The Asia-Pacific Education Researcher (2025) 34:1157–1167
https://doi.org/10.1007/s40299-024-00930-6

REGULAR ARTICLE

Exploring Learner Prompting Behavior and Its Effect on ChatGPT-Assisted English Writing Revision

Myunghwan Hwang[1] · Robert Jeens[2] · Hee-Kyung Lee[1]

Accepted: 7 October 2024 / Published online: 21 October 2024

Abstract This study aims to identify EFL learners' prompting behaviors observed during ChatGPT-assisted English writing revision and examine how these behaviors affect the quality of their revised writings. Specifically, we investigated learners' objectives for using ChatGPT during revision, their prompt-writing approaches, and the alignment between these objectives and approaches. Data from essays, reflection logs, and ChatGPT logs were gathered using a one-group pretest–posttest experiment, with thematic analysis, network analysis and Wilcoxon signed-rank test used for analysis. Results indicated that learners predominantly focused on surface-level features, displayed a lack of prompt approaches, and often employed general, one-size-fits-all, prompts that did not align with their specific objectives. Consequently, while improvements were noted in surface-level aspects of writing, higher-order elements like content, organization, and cohesion showed minimal enhancement. These findings underscore the necessity of incorporating prompt-writing instruction into English writing pedagogy to enhance learners' ability to craft specific, goal-aligned prompts, leading to more productive feedback from AI tools like ChatGPT and facilitating meaningful improvements in higher-order writing skills. This study provides valuable insights for instructors and students, highlighting the pedagogical potential of integrating prompt literacy training in the use of chatbots to enhance English writing skills.

Keywords Prompting behavior · ChatGPT · Prompt literacy · ChatGPT usage objectives · Chatbot-assisted L2 writing · Prompt approaches

✉ Hee-Kyung Lee
heelee@yonsei.ac.kr

Myunghwan Hwang
hwangmh031@naver.com

Robert Jeens
jeensy@yonsei.ac.kr

[1] The Center for Cognitive Science, Yonsei University, Yonsei-ro, Seodaemun-gu, (03722) 50 Seoul, South Korea

[2] Yonsei University, 85 Songdo Gwahak-ro, Yeonsu-gu, Incheon, South Korea

Introduction

The rapid advancement of artificial intelligence has introduced new tools for language learning, with ChatGPT standing out as a notably influential tool in English writing pedagogy. Its increasing use in this field demonstrates its potential to reshape the writing process. Specifically, ChatGPT offers learners various types of support during writing, such as generating ideas, correcting grammar, and enhancing style, all delivered through interactive, human-like conversations (Raheem et al., 2023; Wang, 2024).

The effectiveness of ChatGPT, however, can depend greatly on how learners formulate their prompts, as its responses are directly influenced by the input it receives (Hwang, 2023). This key principle has several significant implications for integrating ChatGPT into English writing instruction. First, learners need to craft prompts that align with their objectives to receive suitable feedback. Second, poorly constructed prompts can lead to irrelevant feedback, missing valuable learning and revision opportunities. Thus, the successful use of ChatGPT in writing hinges on effective learner-ChatGPT interaction, channeled by well-designed prompts.

However, despite the significance of the prompt-driven learner-ChatGPT interactions for successful integration of ChatGPT into English writing pedagogy, the understanding of the interactions driven by prompt between learners

and ChatGPT and their impact on writing outcomes remains minimal. This gap partly exists because prior research has focused mainly on other aspects of learner-ChatGPT interactions, such as ChatGPT's overall impact on writing quality (Maghanmil & Sieras, 2024; Song & Song, 2023), learners' responses to ChatGPT's feedback (Koltovskaia et al., 2024; Lu et al., 2024), and learners' attitudes toward using the tool (Bok & Cho, 2023). While informative, this research does not adequately address how learners formulate prompts and how their prompting behaviors impact writing outcomes. Investigating learners' prompting behaviors and their impact on writing quality is crucial for understanding the dynamics of learner-ChatGPT interactions and, more importantly, for the effective integration of ChatGPT into writing instruction. Therefore, closing this gap will allow both teachers and learners to make better use of ChatGPT to improve writing skills.

