Journal of English Teaching
e-ISSN: 2622-4224 | p-ISSN: 2087-9628
http://ejournal.uki.ac.id/index.php/jet

Volume 12. Number 3, October 2026, pp. 437-455

The Effects of ChatGPT Prompt Engineering Instruction on EFL Writing Performance, Self-Regulation and Self-Efficacy

Yea-Ru Tsai[1]*

Article History:
Received: 31/08/2026
Revised : 22/09/2026
Accepted: 23/10/2026
Available Online: 31/10/2026

Keywords:
GenAI, prompt engineering, EFL writing instruction, self-regulation, self-efficacy

ABSTRACT
The integration of generative artificial intelligence (GenAI) into language pedagogy represents a profound shift in English as a Foreign Language (EFL) instruction. This study evaluated the effects of a structured ChatGPT prompt engineering curriculum on EFL undergraduates' writing performance, writing self-regulation, and writing self-efficacy. Participants were 68 undergraduate EFL students in Taiwan. The experimental group (EG, n = 31) received systematic instruction in ChatGPT prompt engineering, while the control group (CG, n = 37) used ChatGPT during conventional writing course without prompt instruction. Data were collected through pre- and post-intervention timed writing tasks, a self-regulation and self-efficacy questionnaire given to both groups before and after the 18-week intervention, and EG open-ended responses. Both groups improved in writing performance, but the EG significantly outperformed the CG in the posttest. The EG also reported significantly higher posttest self-regulation and self-efficacy, particularly in time management, revision, error correction, and confidence in writing on unfamiliar topics. Open-ended responses revealed intentional, self-directed uses of AI. These findings suggest that structured prompt engineering instruction helps EFL learners use AI strategically and that the pedagogical value of GenAI depends not on mere access to the technology but on deliberate instructional scaffolding.

[1] I-Shou University, Kaohsiung, Taiwan. Email: syrtsai@gmail.com *Corresponding author

INTRODUCTION

Writing is a crucial skill for academic, professional, and personal success, yet it presents significant challenges for English as a foreign language (EFL) learners. Students frequently encounter multiple obstacles in their writing development: persistent writer's block; difficulties with idea generation and organization; and challenges with grammatical accuracy, vocabulary selection, and textual coherence (Cheng & Zhang, 2021; Evmenova et al., 2024; Woo et al., 2025). These challenges are particularly acute in large classroom settings, where insufficient individualized feedback impedes students' writing development and motivation (Latifi et al., 2023).

The emergence of generative artificial intelligence (GenAI) and large language models (LLMs), particularly ChatGPT, represents a significant advance in providing personalized, interactive writing support (Fitria, 2023; Hsiao & Chang, 2024). Serving as intelligent writing assistants, these systems offer real-time suggestions for content development, structural organization, and linguistic refinement. Guo et al. (2022) demonstrated how chatbots can scaffold EFL students' argumentative writing, while Woo et al. (2023b) highlighted their effectiveness in reducing cognitive load and enhancing creative expression. These tools have shown particular promise in fostering meaningful engagement with writing tasks while supporting the development of advanced composition skills (C. Liu et al., 2023; Woo et al., 2025).

At the foundation of effective human-AI interactions lies prompt engineering—the strategic development and refinement of task-specific instructions that optimize AI outputs (P. Liu et al., 2023). Recent studies have demonstrated how well-constructed prompts can significantly enhance writing performance across multiple dimensions, from fundamental aspects such as grammatical accuracy and vocabulary usage to higher-order skills including critical analysis and argumentative coherence (Dang et al., 2022; Woo et al., 2024b). Moreover, as Bozkurt and Sharma (2023) argue, prompt engineering competencies increasingly represent essential skills for the AI era, facilitating effective human-AI collaboration across diverse professional contexts. However, research indicates that many EFL students encounter substantial difficulties with prompt engineering, which significantly limits their ability to leverage LLMs to their full potential (Zhou et al., 2023).

This study investigates the integration of ChatGPT prompt engineering writing instruction in EFL classrooms through the lens of Nazari and Saadi's (2024) two-layered model of prompt engineering. Their framework synthesizes foundational components (task, context, and instructions) with adaptable elements (role, audience, tone, and examples), establishing a comprehensive approach to scaffold learners' engagement with LLMs. Through the application of this model, this study examines the impact of systematic prompt-based interventions on three key dimensions: EFL students' writing performance, writing self-regulation, and writing self-efficacy, together with their reported use of prompt engineering strategies.

Although research on prompt engineering in language education is expanding rapidly, several gaps remain. First, much of the existing evidence on EFL learners' prompting comes from case studies, descriptive analyses, or short interventions without a comparison group (e.g., Woo et al., 2023a, 2025; Yacoub, 2026). Second, studies that have examined learner-internal outcomes have tended to focus on AI-related constructs such as AI self-efficacy and prompt engineering ability (Woo et al., 2026) rather than on writing-specific self-regulation

and self-efficacy. Third, few studies have contrasted structured, process-based prompt engineering instruction with the unstructured, ad hoc ChatGPT use that is increasingly common in writing classrooms, leaving it unclear whether the benefits of GenAI stem from access to the tool itself or from explicit instruction in how to use it.

In response to these gaps, the purpose of this study is twofold: first, to evaluate whether structured ChatGPT prompt engineering instruction leads to greater gains in EFL writing performance than ChatGPT-assisted instruction without prompt instruction; and second, to examine whether such instruction is associated with higher levels of writing self-regulation and writing self-efficacy. The following research questions guided the study:

RQ1 What is the effect of ChatGPT prompt engineering writing instruction on EFL students' writing performance?

RQ2a. What is the effect of ChatGPT prompt engineering writing instruction on EFL students' writing self-regulation?

RQ2b. t is the effect of ChatGPT prompt engineering writing instruction on EFL students' writing self-efficacy?

LITERATURE REVIEW

Prompt engineering and EFL writing

Prompt engineering has emerged as a critical area of inquiry in language education, particularly within English as a foreign language (EFL) contexts. In the context of LLM tools, a prompt comprises a set of structured instructions designed to guide the tool's output generation (White et al., 2023). Prompt engineering encompasses the strategic development and optimization of these instructions to elicit desired responses in artificial intelligence (AI) and educational settings (P. Liu et al., 2023). Through sophisticated input design, prompt engineering enhances LLM tools' effectiveness in achieving learning outcomes, fostering creativity, and facilitating problem solving through AI-generated outputs (Woo et al., 2024b).

