on motivation, cognitive load, and satisfaction. Results revealed no significant difference in students' motivation before and after the workshop, but mean motivation scores increased slightly. Students reported high cognitive load during the writing task, especially during prompt engineering. However, students expressed high satisfaction with the workshop overall. Findings indicate ChatGPT's potential to engage EFL students in the writing classroom, but its use can impose heavy cognitive demands. To ensure that ChatGPT use supports EFL writing without overwhelming students, educators should consider an iterative design process

for activities and instructional materials and carefully scaffolding instruction, especially for

prompt engineering.

Keywords: ChatGPT; cognitive load; English as a foreign language (EFL); prompt

engineering; student motivation; student perceptions

Teaching EFL Students to Write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process

1. Introduction

Generative artificial intelligence (AI) language models (LMs) such as OpenAI's GPT-

2, GPT-3 and GPT-4 have captivated educators' interest, because they can generate large

chunks of coherent text indistinguishable from human writing (Brown et al., 2020) and

proficiently perform a variety of natural language processing tasks when instructed or

prompted (Ouyang et al., 2022). Furthermore, ChatGPT has popularized interaction with

LMs through a chatbot interface, that is, a conversational user interface that enables people to

engage in meaningful verbal or text-based exchanges with an LM (Kim et al., 2022). As

ChatGPT has captured popular imagination, ChatGPT is used as a catchall phrase for

chatbots that use transformers-based LMs (Vaswani et al., 2017).

ChatGPT enables students to write with a *machine-in-the-loop,* which refers to a

collaborative process between a student and a chatbot to complete a writing task. As defined

by Clark et al. (2018) and illustrated in Figure 1, the process is iterative. First, a student

prompts or delivers a set of instructions to guide ChatGPT such as a question, an imperative

statement or an excerpt from a text. Based on its understanding of the student's prompt,

ChatGPT generates output. The student then evaluates the output, accepting, rejecting or modifying ChatGPT's output for integration into the student's written composition. The cycle loops until the completion of the writing task with the student retaining full control over the written composition. Having previously been applied to creative writing, researchers found writers appreciated ChatGPT's fresh ideas, which helped overcome writer's block, while still maintaining ownership over their work. However, the quality of written compositions have not necessarily improved with ChatGPT suggestions (Clark et al., 2018; Calderwood et al., 2020).

Fig 1 Machine-in-the-loop system characteristics with arrows indicating relationships between a student, ChatGPT, and ChatGPT output to complete student writing
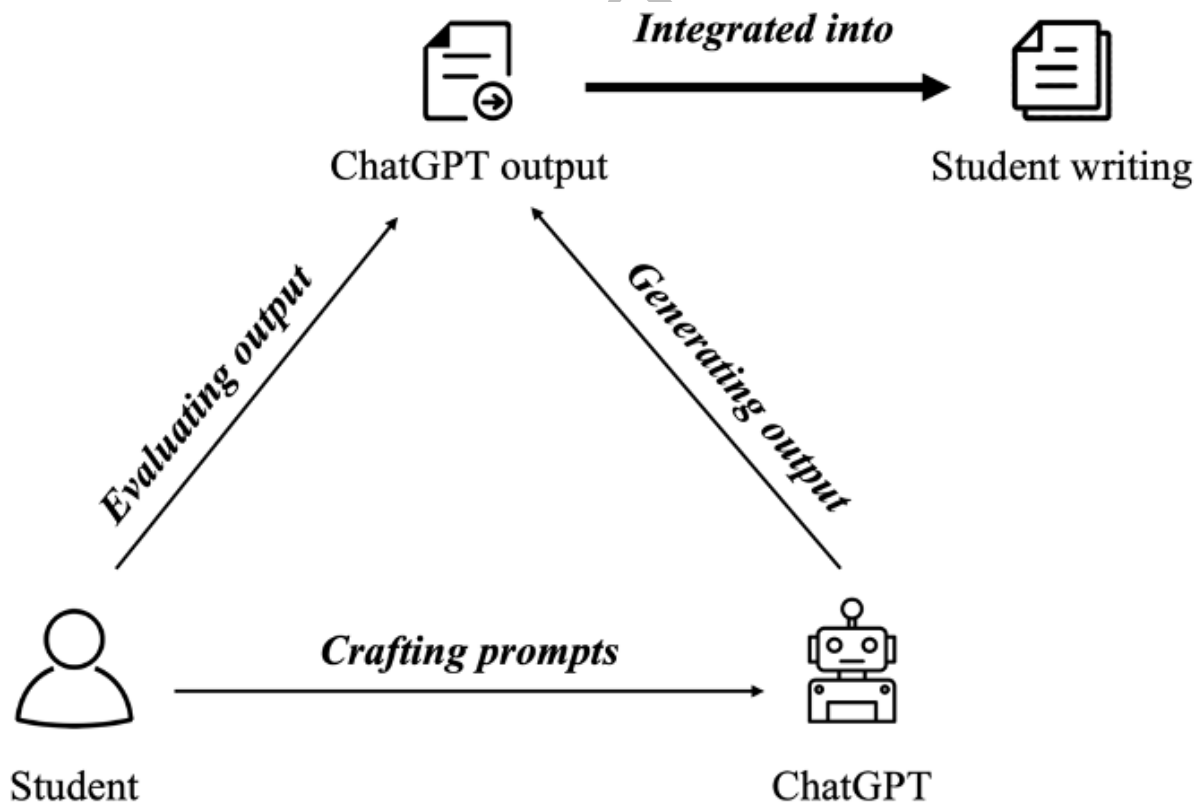
Notwithstanding ChatGPT's potential benefits, the integration of ChatGPT into the English as a foreign language (EFL) writing classroom remains largely unexplored in terms

of students' experiences and perceptions. This study aims to fill this research gap in the

context of Hong Kong secondary school students learning to compose a written EFL

composition with ChatGPT support. The objective is to explore how students perceive their

experience of learning this innovative writing approach in terms of their motivation to learn

about ChatGPT, cognitive load, and satisfaction with the learning process. These aspects are

critical as they directly impact students' learning behaviors, engagement, and ultimately, their

writing outcomes. Furthermore, understanding these aspects can provide valuable insights for

educators, informing instructional approaches for integrating ChatGPT into the EFL writing

classroom. The overarching question guiding this research is: How do EFL students perceive

learning to write with ChatGPT in a classroom context?

1.1. Potential of ChatGPT in the EFL Writing Classroom

In the EFL writing classroom, students can face difficulty in retrieving intended

English words and take time to translate ideas from their first language to English (Gayed et

al., 2022). Students can struggle to generate ideas independently (Author 1 et al., 2023) and

may not have sufficient and effective engagement with peer feedback in the writing process

(Zhang & Hyland, 2023), although collaborative writing is an effective pedagogical practice

(Li & Zhang, 2023) and students' quality of writing can benefit from collaboration (Hsu,

2023).

The implementation of ChatGPT in an EFL writing classroom may support learning

opportunities for students. This is because a chatbot can act as an ideal collaborative partner

for EFL students (Author 3 et al., 2022), and ChatGPT is highly capable of natural language

tasks such as brainstorming ideas, generating texts, answering questions, rewriting texts and

summarizing texts (Ouyang et al., 2022). Conceptual studies have explored the use of

ChatGPT in EFL writing classrooms, suggesting hypothetical use cases. For instance, Hwang

and Chen (2023) suggested the potential application of students using ChatGPT as a proofreader for academic writing in EFL courses. Su et al. (2023) explored the potential of ChatGPT in assisting students with preparing outlines, revising content, proofreading, and reflecting.

However, some EFL teachers fear that students may become dependent on ChatGPT and its dubious suggestions (Ulla et al., 2023). ChatGPT could reinforce biased ideas (Mohamed, 2023). Additionally, students could use ChatGPT with neither much effort nor student input to complete writing assignments, undermining students' acquisition of English and writing skills (Gayed et al., 2022), and critical and creative thinking (Barrot, 2023). Empirical studies featuring actual use cases of ChatGPT support in the EFL writing classroom show mixed results. For instance, Cao and Zhong (2023) compared ChatGPT feedback, EFL teacher feedback and student feedback for improving 45 university students' written translation performance and found ChatGPT feedback was less effective than other feedback types in improving performance. On the other hand, Athanassopoulos et al. (2023) examined ChatGPT's effectiveness as a writing vocabulary and grammar feedback tool for eight, 15-year old migrants and refugees. After writing a task and receiving improved versions of their writing generated by ChatGPT, the students could increase the total number of words, the unique words and the number of words per sentence when writing a similar task.

1.2. Genre Writing and Prompt Engineering as Genre in the EFL Classroom

How teachers should approach the instruction of writing with ChatGPT in an EFL classroom is a complex issue. From an EFL teaching and learning perspective, a teacher adopting an explicit, instructional approach to EFL writing appears necessary for whether implementing ChatGPT intentionally benefits or hinders students' acquisition of knowledge and skills. In this regard, although process writing has been a popular, inductive writing

strategy, Hyland (2007) has argued it has limited value for EFL learners who lack access to cultural knowledge that facilitates effective, independent writing. Instead, this study approaches EFL students' acquisition of writing through *genre*, which emphasizes communicating effectively through different types of texts, and their specific conventions, language features, and structures (Hyland, 2019). As illustrated in Figure 2, a genre-approach to writing instruction is explicit, including stages such as a teacher modeling a genre, joint construction of a text in the genre, and a student's independent construction of a text in the genre. Like a conventional EFL teacher, ChatGPT could support students at each stage by, for example, generating model texts of a genre, identifying the genre's linguistic features, collaboratively writing sections of a text with a student, and suggesting vocabulary, grammar and outlines and providing feedback for a student's independent construction of a text.