In response to this need, the current study investigated prompting behaviors and their impact on writing quality by analyzing the prompt-driven interactions between college EFL learners and ChatGPT during the revision stage of writing, aiming to better understand these interactions by isolating external factors and concentrating solely on this phase (Butler & Britt, 2011). Specifically, the study utilized network analysis to assess learners' prompting behaviors, which, in the present study, were operationalized as the objectives for using ChatGPT, the approaches taken in formulating prompts, and the alignment between these objectives and the prompt approaches employed. The study also examined whether the prompting behaviors could influence writing outcomes, based on the hypothesis that the quality of writing revised through learner-ChatGPT interaction would be affected by learners' prompting behaviors. The specific research questions are as follows:

1. What types of prompting behaviors occur in learner-ChatGPT interaction during the revision phase of English writing?
2. How do these prompting behaviors impact the quality of revised English writings?

Literature Review

Learner-ChatGPT Interactions in Assemblage Framework

Recently, there is a growing preference for understanding the learner-ChatGPT interactions using an assemblage framework. It reflects the reciprocal nature of the learner-ChatGPT interactions, where interactional meaning and outcomes are co-constructed through a network of entities involved (Hwang, 2023; Hwang, Lee, & Shin, 2023; Luther, Kimmerle, & Cress, 2024).

More specifically, learners and ChatGPT, in the assemblage perspective, are not simply sender and receiver but co-participants in a dynamic interaction. Learners bring their objectives, prior knowledge, and unique linguistic challenges, while ChatGPT contributes its algorithmic processing, response generation, and adaptation based on the data it was trained on. These components interact in a non-linear manner, creating an emergent process where both the learner's goals and ChatGPT's responses evolve over time (Han et al., 2024; Hwang, 2023). Such interaction thus becomes a collaborative system, where meaning and understanding are not merely transmitted but constructed through the interplay of different elements.

A central aspect of this assemblage is the concept of the prompting behavior—a cognitive process of translating the users' objectives into the language which generative AI systems can interpret. In light of an assemblage framework, the prompting behavior is not merely a static channel through which information passes but a crucial part of the interaction's success. It acts as a mediator that shapes the quality of responses generated by ChatGPT and reflects the learner's evolving understanding of how to interact with AI systems (Ekin, 2023; Giray, 2023).

Furthermore, the prompting behavior involves understanding the relational dynamics between the learner's objectives, the affordances of ChatGPT, and the broader learning context (Arvidsson & Axell, 2023; Heston & Khun, 2023; Schmidt et al., 2023). In the assemblage, each prompt can serve as a node that connects various elements—cognitive, linguistic, and algorithmic. Thus, the learner's ability to adapt their prompts based on objectives, ongoing feedback and outcomes is essential for maintaining a fluid, productive interaction. Conversely, prompt failures or misalignments are not simply breakdowns in communication but opportunities for the learner to recalibrate and reshape the interaction contributing to the co-construction of knowledge.

In conclusion, for navigating this complex system, learner-ChatGPT interactions from an assemblage perspective highlight the importance of the prompting behavior as a bridge connecting learners and ChatGPT. As learners craft and refine their prompts, well aligning with their objectives, they not only improve their communication with ChatGPT but also actively participate in shaping the learning outcomes within the assemblage.

Prompt Design: Challenge, Strategies and Its Impact

Given the important role that prompts play in learner-ChatGPT interactions as mentioned earlier, there is increasing interest in prompt literacy, an ability underlying learners' prompting behaviors, with growing consensus in AI-assisted

language learning that enhancing learners' prompt literacy is crucial for better engagement with generative AI, including ChatGPT. Although prompt literacy is a relatively new concept without an established definition, Hwang, Lee, and Shin (2023) suggests it involves the skills required to engineer prompts to align with user objectives, critically evaluating AI-generated feedback.