The efficacy of natural language generation (NLG) tools is significantly influenced by prompt composition and structure (Woo et al., 2024a). Subtle variations in input length, keyword selection, or syntactic organization can substantially impact output quality. P. Liu et al. (2023) and Reynolds and McDonell (2021) demonstrated that well-constructed prompts enable AI systems to execute complex tasks, including creative writing and decision-making, with remarkable precision. However, the sensitivity of AI models to nuanced changes in prompt structure presents significant challenges, particularly for nonexpert users (Zhou et al., 2023). These complexities underscore the necessity of providing structured guidance and experiential learning opportunities in prompt creation within AI-assisted learning environments.

Empirical evidence demonstrates the transformative potential of prompt engineering in EFL writing across multiple dimensions: writing quality enhancement, creative development, and learner engagement (Rad et al., 2024). For example, Woo et al. (2023a) documented how iterative prompt refinement enabled students to develop more sophisticated drafts while engaging with higher-order cognitive skills. Huang et al. (2023) revealed that strategically designed prompts facilitated the production of more accurate and contextually appropriate texts, marked by enhanced grammatical precision and lexical diversity. The impact extends beyond technical proficiency to creative and analytical

capabilities. Woo et al. (2025) illustrated how iterative prompting enabled students to construct cohesive narratives, positioning AI as an active collaborator rather than a passive tool. Huang et al. (2023) emphasized how structured prompts encouraged critical thinking as students revised AI-generated content to meet academic standards.

Prompt engineering also bears on learners' cognitive load. Although structured prompts may free learners to focus on content development and organization, Woo et al. (2024c) found that Hong Kong secondary EFL students reported high cognitive load when writing with ChatGPT, particularly during prompt engineering, and called for carefully scaffolded prompting instruction. Moreover, prompt engineering promotes engagement and satisfaction by fostering interactive, iterative writing experiences. Tsai et al. (2024) noted that ChatGPT's immediate feedback heightened students' enthusiasm, while Woo et al. (2024c) reported that students were highly satisfied with learning to write with ChatGPT, even though their motivation did not change significantly. Tailored prompt engineering further enhances alignment with individual learning needs. Hegazy (2024) demonstrated that linguistically parameterized prompts helped students produce texts aligned with their proficiency levels. These findings highlight the multifaceted benefits of integrating prompt engineering into EFL instruction.

Prompt engineering instruction and writing self-regulation and self-efficacy

The integration of prompt engineering into EFL writing instruction has demonstrated significant potential not only in improving writing performance but also in enhancing students' self-regulation and self-efficacy in writing. Writing self-regulation refers to learners' ability to plan, monitor, and evaluate their writing processes, while self-efficacy pertains to their confidence in executing writing tasks effectively. These two constructs are crucial for fostering independent and skilled writers, particularly in the context of EFL learners who often face challenges in generating ideas, organizing content, and adhering to linguistic norms (Rahimi & Fathi, 2022; Woo et al., 2024c).

Research has shown that prompt engineering instruction can act as a scaffold to support the development of self-regulation strategies in writing. For instance, structured prompts help students break down complex tasks into manageable stages, enabling them to focus on goal setting, task planning, and iterative revision (Han & Hiver, 2018; Hwang et al., 2023). Huang et al. (2023) emphasized the role of explicit prompt strategies in facilitating cognitive engagement, as students learn to critically analyze AI-generated outputs and integrate them effectively into their work. Similarly, Dang et al. (2022) found that guided prompt creation encourages learners to reflect on their writing goals and adjust their approaches dynamically, which is a key component of self-regulation.

In addition to fostering self-regulation, prompt engineering instruction positively impacts students' writing self-efficacy. When learners use well-designed prompts, they experience immediate success in generating coherent and contextually appropriate texts, which reinforces their confidence in their writing abilities (Woo et al., 2023a). Nazari and Saadi's (2024) two-layered model of prompt engineering provides a theoretical basis for this phenomenon by emphasizing foundational components such as task and context alongside adaptable elements such as tone and examples. This comprehensive approach not only improves the quality of writing but also empowers learners by demonstrating the effectiveness of their inputs in influencing AI outputs.

Empirical studies illustrate these effects. Z. M. Liu et al. (2026) found that EFL students using a structured prompt framework reported higher levels of self-efficacy, as they perceived greater control over the writing process and recognized improvements in their outputs. Similarly, Hegazy (2024) highlighted that linguistically parameterized prompts aligned with students' proficiency levels helped them overcome apprehension about tackling unfamiliar topics, thereby boosting their confidence.

However, challenges remain in ensuring that prompt engineering instruction is accessible and effective for all learners. Zhou et al. (2023) noted that many students struggle with the difficulties of prompt creation, which can hinder their ability to fully leverage AI tools. To address this, pedagogical interventions that provide explicit training in prompt strategies are essential. By equipping learners with the skills to craft, evaluate, and refine prompts, educators can enhance both self-regulation and self-efficacy in writing tasks, ultimately supporting students in becoming more autonomous and proficient writers.

Structured versus unstructured use of ChatGPT in EFL writing

A distinction should be drawn between structured, process-based prompt engineering and the unstructured, ad hoc use of ChatGPT that typically occurs when learners are given access to the tool without guidance. In unstructured use, learners tend to issue brief, underspecified requests (e.g., "correct my essay" or "write an essay about the internet") and accept the default output with little scrutiny. Because the quality of AI output depends heavily on the specificity of the prompt (Sawalha et al., 2024; Zhou et al., 2023), such requests often yield generic text; and because the AI completes much of the cognitive work, learners risk passively copying and pasting AI-generated content and bypassing the planning, monitoring, and evaluating phases that constitute writing self-regulation (Han & Hiver, 2018). Unstructured use may therefore improve the surface quality of a given text without fostering the learner's own strategic control over the writing process. Structured prompt engineering, in contrast, requires learners to specify the task, context, role, audience, tone, and constraints of each request (Nazari & Saadi, 2024). Composing such prompts obliges learners to articulate their writing goals before consulting the AI (planning), to compare the AI output with those goals (monitoring), and to decide what to accept, modify, or reject (evaluating). When these prompting moves are mapped onto the stages of the writing process, AI assistance becomes a scaffold for self-regulated writing rather than a substitute for it, and successful, learner-controlled interactions with the AI provide mastery experiences that can strengthen writing self-efficacy (Woo et al., 2023a). This contrast between unstructured access and structured, process-based prompt instruction forms the basis of the comparison in the present study.