Fig 2 Three sequential stages of a genre approach to writing and ChatGPT's support at each stage
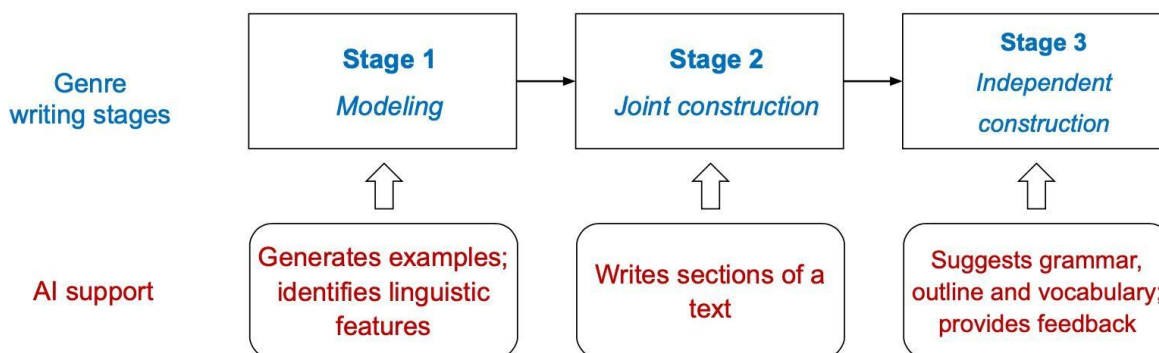
When students write with a machine-in-the-loop, the effect of ChatGPT on students' knowledge and skill development in genre writing depends on how well students give instructions or prompts for ChatGPT. Proficient crafting of prompts or prompt engineering can significantly enhance the quality of ChatGPT's generated output and the overall effectiveness of the interaction with ChatGPT (Reynolds & McDonell, 2021). Since

constructing appropriate prompts is not straightforward for non-technical users (Zamfirescu-Pereira et al., 2023), and ChatGPT prompts are an emergent genre, scholars have proposed example prompts for hypothetical ChatGPT use cases in the literature (Hwang & Chen, 2023; Kohnke et al., 2023; Su et al., 2023).

The implication for the EFL writing classroom is that ChatGPT's capability to support students at different stages of genre writing would depend on teachers not only developing students' knowledge and skills of the target text type but also developing student's prompt engineering knowledge and skills. Furthermore, the authors anticipate prompt engineering instruction could compose a significant part of students' learning to write with ChatGPT. For instance, teachers could orient students towards what AI is, what a chatbot is, ChatGPT capabilities, exemplary prompts to unlock ChatGPT capabilities and vocabulary and grammar for students to independently construct prompts to unlock ChatGPT capabilities. Given the iterative nature of the machine-in-the-loop writing process, students' may spend much time crafting prompts. Thus, exploring EFL students' perceptions during the prompt engineering phase of genre writing could inform more effective instruction to develop students' prompt engineering knowledge and skills.

In summary, previous research has suggested potential for ChatGPT to support EFL students' writing yet realizing that potential in the classroom may require not only effective writing instruction but also effective prompt engineering instruction. Furthermore, ChatGPT may convey benefits for EFL students' genre writing if ChatGPT does not replace the teacher, but rather students use ChatGPT alongside teacher instruction to ethically and effectively develop writing skills (Shaikh et al., 2023). Empirical research on EFL student perceptions is a means to evaluate student experiences when learning to write with ChatGPT a classroom context.

1.3. Student Perceptions about Learning to Write with ChatGPT

While ChatGPT shows potential to support students' writing, it is crucial to understand students' perceptions about learning to write with ChatGPT in their classroom context. Student perceptions encompass students' subjective assessment of their learning environment (e.g. curriculum; instructional methods and materials; and other services and contextual factors) (Biggs, 1999), and can significantly influence their learning behaviors, engagement, and ultimately academic achievement. This is because positive perceptions may foster a deep learning approach, whereas negative perceptions may facilitate a surface learning approach. The following literature review elaborates three aspects of student perceptions that are often examined to evaluate learning environments.

1.3.1. Motivation to Learn

Motivation to learn refers to students' desire and willingness to engage with the learning materials and activities (Keller, 1987). It is an important factor influencing how students approach and persist with learning tasks. It's especially crucial in the context of EFL writing, a challenging task demanding cognitive effort and continual practice. In the case of writing with ChatGPT, students' motivation can be influenced by their perceived usefulness and ease of use of the technology (Davis, 1989). Motivation to learn is often evaluated through questionnaire items. Hwang and Chang (2011) found their formative assessment-based mobile learning environment improved students' learning motivation toward the target content, the authors suggesting appropriate challenges had motivated students during the learning process. Shim et al. (2023) found their experiential chatbot workshop was instrumental in positively motivating their students to learn chatbot competencies. Kim and Lee (2023) found socio-economically disadvantaged Korean middle school students were far more motivated to learn about AI than students who were not socio-economically

disadvantaged. On the other hand, Hwang et al. (2013) found a concept map-embedded game did not have a significant impact on students' learning motivation when compared to a digital game without a concept mapping strategy. Alternatively, Jeon (2022) adopted qualitative methods to explore how chatbots affected EFL primary students' motivation to learn English, identifying chatbot affordances and limitations that facilitated and decreased, respectively, students' motivation to learn English through chatbots. Similarly, Chan and Hu (2023) asked Hong Kong university students open-ended questions to collect data on students' willingness to use ChatGPT and found most participants were motivated to use it, identifying several reasons.

1.3.2. Cognitive Load

Cognitive load theory (Sweller, 1988) posits that people's capacity to process information during learning is limited. In this way, a heavy cognitive load impedes learning but a manageable level of cognitive load facilitates it. Furthermore, cognitive load is a multidimensional concept comprising two components (Paas, 1992). Mental load refers to the load imposed by task demands. Mental effort refers to the amount of cognitive capacity allocated to address the task demands. Sweller et al. (1998) elaborated a cognitive architecture and proposed that when designing instruction, information should be organized and presented in a way to reduce cognitive load on working memory and increase knowledge stored in long-term memory. In evaluating innovative educational technology approaches, researchers have evaluated cognitive load in students through surveys and have found, for example, a formative assessment-based mobile learning environment could improve learning achievement with appropriate cognitive load (Hwang & Chang, 2011); and a concept map-embedded game also improved students' learning achievement and decreased their cognitive load (Hwang et al., 2013).

1.3.3. Satisfaction with Learning

Satisfaction is a basic measure of how participants react to a program or learning process. It can be characterized as either positive or negative. Importantly, although satisfaction with the learning process does not ensure learning, dissatisfaction may impede learning (Kirkpatrick & Kirkpatrick, 2006). When designing instruction, high satisfaction can validate standards of performance for future programs. Satisfaction is often evaluated quantitatively through questionnaires. For instance, Fisher et al. (2010) found that teachers who participated in a virtual professional development program were as satisfied as teachers who participated in an in person program; and that students were satisfied by the instruction from both groups of teachers. Shim et al. (2023) found 91% of their students were satisfied with an experiential learning chatbot workshop with no students indicating dissatisfaction. With regards to ChatGPT, Amaro et al. (2023) found that their cohort of Italian university students exhibited a high level of satisfaction during a guided interaction with ChatGPT. However, they also observed that satisfaction levels decreased because students became aware of ChatGPT's ability generate false information, particularly when students' awareness arose early in the interaction. Escalante and Barrett (2023) conducted a study in which 43 university EFL students received writing feedback from both human tutors and ChatGPT over a six-week period. The students reported similar levels of satisfaction with the feedback from both sources. Alternatively, in a mixed-methods study by Belda-Medina and Calvo-Ferrer (2022), 176 Spanish and Polish undergraduates interacted with three AI chatbots over a four-week period. Through analysis of survey data, the researchers found gender-related differences in levels of satisfaction and by analysis of students' written reports to open-ended questions, identified key factors for students' satisfaction.

To conclude, after the literature review, the overarching research question is operationalized into three questions, each addressing a particular aspect of student perception:

- RQ1: How does the use of ChatGPT in writing impact EFL students' motivation to learn about ChatGPT?

- RQ2: What is the cognitive load experienced by EFL students when writing with ChatGPT?

- RQ3: How satisfied are EFL students with the experience of writing with ChatGPT?

2. Methods

2.1. Context and Sample

This research used a convenience sample. Twenty-one students voluntarily participated in the study, where they were provided information about the study's objectives and tasks, their rights as participants, and the option to withdraw their participation at any point during the study. They were informed in English and Chinese language, verbally and in text, and were allowed to raise any questions or concerns about their participation with the researchers. No students declined participation.