More specifically, prompt-engineering is defined as the skill of crafting, refining, and optimizing prompts to facilitate effective interactions between users and generative AIs, with the aim of generating desired outputs from large language models (LLMs) (Ekin, 2023; Giray, 2023; Heston & Khun, 2023). Despite its significance, prompt creation remains challenging even for AI experts, with several issues arising in the prompt-engineering efforts of non-AI experts, such as learners (Zamfirescu-Pereira, Wong, Hartmann, & Yang, 2023). For example, non-AI experts often stop after just one result, fail to explore the correct instructions for the desired outcome, show a preference for direct instructions over examples, and struggle to construct clear and specific queries. The reasons behind these issues require further exploration, but it is expected that factors like the absence of structured prompt flows and guidance, learners' assumptions of human-like comprehension from generative AI, and a limited grasp of generative AI and LLM functionality are influencing users' lack of proficiency in prompt design (Han, et al., 2023; Mollick & Mollick, 2023).

On the other hand, several prompting strategies have been developed to support non-AI experts in their prompt-engineering practices and address various issues in crafting prompts (Arvidsson & Axell, 2023; Ekin, 2023; Giray, 2023; Heston & Khun, 2023; Schmidt et al., 2023). These include providing clear and specific instructions, applying explicit constraints like structure and length, offering detailed context, incorporating examples, creating domain-specific prompts, and engaging in multi-turn interactions with ChatGPT to refine prompts. These strategies help users develop prompts that clearly define specific objectives, thereby improving their ability to communicate effectively with generative AI.

However, how different prompt designs created by learners impact educational outcomes remains unclear. Taking a glimpse at the impact of prompting behaviors on educational outcomes from a limited number of prior studies, Park (2024) found that when learners were given detailed prompt samples, the quality of their English revisions improved compared to when they created prompts on their own. Similarly, in the area of multimodal generative AI such as DALL·E, which is closely related to prompting LLMs, several reports suggest that more specific and detailed prompts result in better images (Kwon, 2024). To summarize, it can be said that the more detailed and specific a prompt is, the more likely ChatGPT will generate responses that align with the user's objectives. This finding signifies a reasonable hypothesis that the responses learners receive from ChatGPT may vary in quality depending on the effectiveness of their prompting behaviors, and, in turn, the quality of these responses could impact how learners make revisions.

Methods

Participants

This study involved 11 multinational students from an advanced English writing course at a Korean university, all of whom volunteered to participate. The participants' English learning experience ranged from a minimum of 7 years to a maximum of 15 years, with an average of around 10 years. They were enrolled in the international program at this university, where all coursework is conducted in English. Due to the challenging nature of this learning environment, the admission process includes an evaluation of English proficiency and academic ability through official English test scores and a comprehensive English interview. Consequently, according to the Common European Framework of Reference (CEFR) (Council of Europe, 2024), which provides a self-assessment tool for profiling language skills, the English proficiency of the participants in this study is considered to be at least at the B2 level. Furthermore, all participants mentioned that they had used ChatGPT for various purposes, such as school assignments, information searches, and programming assistance, before joining this experiment.

Demographic analysis showed that the group consisted exclusively of first-year students, aged between 19 and 21. Their fields of study were diverse, including 2 (18.2%) students in the humanities, 3 (27.3%) in social sciences, 4 (36.3%) in business, and 2 (18.2%) in interdisciplinary studies. Regarding gender, there was a significant female predominance, with a distribution of 2 males (18.2%) to 9 females (81.8%). A review of the participants' countries of origin shows that there are 3 (27.3%) from China, 3 (27.3%) from Uzbekistan, 2 (18.1%) from Russia, 1 (9.1%) from Austria, 1 (9.1%) from Mongolia, and 1 (9.1%) from Taiwan.