RESEARCH METHODS

Participants

The participants of the present study consisted of 68 undergraduate EFL students enrolled in two English writing classes at a university in southern Taiwan. One class was randomly assigned as the experimental group (EG, n = 31, 13 males and 18 females), while the other class served as the control group (CG, n = 37, 15 males and 22 females). Because intact classes were used, the study adopted a quasi-experimental pretest–posttest comparison-group design. Students' ages ranged from 20 to 24 years (M = 21.38, SD = 1.35). The two classes were taught by two professors, each with more than ten years of EFL

writing teaching experience. To minimize instructor-related variation, instruction was standardized across the two classes: both classes followed the same syllabus, weekly schedule, instructional materials, essay assignments, and assessment procedures, all hosted on parallel Moodle course sites; the two instructors met regularly throughout the semester to coordinate lesson content and pacing; and the only planned difference between the classes was the prompt engineering component delivered to the EG. In addition, writing performance was scored by two independent raters who were not the course instructors (see Writing performance evaluation), so that instructors' familiarity with their own students could not influence the outcome scores. Possible residual instructor effects are acknowledged as a limitation of the study. All participants had studied English for more than ten years. Ethical approval for data collection and student consent was obtained before the study's implementation.

Instructional design

ChatGPT prompt engineering instruction in the experimental group

The theoretical framework described in the literature review frames prompts as dynamic elements that guide learners through iterative, process-oriented stages of writing. This connection positions prompt strategies as a key scaffolding tool within the writing process, aiding students in achieving higher levels of cognitive and creative engagement with generative AI. The EG course was implemented in a computer lab with the support of the Moodle course management system (CMS). In this study, the EG course was designed based on Nazari and Saadi's (2024) model, systematically applied across various writing process stages to create a scaffolded framework for EFL instruction:

1. Pre-Writing: During brainstorming, prompts leveraging the Task and Context components guide learners to generate ideas. For example, "List three arguments for and against the use of AI in education" provides a structured entry point for argumentative writing.
2. Drafting: The Role and Examples elements are particularly useful at this stage, enabling learners to develop coherent and well-structured texts. A prompt such as "Act as a debate moderator and write an opening statement that summarizes these arguments" encourages learners to organize their ideas effectively.
3. Revising and Editing: The Tone, Audience, and Limits elements scaffold revision by encouraging learners to evaluate and refine their texts. For instance, "Rewrite this paragraph to appeal to an audience of policymakers" fosters critical thinking and stylistic experimentation.
4. Reflecting: Prompts can also support metacognitive reflection, such as "Explain how you improved your introduction based on the AI's suggestions." This activity deepens learners' understanding of the writing process and enhances their ability to self-monitor and evaluate their work.

By integrating Nazari and Saadi's (2024) model into EFL writing instruction, educators can scaffold learners' development at each stage of the writing process, fostering linguistic accuracy, creativity, and autonomy. This structured approach ensures that AI tools serve as effective mediators, enhancing both the quality of learners' writing and their engagement with the learning process.

ChatGPT-assisted writing instruction for the control group

Students in the CG received conventional English writing instruction with the support of ChatGPT. The objective of this course was to help students learn to express themselves coherently in academic English writing. This course provided samples of academic writing and appropriate practice materials for essay or report writing at an intermediate or advanced level. Comprehensive instruction in rhetoric and sentence structure was provided. Students used ChatGPT in online writing sessions to enhance their writing skills across different essay types, such as process essays, cause/effect essays, comparison/contrast essays, and argumentative essays. Specific skills of paraphrasing and summarizing were also taught. This course was implemented with the support of Moodle. The key difference was that no instruction for prompt strategies was offered to the CG, while the EG received specific prompt engineering instruction during each stage of the writing process. In other words, CG students used ChatGPT in an unstructured, self-directed manner, deciding for themselves when and how to prompt it, whereas EG students were taught to construct structured prompts and to use them purposefully at specific stages of the writing process.

Treatment Fidelity

Several measures were taken to keep the two instructional conditions distinct. Prompt engineering materials, including the explanations of Nazari and Saadi's (2024) prompt components and the stage-specific prompt examples, were posted only on the EG's Moodle course site; the CG's course site contained no prompt-strategy materials. In the CG, ChatGPT was introduced as a general writing aid that students could consult freely, but no class time was devoted to how prompts should be constructed. The CG instructor was briefed on the study design and refrained from teaching prompt strategies. Because the two classes were separate course sections that met at different times, opportunities for informal sharing of the EG's prompting techniques were limited. Nevertheless, the CG students' individual ChatGPT transcripts were not systematically collected, so informal diffusion of prompting techniques cannot be ruled out entirely. Importantly, any such diffusion would have made the CG more similar to the EG and would therefore have narrowed, rather than widened, the between-group differences. The effects reported below can thus be regarded as conservative estimates of the impact of explicit prompt engineering instruction.

Data collection methods and instruments

Writing performance evaluation

The writing performance measurement consisted of a pretest and a posttest in which the participants were asked to perform a writing task. For both tests, the students wrote an essay as part of their regular assignments. There were four essay assignments in the course. Essay 1 and essay 4 served as the pre- and posttests, respectively. Both tests were conducted as in-class timed writing assignments, requiring students to write an essay within two hours. In the pretest (essay 1), the participants described their expectations for college education or what they would like to learn at university. In the posttest (essay 4), they delivered their opinions about the impact of the internet on society. Two EFL teachers with at least ten years of teaching English composition evaluated the written work of all participants in the EG and CG. The two raters scored each essay independently and were blind to group membership.

To assess the students' writing performance, the rubric created by Rakedzon and Baram-Tsabari (2017) was adapted. The original tool is a 20-point writing assessment instrument that evaluates four key dimensions: content organization, vocabulary, spelling, and grammar accuracy, with each criterion scored on a 5-point scale, ranging from excellent to poor. In the present study, this rubric was adapted to a 100-point scale, with each of the four criteria scored out of 25 points, while retaining the original descriptors for nuanced evaluation. Content organization assesses the logical flow and connectivity of ideas, vocabulary examines the precision and appropriateness of word choice, spelling tracks the impact of orthographic errors, and grammar accuracy evaluates the technical precision of language use.