The participants in this study were students from an all-girls secondary school in Hong Kong where the first author worked as an English as a Foreign Language (EFL) teacher. The school's students have academic achievement ranging from the 44th to the 55th percentile based on their results in the secondary school entrance exams (Lee & Chiu, 2017) and compared to peers in the school's geographic district. The demographic information of the sample is shown in Table 1. The average age was 13.6 years. Seven students lived in public housing, indicating a lower socio-economic status background in Hong Kong. Student's EFL writing proficiency was defined by their last EFL writing exam mark. As the

school's passing mark for a writing exam is 50 out of 100, the majority of students (n=11)

were mediocre writers scoring between 40 and 60.

Table 1. Participants' Demographic Information

Demographic Profile	Classification	Number	Percent
Gender	Male	0	0.00%
	Female	21	100.00%
Secondary Grade	1	2	9.52%
	2	16	76.19%
	3	2	9.52%
	4	1	4.76%
	5	0	0.00%
	6	0	0.00%
Age	12	1	5.00%
	13	9	45.00%
	14	8	40.00%
	15	1	5.00%
	16	1	5.00%
Public Housing	No	13	65.00%
	Yes	7	35.00%
Writing Exam Score	Below 40	3	15.00%
	40-50	4	20.00%
	50-60	7	35.00%
	60-70	4	20.00%
	70-80	1	5.00%

80-90	1	5.00%
90-100	0	0.00%

Note. One participant did not share demographic information besides gender and grade level.

60% of students (n=12) reported having used ChatGPT prior to the workshop, and 40% (n=8) reported not having used ChatGPT. However, only 25% of students (n=5) reported that they had used ChatGPT to complete English language homework, suggesting the majority of students have no experience with ChatGPT use cases in the EFL writing classroom.

2.2. Materials and Procedure

The study took place in the school's STEM classroom on July 5, 2023 and repeated on July 6. Six students attended on July 5 and 16 on July 6. The study's environment for learning to write with ChatGPT took the form of a human-AI creative writing workshop. Each workshop lasted one-hour, 45-minutes. Because writing with ChatGPT in the EFL classroom is novel, the authors' developed the workshop activities and materials by design-based research (DBR) (Wang & Hannafin, 2005), that is, a flexible and systematic methodology that can improve educational practice iteratively through design, development, implementation and analysis. The authors adopted an outcome-based learning design, that is, a framework for describing learning environments and learning activities (Conole & Wills, 2013). First, the researchers designed the workshop's purpose and intended learning outcomes (ILOs), that is, what students should achieve by the end of the workshop. Then the authors designed the learning activities, that is, basic units of interaction with or among learners. Table 2 summarizes the workshop design, which comprises its (1) title, (2) purpose, (3) ILOs, (4) learning activities, and (5) materials and resources. By evaluating student perceptions, the learning design can be improved for subsequent implementations.

Table 2. Workshop learning design

Title	How to use ChatGPT effectively to <mark>attempt</mark> a writing task
Time	1 hour 45 minutes
Purpose	To provide hands-on experience with prompt engineering
Intended learning outcomes	1. I understand what a chatbot and a prompt are. 2. I can access and use an app's chatbots. 3. I can engineer prompts so I get what I want. 4. I can integrate my words and chatbot output to write my best for a writing task.
Learning activities (minutes)	1. Introduction to Workshop, AI and Language Models (10) 3. Interacting with Chatbots (10) 4. Defining Prompt Engineering (10) 5. Task introduction (10) 6a. Opening POE on an iPad (5) 6b. Thinking-aloud Protocol introduction (5) 7. <mark>Attempting</mark> a Writing Task with ChatGPT (45) 8. Reviewing Concepts and Reflecting (15)

Materials	1. Generative AI tools on POE app on iPads
	2. Google Docs
	3. Shared Google Drive folder:
	   a. Contest website
	   b. Marking scheme
	   c. Pre- and post-workshop questionnaires
	   d. Workshop slidedeck

In the workshop, the authors introduced students to the genre of effective written communication with chatbots before introducing students to the writing task that they would attempt with ChatGPT. (1) The concept of *chatbots* was introduced using an inductive approach by showing a chatbot screenshot and asking students, "What are you looking at?" (2) Students were asked to interact with a chatbot, before asking students what this type of generative AI is and how to interact with it. (3) The features of chatbots were introduced, including turn-taking and memory. (4) The principles for chatbot prompting such as the garbage-in-garbage-out principle were introduced by showing a chatbot screenshot to students and asking, "What is a problem with this conversation?" (5) For students to take advantage of ChatGPT's novel capabilities and to get desired output, the concepts of *prompts* and *prompt engineering* were defined. The authors introduced different ChatGPT use cases for writing and example prompts for those use cases based on classmates' actual prompts and theoretical prompts from a literature review. The use cases included asking ChatGPT to act as a particular role, to act as a search engine, to analyze a text input, to answer a question, to auto-complete a text input, to explain its reasoning for its text output, to paraphrase a text input, to provide additional information to its text output, to summarize a text input, and to translate a text input. The authors did not introduce a use case of prompting ChatGPT to generate a complete composition that replaces human effort in writing. The instructional materials such as the slide deck (see Supplemental Material) were delivered in English by the first author. At the same time, the first author's colleague provided simultaneous spoken translation in Cantonese Chinese language.

After the introduction to prompt engineering for ChatGPT, students began writing with ChatGPT and other state-of-the-art chatbots on school-supplied iPads, on which the Platform for Open Exploration (POE) app was loaded. At the time of study, the app granted

free access to ChatGPT and five other chatbots (i.e. Sage, GPT-4, Claude+, Claude-instant, and Google-PaLM) that rely on commercial LMs hundreds of billions of parameters in size. Figure 3 shows the POE app interface on iPad from which students could select from the six chatbots.

Fig 3 The POE app interface with six chatbots accessible on the left side of the interface
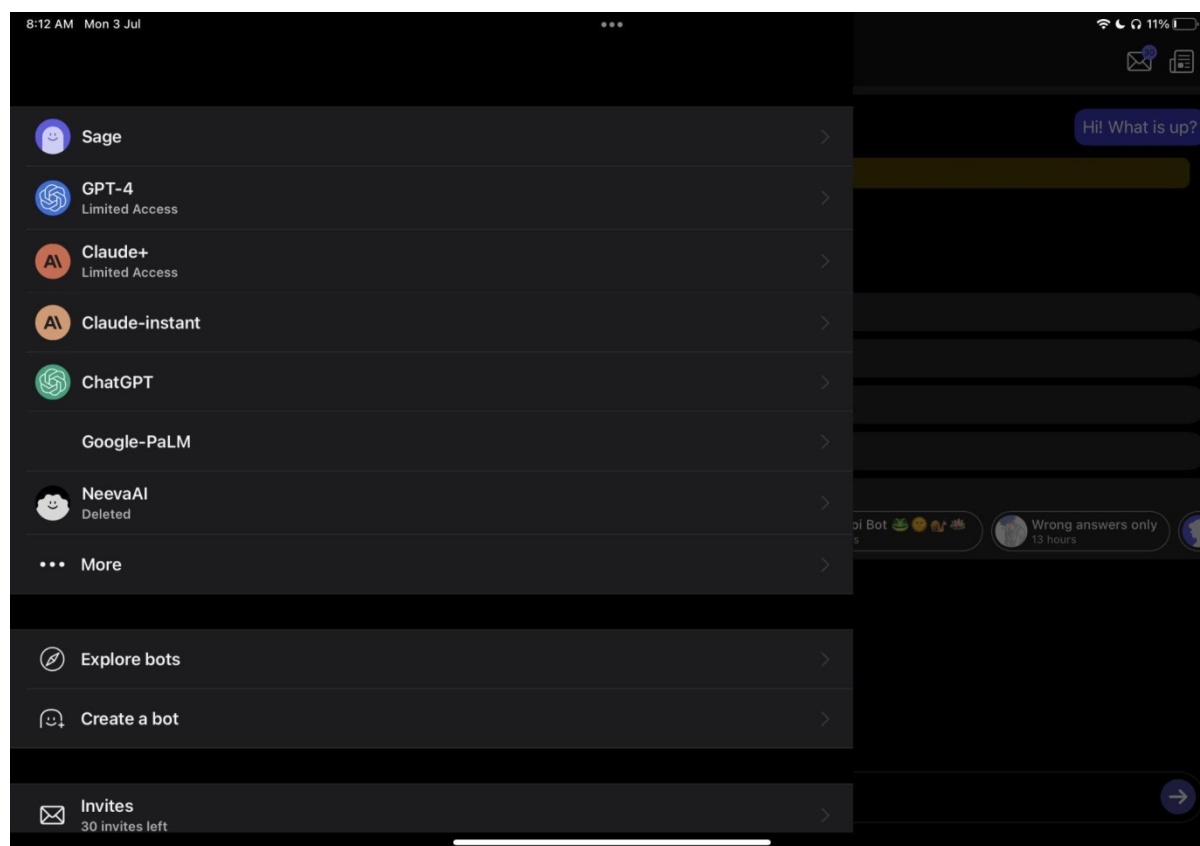
The students were given 45 minutes to attempt a writing task using ChatGPT and other POE chatbots. The task was designed for students to demonstrate the range of writing skills and genre assessed in their EFL school curriculum, and to compel students to engage ChatGPT. (1) Students were instructed to write either a feature article or a letter to the editor. Figure 4 shows the prompts selected by the authors, taken from the 2023 Hong Kong university entrance examination for the EFL subject area (HKDSE), writing paper, which Hong Kong secondary school students take in their final year. (2) Students were instructed to write no more than 500 words on Google Docs, using their own words and words generated from POE chatbots. Students could prompt any POE chatbot in any way possible, as many times as necessary and use any chatbot output. (3) Students were instructed to differentiate their own words from AI words in their writing by highlighting words from each chatbot in a

specific color. Figures 5 and 6 show a completed feature article and letter to the editor,

respectively, following the color-coding scheme.

Fig 4 Two writing tasks, one for a feature article and the other for a letter to the editor task,

each task comprising a topic heading, a picture and a written prompt

Task
For questions 1-2, choose ONE question. Write no more than 500 words on a Google Doc.

1. **Learning English through Sports Communication**

While tai chi is a popular activity in Hong Kong, it is less known in some parts of the world.

Write an article for *International Travel* magazine introducing the benefits of tai chi to tourists.

2. **Learning English through Poems and Songs**

Contestants in a song-writing competition complained when a song created by artificial intelligence (AI) software won first prize.

Write a letter to the editor on whether computer-generated songs should be allowed to compete in such competitions.

Fig 5 A feature article written by a student with words from the Google-Palm chatbot in grey

and the student's words in black

Tai chi is a popular activity in Hong Kong, it is less known in some parts of the world.Today, I will introduce the benefits of Tai chi to you.

Tai chi is a Chinese martial art that has been practiced for centuries. It is a gentle, flowing exercise that is said to have many health benefits. Tai chi can help to improve balance and coordination, reduce stress and anxiety, and strengthen the immune system.

First,Improved focus and concentration: Tai chi can help to improve focus and concentration by teaching you how to control your breath and relax your mind. When you are relaxed, you are more able to focus on the task at hand.
And then,Reduced stress and anxiety: Tai chi can help to reduce stress and anxiety by teaching you how to relax your muscles and mind. When you are stressed or anxious, it is difficult to focus on your writing. Tai chi can help you to manage your stress and anxiety so that you can focus on your writing.
Then, Improved flexibility and range of motion: Tai chi can help to improve flexibility and range of motion by teaching you how to move your body in a controlled and relaxed manner. When you are flexible and have a good range of motion, it is easier to write for long periods of time

Finall,hope everyone can know more about tai chi because tai chi can make we healthy

Fig 6 A letter to the editor written by a student and chatbots with the student's words in black, words from the Sage chatbot in purple, ChatGPT in green, GPT-4 in blue, Claude+ in red and Google-Palm in grey

Dear Editor,

I am writing to express my thoughts on a topic that has been stirring the emotions of music enthusiasts worldwide: the inclusion of computer-generated songs in music competitions. As advances in artificial intelligence (AI) continue to infiltrate various creative domains, it is essential that we critically evaluate the implications of such developments and determine whether they align with the spirit and objectives of these competitions.

While artificial intelligence is advancing rapidly and algorithms can now produce songs that sound highly realistic, they lack something fundamental that human songwriters possess—creativity, emotion, and life experiences. The songs that move us and stand the test of time are written by humans, drawing on their emotions, experiences, and creativity. Computer-generated songs may be good at mimicking styles and forms, but they do not evoke feelings in the same way.

Furthermore, allowing computer-generated music to compete in music competitions would be unfair to human musicians who have worked hard to hone their craft and develop their skills. Music competitions should be a platform for human creativity and talent to shine, not a showcase for technological advancements.

Of course, technology can be used to enhance and complement human creativity in music, such as in the production and recording process. However, it should not be used as a replacement for the essence of music itself.

In conclusion, I strongly believe that computer-generated music should not be allowed to compete in music competitions. Let us continue to celebrate the human creativity and artistry that makes music such a special and inspiring form of expression.

Sincerely,
Zita

The research team had monitored student progress as students attempted the writing task during the workshop. Students were not required to complete the task during the workshop as the students and research team had agreed on a task completion deadline after the workshop.

2.3. Data Collection

This mixed method study followed an embedded design (Creswell & Clark, 2007) where two sets of quantitative data and one set of qualitative data were collected in a workshop (see Table 3). In sum, the quantitative data from the pre-workshop questionnaire and the qualitative data from the think aloud protocols were collected to support quantitative data collected from the post-workshop questionnaire.

Table 3. Data Sources and Purposes

Sequence	Data source	Purpose	Related research questions
1	Pre-workshop questionnaire	To establish a baseline for student motivation to learn about ChatGPT	RQ1
2	Think aloud protocols	To explore immediate cognitive load in situ and to supplement other cognitive load data	RQ2
3	Post-workshop questionnaire	To retrospectively explore student motivation, cognitive load and satisfaction	RQ1, RQ2, RQ3

2.3.1. Pre-workshop Questionnaire

To collect data on students' learning motivation about ChatGPT before writing the task, a pre-workshop questionnaire, which also collected student background information, was developed. The learning motivation part comprised seven items with a six-point rating scheme (see Appendix), which was adapted from a measurement tool developed by Hwang and Chang (2011) to assess the motivation of fifth-grade primary school students towards a local culture course. The original scale has undergone thorough review, adoption, and adaptation by researchers (e.g., Cai et al., 2014; Huang et al., 2023) in diverse contexts to

evaluate students' motivation. In accordance with their procedures, the authors modified the scale by replacing the course name with terms relevant to learning ChatGPT to ensure content validity. Furthermore, a pilot study involving 46 participants was conducted to establish the construct validity of this questionnaire. The results of a confirmatory factor analysis (CFA) yielded favorable indices: $X2/df = 1.02$, $P(CMIN) = 0.422$, $CMIN/DF = 1.018$, root mean square error of approximation (RMSEA) = 0.022, and comparative fit index (CFI) = 0.989, thereby confirming the construct validity. The reliability of the learning motivation questionnaire in this study was found to be 0.95, indicating a high level of internal consistency. As students are taught Chinese language and English language literacy in school, the questionnaire was delivered in English language and traditional Chinese language text. Students completed the questionnaire at the workshop, before the delivery of instructional materials. The questionnaire was introduced to students verbally, in English and in Cantonese Chinese, most students' mother tongue. The research team monitored students while they completed the questionnaire and was available to answer any questions.

2.3.2. Think Aloud Protocols

To collect data on students' cognitive load during the prompt engineering phase of writing the task, thinking aloud (TA), a research method where a student speaks their thoughts and feelings during an activity (Ericsson & Simon, 1993), was utilized. Scholars (Charters, 2003; Yoshida, 2008) have claimed that think aloud protocols provide insights into students' cognitive load from demanding language tasks that can influence working memory and verbalization. In this way, students may not suffer great cognitive load if they can speak effortlessly and fluently.

The authors randomly sampled nine students for the think-aloud method. Not least because of the smaller sample, this data was supplementary to retrospective data collection

from a larger sample. Furthermore, the authors took a pragmatic view (Cotton & Gresty, 2006) to students' think-aloud protocols, actively moderating them. At the workshop, before students attempted the task, the selected students were briefed on think-aloud protocols in English language and Cantonese Chinese language; and the authors demonstrated a protocol. The fourth author administered the think-aloud protocols, spending six minutes with each student, video-recording students' iPad screens and iteratively asking students when they arrived at specific interaction points with a POE chatbot, (1) What do you think about this prompt? (visual cue: student has cursor in chatbot input box or is typing in chatbot input box) (2) What do you think about this output? (visual cue: chatbot has completed its output; and student is not typing anything) (3) How do you feel? (visual cue: student appears to have stopped answering question two) Students could answer in either or both English language and Cantonese Chinese language.

Of the nine think-aloud protocols video-recorded on July 5 and 6, only the five collected on July 5 had sound. These five protocols were transcribed for each thought, the sequence of the thought, the timestamp on the video recording, and the chatbot and prompt used at the time.

2.3.3. Post-workshop Questionnaire

A post-workshop questionnaire was the primary method to collect data on (1) learning motivation, (2) satisfaction, and (3) cognitive load. The seven learning motivation items were the same as those administered in the pre-workshop questionnaire except in the post-workshop items, the term "ChatGPT" had been replaced with the phrase "ChatGPT and other POE chatbots." For instance, item 1 in the post-workshop questionnaire was, "I think learning ChatGPT and other POE chatbots is interesting and valuable [□□□□□ ChatGPT □□□ POE □□□□□□□□□□□□]." These terms were replaced because by the end of the task, students

had been introduced to additional chatbots besides ChatGPT. The 14 satisfaction items were adapted from Fisher et al. (2010) (see Appendix). The eight, cognitive load items with a six-point Likert rating scheme were developed based on the measures of Paas (1992) and Sweller et al. (1998) (see Appendix). To ensure content validity, two experts with knowledge and expertise related to the construct of "satisfaction" were invited to evaluate the items for relevance, clarity, and comprehensiveness. Both experts confirmed the acceptability of the questionnaire, supporting its content validity. Additionally, the CFA on the pilot study showed that X2/df = 1.14, P(CMIN)=0.21, CMIN/DF =1.14, RMSEA =0.07, and CFI =0.97, confirming the construct validity. The Cronbach's alpha value for this satisfaction questionnaire was 0.98, indicating a high level of consistency.