Comparability of the Pretest and Posttest Tasks

The pretest and posttest essays were embedded in the regular course assignments and therefore addressed different topics: the pretest essay (Essay 1) asked students to describe their expectations for college education, whereas the posttest essay (Essay 4) asked them to argue for their opinions about the impact of the internet on society. Because the two prompts differ in topic and rhetorical demand, three steps were taken to limit the influence of task differences on the interpretation of the results. First, both tasks were general-interest topics requiring no specialized knowledge and were administered under identical conditions (in class, timed, two hours). Second, the analytic rubric contains no genre-specific criteria; its four dimensions (content organization, vocabulary, spelling, and grammar accuracy) capture general rhetorical and linguistic competence that applies across essay types, and the same rubric and descriptors were applied to both essays. Third, and most importantly, the primary inference of the study rests on between-group comparisons at each testing point rather than on within-group pretest–posttest gains. Because the EG and CG wrote on exactly the same prompt at each testing point, any difference in prompt difficulty affected both groups equally and cannot account for the between-group difference observed at the posttest. Within-group changes from the pretest to the posttest, in contrast, may partly reflect differences in task difficulty and are therefore interpreted descriptively and with caution.

Inter-Rater Reliability

Before scoring, both raters were trained on the adapted rubric using sample essays that were not part of the dataset. Inter-rater reliability was estimated with the intraclass correlation coefficient (ICC; two-way random effects, absolute agreement). Agreement between the two raters was ICC = .82 for the pretest essays and ICC = .86 for the posttest essays, indicating good reliability (Koo & Li, 2016). Essays on which the two raters' total scores differed by more than 10 points were discussed until agreement was reached, and the mean of the two raters' scores was used in the analyses.

Scale of Writing Self-Regulation and Self-Efficacy

A questionnaire adapted from Han and Hiver (2018) was used, integrating two complementary constructs to assess students' perception of their writing self-regulation and self-efficacy before and after the intervention. The first section evaluates learners' strategies for managing their writing goals and processes. This writing self-regulation scale includes

nine items focusing on key aspects such as goal setting, task planning, time management, progress monitoring, feedback seeking, and maintaining motivation, all rated on a 5-point Likert scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). This section demonstrated strong reliability, with a Cronbach's alpha of .85. The second section assesses learners' confidence in their ability to perform various writing tasks. This seven-item scale addresses skills such as idea generation, organization, vocabulary use, grammar accuracy, error correction, tackling unfamiliar topics, and belief in improvement through practice. Also rated on a 5-point Likert scale, it showed a reliable Cronbach's alpha of .83. Together, the two subscales provide a comprehensive measure of EFL learners' perceptions of their writing self-regulation and self-efficacy. In addition, the EG's post-intervention questionnaire included three open-ended questions about their use of prompts (see Analysis of open-ended questions).

Data Collection Procedures

The intervention curriculum was delivered in two essay writing courses offered at the university where the participants were studying. The EG received ChatGPT prompt engineering writing instruction, whereas the CG received ChatGPT-assisted instruction with the same core learning materials as the EG, excluding the prompt engineering materials. Both classes were implemented in a Moodle CMS. Each course was offered three hours per week and lasted for 18 weeks. Following the course introduction, the writing pretest and the self-regulation and self-efficacy questionnaire were administered to both groups in week three. Students' essay 1 and essay 4 served as the pre- and posttest for the evaluation of writing performance. The same questionnaire was administered again to both groups during week 16. Consent forms were provided to all the participants for the survey.

Data Analysis Methods

Quantitative data were analyzed using SPSS 21. To address RQ1, independent-samples t tests were used to compare the EG's and CG's writing scores at the pretest and the posttest. To address RQ2a and RQ2b, independent-samples t tests were likewise used to compare the two groups' self-regulation and self-efficacy scores (total and item level) at each testing point. Levene's test was used to check the homogeneity of variance assumption; where it was violated, the t statistic for unequal variances (Welch's correction) was reported. Cohen's d was calculated as a measure of effect size and interpreted as small (0.20), medium (0.50), or large (0.80) (Cohen, 1988). Qualitative data from the open-ended questions were analyzed using a systematic thematic coding process. Written responses were thoroughly reviewed and coded by identifying meaningful elements, after which similar codes were grouped to identify patterns or themes. Frequencies were calculated based on the number of responses or stage mentions assigned to each category, as appropriate. To establish the trustworthiness of the coding, two researchers independently coded all responses using a jointly developed codebook; initial inter-coder agreement was 87%, and all disagreements were resolved through discussion until consensus was reached. The resulting themes were reviewed through peer debriefing with a colleague not involved in the study, and representative excerpts are reported verbatim so that readers can evaluate the interpretations.

FINDINGS

The results of writing performance (RQ1)

Descriptively, writing performance increased in both groups. The mean score of the EG increased from 70.47 (SD = 12.70) on the pretest to 78.71 (SD = 11.49) on the posttest, whereas the mean score of the CG increased from 66.73 (SD = 13.81) to 71.93 (SD = 14.33). An independent-samples t test indicated no significant difference between the EG (n = 31) and the CG (n = 37) at the pretest, t(66) = 1.15, p = .253, d = 0.28. At the posttest, the EG scored significantly higher than the CG, t(66) = 2.12, p = .037, d = 0.52, indicating a medium effect in favor of the EG after the implementation of the prompt engineering instruction.

Two features of these results support this interpretation. First, the between-group effect size increased from small at the pretest (d = 0.28) to medium at the posttest (d = 0.52), and the EG's mean gain (8.24 points) was larger than that of the CG (5.20 points). Second, because both groups wrote on the same topic at each testing point, the posttest difference cannot be attributed to differences in prompt difficulty. As noted in the Method section, however, the pretest and posttest essays addressed different topics, so the within-group gains are interpreted descriptively rather than as precise estimates of growth.