The cognitive load questionnaire consists of five items related to mental load and three items pertaining to mental effort. The original questionnaire, as presented in Hwang et al. (2013), was initially developed to assess the mental load and mental efforts of sixth-grade primary school students engaged in a game-based learning activity. Since then, this scale has been extensively examined, adopted, and adapted by researchers in various learning contexts to explore students' cognitive load (e.g., Hsu, 2017; Dong et al., 2020). Following their methodological guidelines, minor adjustments were made by substituting the term "learning activity" with "workshop" to measure students' cognitive load, thereby ensuring content validity. Furthermore, the CFA conducted during the pilot study revealed that X2/df = 1.08, P(CMIN) = 0.37, CMIN/DF = 1.08, RMSEA = 0.058, and CFI = 0.99, thereby confirming the construct validity. The dimensions of mental load and mental effort exhibited high levels of internal consistency, with Cronbach's alpha values of 0.97 and 0.95, respectively. The questionnaire was delivered in English language and traditional Chinese language text. Students completed the post-workshop questionnaire at the end of the workshop on Google

Forms. Like the pre-workshop questionnaire, the post-workshop questionnaire was introduced to students verbally, in English and in Cantonese Chinese, the research team monitored students while they completed the questionnaire.

2.4. Data Analysis

To investigate EFL students' learning motivation, cognitive load, and satisfaction after their active participation in the study, the authors analyzed post-workshop questionnaire data, employing basic descriptive statistics, including mean, standard deviation, minimum, and maximum values.

To further investigate EFL students' learning motivation, the descriptive statistics were applied to the pre-workshop questionnaire data. In addition, the Wilcoxon signed-rank test was employed to assess students' motivation changes from pre-workshop to post-workshop, given that the data did not adhere to a normal distribution.

To further investigate EFL students' cognitive load, the authors analyzed students' think-aloud protocols, employing descriptive statistical measures such as the mean number of turns per student and the mean number of spoken words per turn, and standard deviation, minimum, and maximum values. This quantitative analysis was supplemented with representative quotes from students' think aloud protocols. This analysis provides a fine-grained perspective to students' cognitive load during the prompt engineering phase of writing with ChatGPT.

3. Results

3.1. Students' Learning Motivation

Twenty-one students had answered the post-workshop questionnaire and their median scores were found to be 35.00, within a range of 21 to 42. Twenty students had answered the pre-workshop questionnaire and their median scores for learning motivation among the

students were observed to be 33.50, with a range spanning from 26 to 42. The results of the

Wilcoxon signed-rank test, comparing the pre- and post-workshop learning motivation of the

20 EFL student cohort, are presented in Table 4. The result of the Wilcoxon signed-rank test

(Z = 1.085, p = 0.278) indicates no significant difference in learning motivation between the

pre- and post-workshop phases. However, it is noteworthy that the students exhibited an

enhanced motivation to engage with ChatGPT and other POE chatbots in the post-workshop

setting, as evidenced by a mean of 34.750 (SD = 6.604) in contrast to the mean of 33.850 (SD

= 5.631) in the pre-workshop context.

To gain deeper insights into the evolution of students' learning motivation, the authors

undertook a visualization of their ratings across specific items, illustrated in Figure 7. The

analysis reveals an increase in learning motivation across five of the seven items. For

instance, the average scores for items 1 and 2 exhibited a rise from 4.70 and 4.85 to 5.00,

reflecting a heightened perception of the value and interest associated with learning about

ChatGPT and other POE chatbots along with an augmented desire for further learning.

Remarkably, although item 3 experienced a decrease in score, it remained at a conspicuously

high level, with a rating of 5.00. In conclusion, the analysis of learning motivation partially

corroborates that interactions with ChatGPT and other POE chatbots in the context of EFL

writing have the potential to amplify students' motivation to advance their proficiency in

utilizing chatbots.

Table 4. The Result of the Wilcoxon Signed-rank Test Comparing the Pre- and Post-workshop Motivation

	N	Median (range)	Mean (SD)	Z	p
Pre-workshop	20	33.50 (26-42)	33.850 (5.631)		
				1.085	0.278
Post-workshop	20	35.00 (21-42)	34.650 (6.604)		

Note. One workshop participant did not complete the learning motivation questionnaire.

Fig 7 Students' learning motivation on each item with pre-workshop result on the left and post-workshop result on the right for each pair of bars
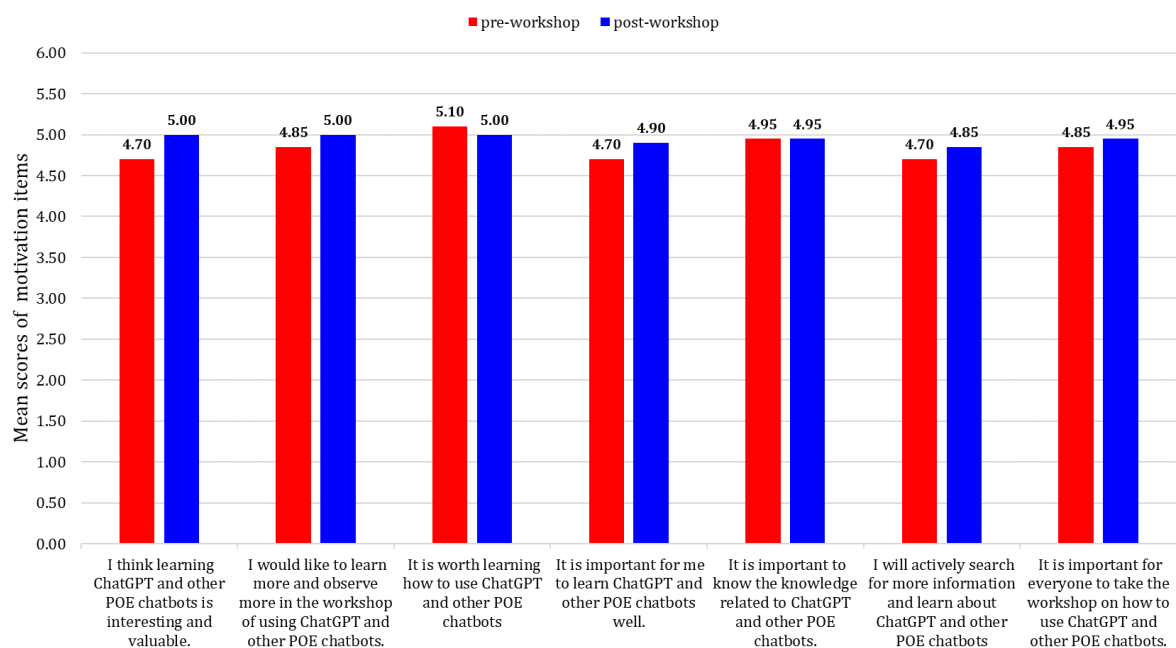
3.2. Students' Cognitive Load

Table 5 and Figure 8 describe data from the post-workshop questionnaire about EFL students' retrospective, self-reported cognitive load during the workshop. Intriguingly, students reported a relatively high level of cognitive load, with six out of the eight items returning an average rating of four out of six points. For instance, students attested to the challenging and effort-intensive nature of the workshop's questions and tasks, a sentiment

mirrored in their responses to items 2, 3, and 7. This cognitive load analysis suggests that students faced challenges when learning to write with ChatGPT and other POE chatbots to attempt a writing task.

Table 5. The Descriptive Statistics Regarding Students' Cognitive Load

Items	1	2	3	4	5	6	7	8
Mean	4.05	4.10	4.14	3.95	4.10	4.10	4.29	3.90
SD	1.56	1.41	1.35	1.43	1.58	1.45	1.38	1.64
Min	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00
Max	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00

Fig 8 Students' cognitive load scores on each item
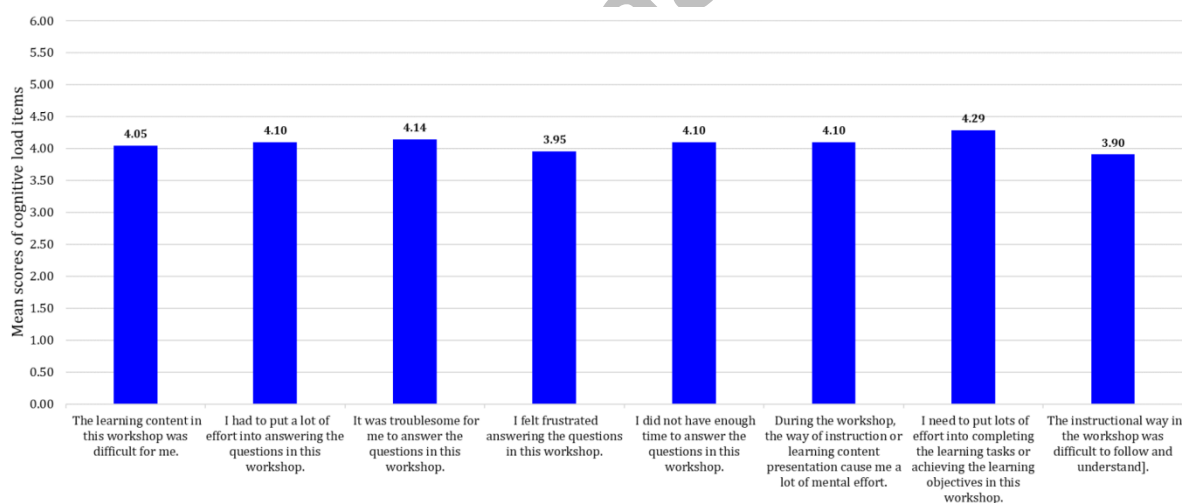
Analysis of five students' think aloud protocols during the prompt engineering phase of their writing with ChatGPT provides some corroborative evidence for students' high cognitive load. In general, students were not speaking effortlessly and freely. Figure 9 illustrates the number of spoken turns that each student took during a six-minute timespan. Students took on average 13 turns, with a range from five turns (n=1) to 17 turns (n=2). Figure 9 also illustrates the average number of words that each student spoke per turn. While

Students 1A, 2A and 3A wrote ChatGPT prompts exclusively in English language and delivered think aloud protocols exclusively in English language, Students 3A and 3B wrote ChatGPT prompts exclusively in Chinese language and delivered think aloud protocols almost exclusively in Chinese language. Therefore, the authors translated these students' words into the English language before preparing descriptive statistics. Students spoke on average 11 words per turn, with a range from five words per turn (n=1) to 25 words per turn (n=1).

Fig 9 A stacked bar chart for each student with each bar on the bottom showing the number of turns a student had taken and on the top showing the mean number of words spoken per turn
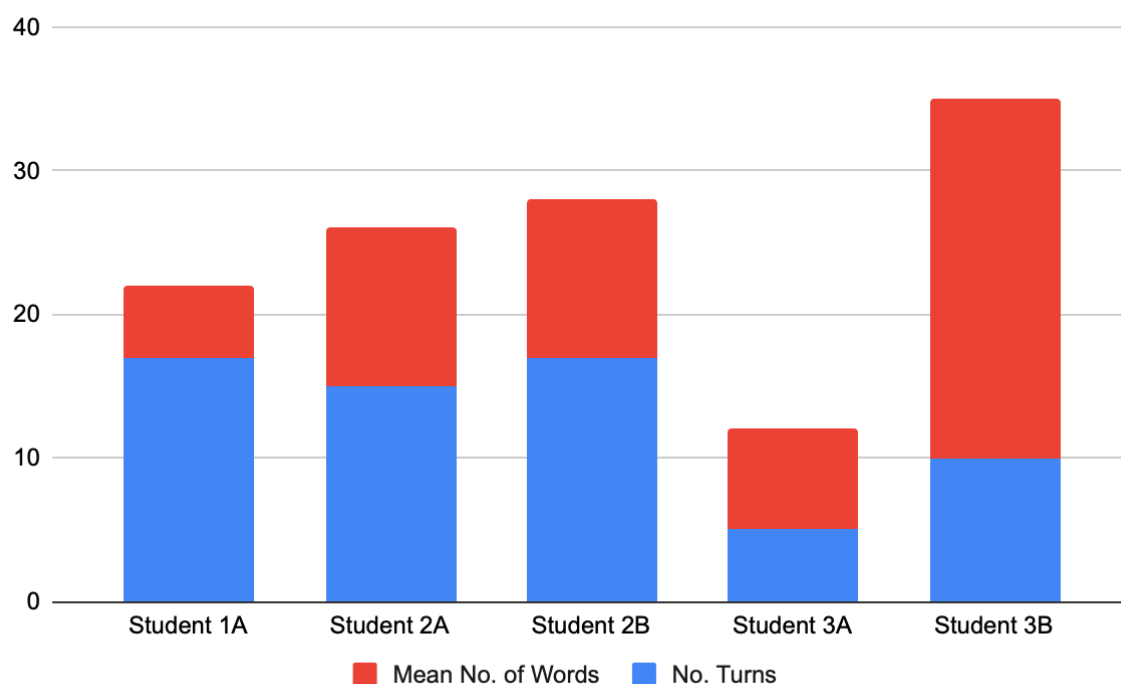
Table 6 shows representative turns for each student. Each student's turns represent a complete instance of a prompt engineering phase, showing their answers to the questions (1) what do you think about this prompt? (2) what do you think about this output? and (3) how

do you feel? Each student's turns were also selected as the most representative of the average

number of words that the student spoke per turn.

Table 6. Representative Turns

Student	Question	Representative Turn
1A	1	I think it's normal.
1A	2	It's not what I want.
1A	3	Slightly worse than before.
2A	1	I think the prompt is so a little bit amazing
2A	2	I think the output might be a little bit OK.
2A	3	I think the output is a bit OK for helping me to do some writing
2B	1	The prompt is better than the first bot and... is exactly what I'm looking for.
2B	2	The output is better than the first bot and I might use this bot like more than the others.
2B	3	Slightly happier.
3A	1	I find it difficult to evaluate
3A	2	What it (ChatGPT) answered was somewhat useful, as if it was actually discussing something with a person.
3A	3	Quite good
3B	1	I think I may have said too much, but it's not very clear. It (ChatGPT) may not be able to get it at the moment. Maybe if you ask me to repeat

|		it, it will get it.
3B	2	It (ChatGPT) will help me translate. It (The output) will be very useful.
3B	3	It (ChatGPT) may sometimes need to get clearer instructions. It can do what you want first, but they are all useful. Maybe I don't express them clearly.

3.3. Students' Satisfaction

Table 7 and Figure 10 offer insights into students' satisfaction concerning the workshop where they learned to write with ChatGPT. On the whole, students expressed a high level of satisfaction throughout the workshop, as all survey items garnered an average rating surpassing 5.40 on a 7.0-point scale. As delineated in Figure 10, students conveyed a robust sense of engagement and enjoyment in the workshop, as evidenced by their responses to items 8, 10, and 12. Furthermore, they reported a noteworthy level of focus, enthusiasm, and confidence in their ability to assimilate, retain, and apply the workshop content, exemplified by their responses to items 1, 2, 3, 4, 6, 7, and 11.

Table 7. The Descriptive Statistics Regarding Students' Satisfaction

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Mean	5.48	5.71	5.76	5.67	5.48	5.67	5.71	5.43	5.57	5.48	5.67	5.48	5.52	5.67
SD	1.29	1.15	1.26	1.28	1.29	1.20	1.19	1.36	1.25	1.33	1.15	1.33	1.33	1.24
Min	3.00	4.00	4.00	4.00	4.00	4.00	4.00	3.00	4.00	3.00	4.00	3.00	3.00	4.00
Max	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00

Fig 10 Students' satisfaction scores on each item
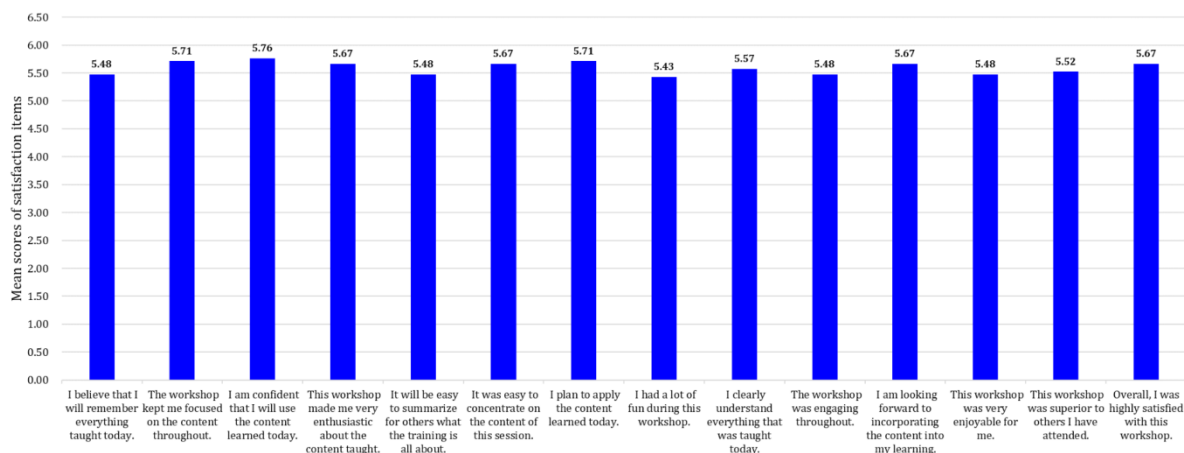
4. Discussion

This study has explored EFL students' perceptions about learning to write with ChatGPT in terms of students' motivation to learn, cognitive load, and satisfaction. The specific sample and context are Hong Kong EFL secondary students in a workshop where they were introduced to ChatGPT and prompt engineering, and attempted a 500-word writing task using ChatGPT for support. The results from the pre- and post-workshop questionnaires and from think aloud protocols provide insights into how using ChatGPT in writing impacts EFL students' motivation to learn about ChatGPT, EFL students' cognitive load when writing with ChatGPT and students' satisfaction with the experience of writing with ChatGPT. The following are the major findings.