Analysis of Self-Regulation Scale (RQ2a)

Table 1. Descriptive Statistics and t Test for Posttest Self-Regulation by Group

					Independent-samples t test		
Self-regulation items	EG (n = 31)		CG (n = 37)				
Item	M	S.D.	M	S.D.	t	p (two-tailed)	*d*
1. I will set specific goals for improving my writing.	4.17	0.59	3.81	1.19	1.60	.116	0.37
2. I will plan my writing tasks before starting them.	4.13	0.68	4.03	1.28	0.43	.670	0.10
3. I will manage my time effectively while completing writing assignments.	4.47	0.78	3.58	1.25	3.45	.001**	0.83
4. I will monitor my progress during the writing process.	3.83	0.38	3.61	1.20	1.05	.300	0.24
5. I will revise my drafts to improve the quality of my writing.	4.37	0.72	3.83	1.21	2.13	.037*	0.53
6. I will use strategies to overcome challenges in writing.	4.67	0.61	4.22	1.46	1.67	.102	0.39
7. I will seek feedback to enhance my writing.	4.43	0.63	3.72	1.28	2.94	.005**	0.69
8. I will adapt my writing strategies based on the task requirements.	4.13	0.51	3.75	1.23	1.71	.094	0.40
9. I will stay motivated to complete my writing tasks, even when they are challenging.	4.27	0.74	3.72	1.28	2.06	.043*	0.51
Total	4.27	0.33	3.81	1.16	2.29	.027*	0.52

Note: * p < .05, ** p < .01, *** p < .001

Writing self-regulation was evaluated with a nine-item questionnaire administered before and after the instructional intervention. Independent-samples t tests were conducted to compare the EG and CG at the pretest and the posttest, respectively; Levene's test was

used to check the homogeneity of variance assumption, and Welch-corrected t values were reported where variances were unequal. In the pretest, the EG and CG did not differ significantly on the total self-regulation score, t = 1.59, p = .117, d = 0.39, or on eight of the nine items. The only baseline difference was found on Item 7 (t = 3.55, p < .001, d = 0.86), where the EG scored higher than the CG. Overall, the two groups' writing self-regulation scores were generally comparable prior to the intervention. At the posttest, the EG significantly outperformed the CG on the total self-regulation construct, t = 2.29, p = .027, d = 0.52, as presented in Table 1.

At the item level, the EG scored significantly higher than the CG on Item 3 (managing time), Item 5 (revising drafts), Item 7 (seeking feedback), and Item 9 (staying motivated), with effect sizes ranging from medium to large (d = 0.51 to 0.83). Because the EG had already scored higher than the CG on Item 7 (seeking feedback) at the pretest, and the between-group effect size for this item was smaller at the posttest (d = 0.69) than at the pretest (d = 0.86), the posttest difference on Item 7 is not attributed to the intervention. The differences on Items 3, 5, and 9, for which no baseline differences were found, provide clearer evidence of an intervention effect. At the level of the total scale, the between-group effect size increased from d = 0.39 at the pretest to d = 0.52 at the posttest.

Analysis of self-efficacy scale (RQ2b)

Independent-samples t tests were conducted to compare the EG and CG on the writing self-efficacy scale at the pretest and the posttest. At the pretest, there was no significant difference between the EG and CG on the total score (t = 1.35, p = .183, d = 0.31) or on any of the seven items, and the effect sizes were small. The nonsignificant baseline differences suggest that the two groups were comparable in writing self-efficacy before the intervention.

Table 2. Descriptive Statistics and t Test for Posttest Self-Efficacy by Group

	Independent-samples t test						
Self-efficacy items	EG (n = 31)		CG (n = 37)				
Item	M	S.D.	M	S.D.	t	p (two-tailed)	d
1. I know how to generate ideas for writing tasks.	4.06	0.89	3.78	0.85	1.32	.190	0.32
2. I can organize my ideas effectively in written form.	3.71	0.90	3.68	0.91	0.15	.878	0.04
3. I can use appropriate vocabulary to express my ideas clearly.	3.84	0.90	3.70	0.97	0.60	.553	0.15
4. I can apply grammar rules accurately in my writing.	4.03	0.87	3.41	0.83	3.02	.004**	0.74
5.I can identify and correct errors in my drafts.	4.58	0.50	3.49	0.87	6.19	<.001***	1.51
6.I feel confident in writing essays on unfamiliar topics.	3.74	0.77	3.08	0.83	3.37	.001**	0.82
7. I believe I can improve my writing skills with practice.	4.48	0.63	4.08	0.72	2.43	.018*	0.59
Total	4.06	0.57	3.60	0.67	3.04	.003**	0.74

Note: * p < .05, ** p < .01, *** p < .001

As shown in Table 2, the EG scored significantly higher than the CG on the total score at the posttest (t = 3.04, p = .003, d = 0.74), indicating a medium-to-large effect. At the

item level, significant differences emerged for item 4 (grammar accuracy), $t = 3.02$, $p = .004$, $d = 0.74$; item 5 (identifying and correcting errors), $t = 6.19$, $p < .001$, $d = 1.51$; item 6 (confidence writing on unfamiliar topics), $t = 3.37$, $p = .001$, $d = 0.82$; and item 7 (belief that writing can improve with practice), $t = 2.43$, $p = .018$, $d = 0.59$. Together with the pretest results, which showed no significant baseline differences, these findings indicate that the two groups were comparable in writing self-efficacy before the intervention. After the intervention, the EG scored significantly higher than the CG on the total score and on four of the seven items – most notably error identification/correction (Item 5) and confidence with unfamiliar topics (Item 6). The between-group effect size for the total self-efficacy score increased from $d = 0.31$ at the pretest to $d = 0.74$ at the posttest.

Analysis of open-ended questions

The post-intervention questionnaire contained three open-ended items concerning the students' use of prompts in EFL writing: (1) examples of effective prompts, (2) the writing stage at which prompts were most effective, and (3) how students balanced AI assistance with independent writing skill development. Open-ended responses were collected only from the experimental group (EG), as the prompt-strategy instruction and the accompanying writing-stage framework were unique to the EG's prompt engineering-assisted intervention; the control group (CG) did not receive structured prompt guidance or a defined set of writing stages to reference, so these questions had no equivalent basis in the CG.

Effective prompt examples

Students' example prompts were clustered into six major themes. The most frequently reported strategy was grammar, language, and style correction ($n = 8$), followed by comprehensive feedback requests that asked the AI to evaluate an entire essay across thesis clarity, argument strength, and organization in a single query ($n = 5$) and structured brainstorming templates that required the AI to generate several distinct points before synthesizing them into a unified conclusion ($n = 5$).