4.1. Major Findings

The Wilcoxon signed-rank test results revealed no significant difference in students' motivation to learn about ChatGPT from pre- to post-workshop. However, a slight increase in the mean scores for post-workshop motivation suggests that students may have had a more favorable attitude towards learning about ChatGPT after engaging in the workshop activities. The widespread appeal of ChatGPT sparked considerable interest among students, motivating their voluntary participation in the workshop and explaining their initially high levels of

motivation, as in line with the findings of Chan and Hu (2023), who reported a willingness among most students to utilize ChatGPT. This highlights the potential of ChatGPT's novelty and interactive nature to stimulate motivation for EFL writing. The integration of ChatGPT into EFL writing bridges the gap between cutting-edge technology and students' academic learning, establishing relevance and potentially contributing to increased motivation. This integration is also consistent with prior research suggesting that the incorporation of innovative technologies can enhance students' motivation to engage with learning materials (Shim et al., 2023; Kim & Lee, 2023). Furthermore, ChatGPT's features, such as generating helpful content tailored to human needs and facilitating interactive conversations, provide students with a high level of satisfaction, as evidenced by their responses to satisfaction questionnaires. This satisfaction may further explain their enhanced motivation to learn. However, it is important to acknowledge the possibility of a "novelty effect" contributing to the slight increase in motivation, which refers to the heightened motivation or perceived usability of a technology due to its novelty or newness (Koch et al., 2018). To substantiate this hypothesis, a longitudinal study is necessary to explore how students' learning motivation may evolve as they become accustomed to ChatGPT.

Notably, this study has found students experienced heavy cognitive load when writing with ChatGPT. Specifically, the think-aloud protocol evidence suggests that students' experience heavy cognitive load during the prompt engineering phase of writing with ChatGPT. Additionally, the think-aloud protocol evidence suggests the heavy cognitive load stems from neither the demands of writing in English, as students could write ChatGPT prompts in Chinese language, nor from the medium of instruction, as verbal and written instructions were delivered in English language and Chinese language, nor from the think-aloud protocols as students could perform protocols in either English or Chinese language.

On the other hand, it is possible the heavy cognitive load stems from the basic cognitive processes associated with writing (Flower & Hayes, 1981) such as planning, drafting and reviewing. Another possibility is that the workshop's time constraint influenced students heavy cognitive load as Shim et al. (2023) suggested that novices unfamiliar with chatbots need more time to follow instruction and to keep pace in a workshop format. Alternatively, think-aloud protocols have been criticized as providing an artificial and incomplete view of cognitive activity during writing, although they can provide insights into writing and writing response practices (Hyland, 2019).

Importantly, the finding highlights the possibly high cognitive demands of integrating ChatGPT into the EFL writing classroom. Cognitive load theory posits that for effective learning to occur, instructional design should manage cognitive load to prevent overloading the learner's working memory (Sweller et al., 1998). Therefore, this study supports existing recommendations to optimize instruction for AI in the classroom so that students engage level-appropriate material and tasks yet are challenged to advance their cognitive boundaries (Walter, 2024). In the context of students writing with ChatGPT in the EFL classroom, teachers can lead students from comprehension-based tasks to controlled production to more communicative tasks (Nunan, 1989). Teachers can also provide materials on how to craft prompts and schematas by which students can evaluate output, because extensive reading must support EFL writing skills (Hyland, 2019). In addition, teachers may provide students with more time to engage in materials and tasks. By intentionally reducing EFL students' cognitive load, teachers better position their students to benefit from ChatGPT in the writing classroom.

Students expressed high satisfaction with the workshop overall. The analysis of satisfaction partially supports that engagement with ChatGPT and other POE chatbots in the

context of an EFL writing classroom fosters a highly gratifying and enriching educational experience for students. It corroborates prior ChatGPT research where students reported high satisfaction with guided ChatGPT interactions (Amaro et al., 2023) and with writing feedback from ChatGPT (Escalante & Barrett, 2023). The distinctive characteristics of ChatGPT may contribute to the observed high level of satisfaction, as previously mentioned. Firstly, ChatGPT is trained using the reinforcement learning from human feedback method (RLHF; Stiennon et al., 2020). This training approach modifies ChatGPT's language modeling objective, shifting its focus from predicting the next token on a webpage to providing helpful and safe responses based on user instructions (Ouyang et al., 2022, p. 2). Consequently, ChatGPT excels at generating responses that are aligned with human preferences and priorities. In the context of this study on EFL writing, ChatGPT effectively produces responses that cater to students' preferences and facilitate their writing process, thereby explaining their heightened satisfaction. Secondly, ChatGPT exhibits a conversational nature. Previous studies have shown that extended interactions with conversational agents can enhance users' overall experience (Jacq et al., 2016). Similarly, the students' frequent interaction with the human-like ChatGPT in this study can be considered a contributing factor to their elevated satisfaction. However, although high satisfaction is an important predictor of future engagement and positive learning outcomes, it alone does not guarantee learning.

4.2. Implications, Limitations and Future Research

This study has allowed educators to better understand the integration of ChatGPT into the English as a foreign language (EFL) writing classroom in terms of students' experiences and perceptions. Specifically, educators can better understand the potential of ChatGPT to enhance the EFL writing classroom by motivating and engaging students. Furthermore, a workshop format can be a suitable way to integrate ChatGPT into the EFL writing classroom.

However, educators should carefully scaffold instruction, especially in teaching prompt engineering skills, so as to manage students' cognitive load. Educators can consider an iterative design process of activities and instructional materials. Careful design ensures that ChatGPT use supports writing without overwhelming students. Educators may begin by identifying the target text type for students to write, adopting an explicit approach to writing that text type, mapping ChatGPT capabilities to that writing approach and developing genre writing and prompt engineering instructional materials, such as worksheets, and activities such as educators and students jointly constructing prompts. Implementing extensively scaffolded instruction to better integrate ChatGPT in the writing classroom will require more contact time.

Although this exploratory study provides a meaningful window into EFL students' perceptions about learning to write with ChatGPT, its limitations should be considered when interpreting the results. The sample size was relatively small, in terms of number of students, schools and instructional time, which may limit the generalizability of the findings. The sample was all female. Further research could involve larger, and more diverse samples, including males, to further validate the findings. Furthermore, research could explore how students' perceptions evolve with prolonged engagement. For instance, longitudinal studies could examine how students' ability to manage cognitive load improves with prolonged exposure to ChatGPT. Likewise, further research could explore differences in students' perceptions between two types of workshops. For instance, student perceptions from this study's workshop can be compared to those from a workshop with adjusted instruction aimed to reduce cognitive load, such as improved prompt engineering instruction. Another limitation is the reliance on self-report measures, which may not fully capture students' perceptions. Future research could incorporate additional measures of student perceptions,

such as additional observational data besides screen recordings to further validate self-report

measures.

4.3 Conclusion

This mixed-methods study explored Hong Kong secondary school EFL students' experiences and perceptions of learning to write a composition with ChatGPT's support in a workshop format. Key findings revealed that, although not statistically significant, students' mean motivation scores to learn about ChatGPT increased slightly from pre- to post-workshop, suggesting ChatGPT's potential to engage students. However, students reported high cognitive load in the workshop, notably when writing with a machine-in-the-loop. This highlights the need for educators to carefully scaffold instruction and activities to reduce students' cognitive load in the classroom. Nonetheless, students expressed high overall satisfaction with the workshop experience.

This study adds to the empirical research on EFL student experiences and perceptions of writing with ChatGPT. The findings provide insights for educators on the motivational benefits and cognitive demands of integrating ChatGPT in the writing classroom. Furthermore, the study proposes developing EFL students' prompt engineering skills alongside genre writing skills to optimize ChatGPT's support at different stages of learning to write a text type. Overall, the study advances how to enhance EFL classroom writing with ChatGPT integration. Future studies could involve larger, more diverse samples, explore longitudinal effects, and compare varied instructional designs for integrating ChatGPT in EFL writing.

Data Availability Statement

The data that support the findings of this study are available from the first author upon reasonable request.

References

Author 1 et al. (2023)

Author 3 et al. (2022)

Amaro, I., Barra, P., Della Greca, A., Francese, R., & Tucci, C. (2023). Believe in Artificial Intelligence? A User Study on the ChatGPT's Fake Information Impact. *IEEE Transactions on Computational Social Systems*.

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, *3*(2), Article 2. https://doi.org/10.25082/AMLER.2023.02.009

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

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

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher education research & development, 18*(1), 57-75.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Cai, S., Wang, X., & Chiang, F. K. (2014). A case study of Augmented Reality simulation system application in a chemistry course. *Computers in human behavior, 37*, 31-40. https://doi.org/10.1016/j.chb.2014.04.018

Calderwood, A., Qiu, V., Gero, K. I., & Chilton, L. B. (2020). How Novelists Use Generative

Language Models: An Exploratory User Study. *IUI '20: Proceedings of the 25th

International Conference on Intelligent User Interfaces*. ACM IUI 2020, Cagliari,

Italy.

Cao, S., & Zhong, L. (2023). *Exploring the effectiveness of ChatGPT-based feedback

compared with teacher feedback and self-feedback: Evidence from Chinese to English

translation* (arXiv:2309.01645). arXiv. https://doi.org/10.48550/arXiv.2309.01645

Chan, CK., & Hu, W. (2023). Students' Voices on Generative AI: Perceptions, Benefits, and

Challenges in Higher Education. *International Journal of Educational Technology in

Higher Education, 20*(43). https://doi.org/10.1186/s41239-023-00411-8

Charters, E. (2003). The Use of Think-aloud Methods in Qualitative Research An

Introduction to Think-aloud Methods. *Brock Education : A Journal of Educational

Research and Practice*, *12*, 68–82.