Respondent 2 explained that the comprehensive feedback prompt was effective because it "asked for detailed feedback rather than simple corrections" and helped identify weak arguments and clarify ideas. Respondent 14 described a wide range of concrete prompting uses spanning content, structure, mechanics, and genre. He asked the AI to generate examples of two competing argumentative structures so he could compare and choose between them and requested "professional transition words" and synonyms to avoid overusing simple words. He also extended prompting by requesting around ten hook-sentence options for essay openings, selecting the best himself. Several respondents associated AI with a higher-level function by scaffolding idea generation rather than requesting isolated corrections. Respondent 15 mentioned that AI helped him learn to "organize his writing more systematically", turning scattered ideas into coherent paragraph or essay structures.

Writing stage of greatest perceived benefit

Regarding the five stages presented in the questionnaire (Planning, Drafting, Revising, Editing, Reflecting), no student selected Editing or Reflecting as the stage of greatest benefit. Some students named two stages jointly. Revising was identified most frequently (10 of 22

mentions), primarily because students felt less confident in grammar and organization before submission. Respondent 8 pointed out that "my ability to revise the word and grammar is very terrible," and Respondent 12 gave an identical rationale. Respondents 2 and 3 offered a fuller account, mentioning that by the revising stage, they already had their main ideas written down but needed help "improving the organization, clarity, and strength of their arguments."

Planning and Drafting were each named less often, with reasons centered on idea creation and getting started, respectively. Respondent 6 explained that Planning is "the stage where I need to brainstorm several ideas, but with AI's help, I can have more directions," while Respondent 4 said Drafting "can let my ideas show out." The finding that Revising was the largest category suggests that for these students, the AI's greatest perceived value lay in polishing and strengthening an already-drafted text rather than generating it.

Balancing AI assistance and independent skill development

Six themes emerged around how students managed dependence on AI, all reflecting deliberate strategies for preventing overreliance. The dominant strategy was drafting independently first and consulting AI afterward only to check or refine the work (n = 6). Respondent 13 described a three-step cycle: "I first write on my own, then use AI to check and improve it, and finally revise it myself", plainly framing this as a way to improve both writing skill and learning. Respondent 8 extended the same logic to correcting, turning to AI "only when I encounter difficulties" and stating, "I don't copy everything directly from AI."

A second cluster regarded balance as a general principle, using AI "as a tool for guidance rather than having it do all of the work". A related but distinct strategy restricted AI use to a narrow function, mainly vocabulary lookup, leaving the rest of the writing process untouched by AI. Two respondents (Respondents 6 and 12) reversed the dominant order, consulting AI first and then deliberately practicing writing without it so as not to "totally copy AI's answer". A further unique response from Respondent 9 explicitly positioned the AI as an instructional resource rather than a corrective or generative resource, stating, "Ask him to teach me." Furthermore, Respondent 15 warned against overdependence on AI, recommending its use only when genuinely necessary and advocating AI literacy to preserve students' critical and creative thinking.

DISCUSSION

This study investigated the effects of ChatGPT prompt engineering instruction on EFL learners' writing performance and writing self-regulation and self-efficacy. It was found that while both groups improved in writing performance, the EG significantly outperformed the CG in the posttest ($p = .037$, $d = 0.52$). This is consistent with the findings of Duong and Chen (2025), who reported that AI chatbots greatly enhanced writing ability for both high-level and low-level participants. Additionally, Zhang et al. (2025) found that hybrid feedback (AI + human) produced more improvement than AI feedback alone, especially in higher-order abilities such as critical thinking and organizing. These findings suggest that explicit prompt engineering instruction was a key factor in the EG's superior learning outcomes. This is strongly supported by Algobaei and Alzain (2026), who discovered that EFL-tailored prompts led to significantly higher quality feedback, therefore enhancing students' learning quality. Recent research suggests that AI prompt engineering can support EFL writing

development, but its effectiveness depends on how deliberately learners are guided to interact with GenAI. Rather than treating GenAI as an automatic text generator, scholars conceptualize prompting as a form of digital literacy involving task interpretation, rhetorical decision-making, revision, and critical thinking. Direct evidence from EFL contexts indicates that students use prompts for several purposes during writing. Woo et al. (2025) found that secondary learners prompted AI to overcome writer's block, generate or expand ideas, and improve developing stories. Classroom-based research also reports positive effects on writing-related outcomes. For example, Yacoub (2026) implemented a three-session prompt engineering instruction with 28 university EFL students who showed significant improvement in prompt clarity, specificity, and purpose.

Concerning writing self-regulation and self-efficacy, the EG and CG showed comparable levels at the pretest. Following the prompt engineering intervention, the EG reported significantly higher levels of both constructs at the posttest. For self-regulation, the EG scored higher than the CG on the total scale ($t = 2.29$, $p = .027$), with the largest item-level differences in managing time, revising drafts, and staying motivated ($d = 0.51–0.83$). For self-efficacy, the EG also outperformed the CG on the total score ($t = 3.04$, $p = .003$), with the strongest effects on identifying/correcting errors, grammar accuracy, and confidence writing on unfamiliar topics. Taken together, these findings indicate that prompt engineering instruction strengthened both students' prompt strategies during writing and their confidence in specific writing subskills, particularly in error-correction and revision-related behaviors.

Two aspects of the baseline data deserve attention in interpreting these effects. First, although none of the pretest differences in writing performance, total self-regulation, or total self-efficacy reached statistical significance, the EG scored slightly higher than the CG on all three measures at the outset ($d = 0.28$, 0.39, and 0.31, respectively), and significantly higher on one self-regulation item (seeking feedback). For this reason, the posttest difference on that item is not treated as evidence of an intervention effect. Second, and more importantly, the between-group effect sizes increased consistently from the pretest to the posttest on all three outcomes: from $d = 0.28$ to 0.52 for writing performance, from 0.39 to 0.52 for self-regulation, and from 0.31 to 0.74 for self-efficacy. If the posttest differences merely reflected initial differences between the classes, the effect sizes would be expected to remain stable; the widening of the gap, together with the finding that the largest item-level effects emerged on items with no baseline differences (e.g., identifying and correcting errors, $d = 1.51$; managing time, $d = 0.83$), suggests that the EG's advantage grew during the intervention. A plausible mechanism for this pattern lies in how structured prompting manages cognitive load during revision. The revising stage requires writers to keep the global goals of a text in mind while simultaneously diagnosing problems in content, organization, and language, which places heavy demands on working memory, particularly for EFL writers whose linguistic resources are still developing. The EG's training in specifying the task, audience, tone, and limits of a revision request allowed students to delegate the diagnostic search to the AI in a targeted way (e.g., asking it to evaluate thesis clarity and argument strength) while retaining the evaluative decision of what to accept or reject. This division of labor eases the high cognitive load that prompt engineering can impose on EFL writers when it is not scaffolded (Woo et al., 2024c) and is consistent with the EG students' identification of revising as the stage at which AI was most beneficial. It may also explain why the largest self-efficacy effect

emerged for identifying and correcting errors and why revising drafts was among the significant self-regulation items: structured prompts made revision more manageable, and repeated successful revision cycles provided mastery experiences that strengthened students' confidence in their own revising ability.

AI-assisted instruction is generally associated with improvements in EFL writing self-regulation, especially in time management and motivation (Zhang, 2025), and in self-efficacy, with the strongest confidence gains identified in error correction and handling unfamiliar or complicated topics (Hongxia & Razali, 2026). A small but growing number of studies support the finding that explicit prompt engineering instruction, rather than solely AI-assisted instruction, promotes self-regulatory behavior in EFL writing. Sawalha et al. (2024) provide cross-disciplinary evidence that prompting strategies are associated with differences in AI response quality and student learning. Although their study was not conducted in an EFL context, the findings suggest that strategic prompting may support EFL learners' reading comprehension, writing development, revision, and metacognitive engagement by encouraging them to clarify goals, request targeted feedback, and evaluate AI-generated responses. A parallel pattern emerged for self-efficacy: the EG's significantly higher total self-efficacy ($t = 3.04$, $p = .003$, $d = 0.74$), and particularly its largest effects in error correction and confidence with unfamiliar topics, align with Woo et al.'s (2026) findings that prompt engineering instruction considerably improves students' AI self-efficacy and confidence in evaluating AI output.

These quantitative improvements are supported by the EG's open-ended responses. When describing effective prompts, students most frequently mentioned using AI for grammar, language, and style correction ($n = 8$), followed by requests for comprehensive feedback ($n = 5$) and structured brainstorming or idea-generation prompts ($n = 5$). Content/evidence strengthening, organizational scaffolding, and topic research were mentioned less frequently. The view that the intervention promoted intentional, self-directed interaction with AI rather than passive dependence on it is reinforced by these self-reported practices, which are consistent with the item-level self-regulation differences in revising drafts and managing the writing process reported above.

CONCLUSION

Collectively, these convergent findings suggest that prompt engineering instruction may help sustain learner confidence as writing-task complexity increases while also cultivating proactive, self-regulated strategies relevant to sustained writing development (Song & Song, 2023).

The findings of this study carry several implications for EFL writing pedagogy in the age of generative AI. First, the results suggest that explicit instruction in prompt engineering, rather than mere access to AI tools, shapes learning outcomes. The EG's superior performance across writing quality, self-regulation, and self-efficacy indicates that the pedagogical value of GenAI is mediated by learners' capacity to interact with it strategically. This aligns with the growing conceptualization of prompting as a form of digital literacy involving task interpretation, rhetorical decision-making, and revision (Woo et al., 2024b; Yacoub, 2026), rather than as an incidental technical skill. Writing curricula should therefore incorporate structured prompt literacy instruction as a core component, rather than assuming that such competence will develop through unguided exposure to AI tools. Second, the

emergence of a consistent "draft-first-then-consult-AI" pattern among EG students, together with the concentration of perceived AI usefulness at the revising stage, points toward a process-oriented integration of AI into writing pedagogy. Rather than regarding AI as a tool available throughout the writing process without constraints, instructors may consider scaffolding assignments to require independent drafting prior to AI consultation, thereby preserving learners' engagement in idea generation and early-stage composing while channeling AI support toward revision and refinement, treating AI as a supplementary reviewer rather than a substitute for independent effort (Sawalha et al., 2024). Taken together, these implications underscore that the pedagogical benefit of GenAI in EFL writing instruction is not inherent to the technology itself but contingent on how deliberately it is scaffolded, sequenced, and taught within the writing curriculum.

While providing critical insights, this study has several limitations that should be considered when interpreting the findings. First, the study used two intact classes taught by different instructors; although instruction was standardized through a common syllabus, materials, assignments, and Moodle platform, and writing was scored by independent raters, residual instructor effects cannot be completely ruled out. Second, the pretest and posttest essays addressed different topics; the conclusions therefore rest primarily on between-group comparisons at each testing point, and the within-group gains should be interpreted with caution. Third, the CG's ChatGPT interactions were not systematically logged, so informal diffusion of prompting techniques cannot be excluded, although such diffusion would have attenuated rather than inflated the observed differences. Fourth, the self-regulation and self-efficacy measures relied on self-report, and the small baseline differences between the groups were examined through effect sizes rather than controlled statistically. Fifth, the intervention lasted a single semester. Future research should therefore assign instructors across conditions or have a single instructor teach both groups, employ counterbalanced or parallel writing prompts, apply covariance-based analyses that adjust for pretest scores, and collect ChatGPT interaction logs from all participants. Such logs would allow researchers to verify treatment fidelity directly by coding whether CG students' prompts contained the components taught to the EG (e.g., task, context, role, audience, tone, and limits), and to examine how prompt quality, the writing stage at which the AI is consulted, and the extent to which AI-generated text is adopted verbatim relate to writing gains and to self-regulated AI use. Longitudinal designs are also needed to evaluate the long-term sustainability and transferability of prompt-mediated self-regulation skills across writing genres. Additionally, expanding these studies to other EFL settings (e.g., EAP or ESP) will broaden external validity, while integrating multimodal learning analytics will capture the real-time, temporal dynamics of student–AI co-regulation.

ACKNOWLEDGMENTS

This research was supported in part by grants from the National Science and Technology Council (NSTC), Taiwan, under Grant [NSTC 114-2410-H-214-007].