Clark, E., Ross, A. S., Tan, C., Ji, Y., & Smith, N. A. (2018). Creative Writing with a

Machine in the Loop: Case Studies on Slogans and Stories. *23rd International

Conference on Intelligent User Interfaces*, 329–340.

https://doi.org/10.1145/3172944.3172983

Conole, G., & Wills, S. (2013). Representing learning designs – making design explicit and

shareable. *Educational Media International*, *50*(1), 24–38.

https://doi.org/10.1080/09523987.2013.777184

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-

learning. *British Journal of Educational Technology*, *37*(1), 45–54.

https://doi.org/10.1111/j.1467-8535.2005.00521.x

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*. SAGE.

Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the research agenda. *Language learning*, *41*(4), 469-512. https://doi.org/10.1111/j.1753-6405.2007.00096.x

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. https://doi.org/10.2307/249008

Dong, A., Jong, M. S. Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in psychology*, *11*, 591203. https://doi.org/10.3389/fpsyg.2020.591203

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. The MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, *3*, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a

computerized professional development program on teacher and student outcomes.

Journal of teacher education, 61(4), 302-312.

https://doi.org/10.1177/0022487110369556

Hsu, H.-C. (2023). The effect of collaborative prewriting on L2 collaborative writing

production and individual L2 writing development. *International Review of Applied

Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2023-0043

Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter?.

Computers & Education, *106*, 137-149.

https://doi.org/10.1016/j.compedu.2016.12.007

Hwang, G. J., & Chang, H. F. (2011). A formative assessment-based mobile learning

approach to improving the learning attitudes and achievements of students.

Computers & Education, 56(4), 1023-1031.

https://doi.org/10.1016/j.compedu.2010.12.002

Hwang, G.-J., & Chen, N.-S. (2023). Editorial Position Paper: Exploring the Potential of

Generative Artificial Intelligence in Education: Applications, Challenges, and Future

Research Directions. *Educational Technology & Society*, *26*(2).

https://www.jstor.org/stable/48720991

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational

computer game for improving students' learning performance in natural science

courses. *Computers & Education*, 69, 121-130.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal

of Second Language Writing*, *16*(3), 148–164.

https://doi.org/10.1016/j.jslw.2007.07.005

Hyland, K. (2019). *Second Language Writing*. Cambridge University Press.

https://doi.org/10.1017/9781108635547

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., & Paiva, A. (2016, March). Building successful long child-robot interactions in a learning context. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 239-246). IEEE.

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *0*(0), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Keller, J. M. (1983). Motivational design of instruction. In C.M. Reigeluth (Eds.), *Instructional design theories and models: An overview of their current status* (1st ed., pp. 383-434), Routledge.

Keller, J. M. (1987). Strategies for stimulating the motivation to learn. *Performance and instruction*, *26*(8), 1-7. https://doi.org/10.1002/pfi.4160260802

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). *Design principles and architecture of a second language learning chatbot*. http://hdl.handle.net/10125/73463

Kim, S.-W., & Lee, Y. (2023). Investigation into the Influence of Socio-Cultural Factors on

Attitudes toward Artificial Intelligence. *Education and Information Technologies*.

https://doi.org/10.1007/s10639-023-12172-y

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating Training Programs: The Four Levels*.

Berrett-Koehler Publishers.

DOI:10.1016/S1098-2140(99)80206-9

Koch, M., von Luck, K., Schwarzer, J., & Draheim, S. (2018). The novelty effect in large

display deployments–Experiences and lessons-learned for evaluating prototypes. In

Proceedings of 16th European conference on computer-supported cooperative work-

exploratory papers. European Society for Socially Embedded Technologies

(EUSSET).

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and

Learning. *RELC Journal*, 00336882231162868.

https://doi.org/10.1177/00336882231162868

Lee, D., & Chiu, C. (2017). "School banding": Principals' perspectives of teacher

professional development in the school-based management context. *Journal of*

Educational Administration, *55*, 686–701. https://doi.org/10.1108/JEA-02-2017-0018

Li, M., & Zhang, M. (2023). Collaborative writing in L2 classrooms: A research agenda.

Language Teaching, *56*(1), 94–112. https://doi.org/10.1017/S0261444821000318

Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in

enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL

Faculty Members. *Education and Information Technologies*.

https://doi.org/10.1007/s10639-023-11917-z

Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge University Press.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback* (arXiv:2203.02155). arXiv. http://arxiv.org/abs/2203.02155

Paas, F. (1992). Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology*, *84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Reynolds, L., & McDonell, K. (2021). *Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm* (arXiv:2102.07350). arXiv. https://doi.org/10.48550/arXiv.2102.07350

Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the Usability of ChatGPT for Formal English Language Learning. *European Journal of Investigation in Health, Psychology and Education*, *13*(9), 1937–1960. https://doi.org/10.3390/ejihpe13090140

Shim, K. J., Menkhoff, T., Teo, L. Y. Q., & Ong, C. S. Q. (2023). Assessing the effectiveness of a chatbot workshop as experiential teaching and learning tool to engage undergraduate students. *Education and Information Technologies*, *28*(12), 16065–16088. https://doi.org/10.1007/s10639-023-11795-5

Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. F. (2020). Learning to summarize with human feedback. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural*

information processing systems 33: Annual conference on neural information processing systems 2020, neurIPS 2020, December 6-12, 2020, virtual (pp. 3008-3021).

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

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). 'To generate or stop generating response': Exploring EFL teachers' perspectives on ChatGPT in English language teaching in Thailand. *Learning: Research and Practice*, *0*(0), 1–15. https://doi.org/10.1080/23735082.2023.2257252

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Wang, F., & Hannafin, M. (2005). Design-based research and technology-enhanced learning systems. *Educational Technology Research & Development*, *53*, 1042–1629. https://doi.org/10.1007/BF02504682

Yoshida, M. (2008). Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research. *Selected Proceedings of the 2007 Second Language Research Forum*, 199–209. https://www.semanticscholar.org/paper/Think-Aloud-Protocols-and-Type-ofReading-Task%3A-The-Yoshida/d115ddfddd5a9aa044b0c92ed80c3ae69331e913

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3544548.3581388

Zhang, Z. (Victor), & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, *58*, 100784. https://doi.org/10.1016/j.asw.2023.100784

Appendix. Questionnaires

Motivation to learn

1	I think learning ChatGPT is interesting and valuable [□□□□□ ChatGPT □□□□□□□]
2	I would like to learn more and observe more in the workshop of using ChatGPT [□□□□□ ChatGPT □□□□□□□□□□□□ □□□□]
3	It is worth learning how to use ChatGPT [□□□□□□□□ ChatGPT]
4	It is important for me to learn ChatGPT well [□□□□□□□□□□ ChatGPT □□□]
5	It is important to know the knowledge related to ChatGPT [□□□□ ChatGPT □□□□□□□□]
6	I will actively search for more information and learn about ChatGPT [□□□□□□□□□□□□ □ ChatGPT]
7	It is important for everyone to take the workshop on how to use ChatGPT [□□□□□□ ChatGPT □□□□□□□□□□□□□□].

Cognitive load

Mental load

1	The learning content in this workshop was difficult for me [□□□□□□□□□□□□□□□□].
2	I had to put a lot of effort into answering the questions in this workshop [□□□□□□□□□□□ □□□□□□□□□□].
3	It was troublesome for me to answer the questions in this workshop [□□□□□□□□□□□□ □□□□].
4	I felt frustrated answering the questions in this workshop [□□□□□□□□□□□□□□□□□□].
5	I did not have enough time to answer the questions in this workshop [□□□□□□□□□□□□□ □□□□□].

Mental effort

6   During the workshop, the way of instruction or learning content presentation causes me a lot of mental effort [□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□].

7   I need to put lots of effort into completing the learning tasks or achieving the learning objectives in this workshop [□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□].

8   The instructional way in the workshop was difficult to follow and understand [□□□□□□□□□□□□□□□□□□□□].

Satisfaction with learning

1   I believe that I will remember everything taught today [□□□□□□□□□□□□□].

2   The workshop kept me focused on the content throughout [□□□□□□□□□□□□□□].

3   I am confident that I will use the content learned today [□□□□□□□□□□□□□□].

4   This workshop made me very enthusiastic about the content taught [□□□□□□□□□□□□□□□□□].

5   It will be easy to summarize for others what the training is all about [□□□□□□□□□□□□□□□□□].

6   It was easy to concentrate on the content of this session [□□□□□□□□□□□□□□□□□].

7   I plan to apply the content learned today [□□□□□□□□□□□□].

8   I had a lot of fun during this workshop [□□□□□□□□□□□].

9   I clearly understand everything that was taught today [□□□□□□□□□□□□□].

10   The workshop was engaging throughout [□□□□□□□□□□□□□□].

11   I am looking forward to incorporating the content into my learning [□□□□□□□□□□□□□□□□□].

12   This workshop was very enjoyable for me [□□□□□□□□□□□□□].

13   This workshop was superior to the others I have attended [□□□□□□□□□□□□□□□□□□].

14   Overall, I was highly satisfied with this workshop [□□□□□□□□□□□□□□□□□].