REFERENCES

Algobaei, F., & Alzain, E. (2026). Prompt engineering for non-native English learners: A generative AI approach to personalised language feedback. *Social Sciences & Humanities Open*, 13, 102341. https://doi.org/10.1016/j.ssaho.2025.102341

Bozkurt, A., & Sharma, R. C. (2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world. *Asian Journal of Distance Education*, 18(2), i–vii.

Cheng, X., & Zhang, L. J. (2021). Sustaining university English as a foreign language learners' writing performance through provision of comprehensive written corrective feedback. *Sustainability*, 13(15), 8192.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dang, H., Mecke, L., Lehmann, F., Goller, S., & Buschek, D. (2022). How to prompt? Opportunities and challenges of zero- and few-shot learning for human-AI interaction in creative applications of generative models (arXiv:2209.01390). arXiv. https://doi.org/10.48550/arXiv.2209.01390

Duong, T. N. A., & Chen, H. L. (2025). An AI chatbot for EFL writing: Students' usage tendencies, writing performance, and perceptions. *Journal of Educational Computing Research*, 63(2), 406–430. https://doi.org/10.1177/07356331241312363

Evmenova, A. S., Regan, K., Mergen, R., & Hrisseh, R. (2024). Improving writing feedback for struggling writers: Generative AI to the rescue? *TechTrends,* 68(4), 790–802. https://doi.org/10.1007/s11528-024-00965-y

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching*, 12(1), 44–58.

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.asw.2022.100666

Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. Journal of Second Language Writing, 40, 44–59. https://doi.org/10.1016/j.jslw.2018.03.001

Hongxia, H., & Razali, A. B. (2026). Impact of AI-assisted peer feedback on EFL students' self-regulated learning, self-efficacy, and motivation in English academic writing. *International Journal of Instruction*, 19(2), 163–182. https://doi.org/10.29333/iji.2026.1929a

Hegazy, H. M. (2024). A linguistically developed prompt engineering parameters model for enhancing AI's generation of customized ESL reading texts. *Faculty of Education Journal Alexandria University*, 34(3), 501–566.

Hsiao, J. C., & Chang, J. S. (2024). Enhancing EFL reading and writing through AI-powered tools: Design, implementation, and evaluation of an online course. *Interactive Learning Environments*, 32(9), 4934–4949.

Huang, H.-W., Teng, D. C. E., & Wu, R. (2023). Teaching chatbot prompt strategies in EFL essay writing instruction. In *2023 5th International Workshop on Artificial Intelligence and Education (WAIE)* (pp. 12–17). IEEE.

Hwang, W.-Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and recognition technologies to facilitate English as foreign language writing for supporting personalization and contextualization in authentic contexts. *Journal of Educational Computing Research*, 61(5), 1008–1035.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Latifi, S., Noroozi, O., & Talaee, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 31(2), 655–669. https://doi.org/10.1080/10494820.2020.1799032

Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31(9), 5614–5632.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35. https://doi.org/10.1145/3560815

Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2026). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 39(3), 466–490. https://doi.org/10.1080/09588221.2024.2389923

Nazari, M., & Saadi, G. (2024). Developing effective prompts to improve communication with ChatGPT: A formula for higher education stakeholders. Discover Education, 3(1), 45. https://doi.org/10.1007/s44217-024-00122-w

Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32(9), 5020–5040.

Rahimi, M., & Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: A mixed methods study. *Computer Assisted Language Learning*, 35(9), 2627–2674.

Rakedzon, T., & Baram-Tsabari, A. (2017). To make a long story short: A rubric for assessing graduate students' academic and popular science writing skills. *Assessing Writing*, 32, 28–42.

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In Extended abstracts of the *2021 CHI Conference on Human Factors in Computing Systems* (Article 314, pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3411763.3451760

Sawalha, G., Taj, I., & Shoufan, A. (2024). Analyzing student prompts and their effect on ChatGPT's performance. *Cogent Education*, 11(1), Article 2397200. https://doi.org/10.1080/2331186X.2024.2397200

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 29(17), 22427–22445. https://doi.org/10.1007/s10639-024-12722-y

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT (arXiv:2302.11382). arXiv. https://doi.org/10.48550/arXiv.2302.11382

Woo, D. J., Guo, K., & Susanto, H. (2023a). Cases of EFL secondary students' prompt engineering pathways to complete a writing task with ChatGPT (arXiv:2307.05493). arXiv. https://doi.org/10.48550/arXiv.2307.05493

Woo, D. J., Wang, Y., Susanto, H., & Guo, K. (2023b). Understanding English as a foreign language students' idea generation strategies for creative writing with natural language generation tools. *Journal of Educational Computing Research*, 61(7), 1464–1482.

Woo, D. J., Guo, K., & Salas-Pilco, S. Z. (2024a). Writing creative stories with AI: Learning designs for secondary school students. *Innovation in Language Learning and Teaching*, 1–13. Advance online publication. https://doi.org/10.1080/17501229.2024.2384884

Woo, D. J., Susanto, H., Yeung, C. H., Guo, K., & Fung, A. K. Y. (2024b). Exploring AI-generated text in student writing: How does AI help? *Language Learning & Technology*, 28(2), 183–209. https://hdl.handle.net/10125/73577

Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024c). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, 29(18), 24963–24990.

Woo, D. J., Guo, K., & Susanto, H. (2025). Exploring EFL students' prompt engineering in human–AI story writing: An activity theory perspective. *Interactive Learning Environments*, 33(1), 863–882. https://doi.org/10.1080/10494820.2024.2361381

Woo, D. J., Wang, D., Yung, T., & Guo, K. (2026). Effects of a prompt engineering intervention on undergraduate students' AI self-efficacy, AI knowledge and prompt engineering ability: A mixed methods study. *British Educational Research Journal*, 52(2), 1442–1469. https://doi.org/10.1002/berj.70087

Yacoub, E. (2026). Rewriting the curriculum: Integrating AI prompt engineering into EFL writing instruction. *International Arab Journal of English for Specific Purposes*, 9(1), 15–33. https://doi.org/10.34874/PRSM.iajesp-vol9iss1.63271

Zhang, Z. (2025). The role of artificial intelligence tools on Chinese EFL learners' self-regulation, resilience and autonomy. *European Journal of Education*, 60(2), Article e70127. https://doi.org/10.1111/ejed.70127

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2503890

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2023). Large language models are human-level prompt engineers (arXiv:2211.01910). arXiv. https://doi.org/10.48550/arXiv.2211.01910