Research Design

A member of the research team led the class for the experiment, which employed a one-group pretest–posttest design due to the limited number of participants. This design is known to effectively maintain internal validity by administering the intervention to the same individuals, thereby managing threats from individual differences and assessing

effectiveness when control groups are not feasible (Moon & Byun, 2013).

More specifically, the study took place over three weeks. In the first week, participants took part in a workshop, their initial formal learning experience with ChatGPT, where they were introduced to its architecture, features, and its role in improving writing quality using prompt examples. However, to maintain the authenticity of learners' interactions, the workshop intentionally excluded direct guidance on how to formulate prompts. Additionally, participants discussed their experiences with ChatGPT and shared their views on incorporating the tool into their learning process. In the second week, a 30-min pretest required learners to write an essay comparing and contrasting urban and rural lifestyles without the use of dictionaries, ChatGPT, the Internet, or peer discussions, and without any formal instruction on comparison essays. In the third week, learners revised their initial drafts with ChatGPT during two sessions: one outside of class (session 1 without a time limit) and one in class (session 2 for 30 min), with an average of 3.5 and 3 turns with ChatGPT, respectively (i.e., one turn consisting of a learner's request and ChatGPT's subsequent feedback). After each revision, participants documented their reflections in a log.

Data Collection Instruments

We collected three types of data: pre-and post-English writing samples from participants, reflection logs, and authentic interaction logs with student-ChatGPT. Below, we provide a detailed overview of the instruments utilized in the study.

First, participants' pre-and-post-English writing samples were collected to assess the impact of utilizing ChatGPT for revising on the quality of learners' writings. This was done to analyze which areas within the writing assessment demonstrated substantial progress following the intervention.

Second, a reflection log was designed to track reasons why participants utilized ChatGPT in the writing revision process, and the prompts they crafted to meet their objectives. The reflection log contains questions including "*What difficulties did you experience and why did you use ChatGPT during your writing revisions?*" and "*Record the specific prompt you used to overcome these difficulties*".

Third, ChatGPT log history was gathered to reconfirm participants' prompt inputs. Participants shared these logs by emailing their interactions with ChatGPT to the instructor or posting them on the class's online board.

Scoring Procedures

Pre-and post-writing samples were evaluated using the Test in English for Educational Purposes (TEEP) attribute writing scale (Weir, 1990), which measures content, organization, cohesion, vocabulary, grammar, punctuation, and spelling, each rated from 0 to 3. Two independent raters scored the samples: one member of the research team with a Ph.D. in English education and extensive teaching and research experience, and an externally recruited Filipino rater who uses English as a first language, holding a Master's degree and over 12 years of high school English teaching experience.

Cohen's Kappa was employed to evaluate inter-rater reliability, as categorized by Landis and Koch (1977). For the pre-writing samples, the reliability scores were: content 0.62, organization 0.86, cohesion 0.86, vocabulary 0.85, grammar 0.85, punctuation 0.81, and spelling 0.74. For the post-writing samples, the reliability scores were: content 0.82, organization 0.65, cohesion 0.81, vocabulary 0.79, grammar 0.74, punctuation 1.0, and spelling 0.74. The total score reliability for pre- and post-writing samples was 0.95 and 0.92, respectively, indicating almost perfect agreement. These figures demonstrate that the inter-rater reliability between the two raters ranged from substantial agreement (0.61 to 0.80) to almost perfect agreement (0.81–1.00) based on Landis and Koch's classification.

Data Analysis

The data analysis followed several steps. First, the study conducted a thematic analysis to investigate why participants used ChatGPT for revisions and the types of prompts they employed. One researcher examined participants' written responses to questions in the Reflection Log. To improve the reliability of the thematic analysis outcomes, two measures were implemented. First, a coding scheme was developed by identifying 54 individual objectives from 11 participants, which were then categorized into 15 overarching objectives. Similarly, 50 individual prompt approaches were merged into 10 categories. However, four simple complaints about ChatGPT (e.g., "ChatGPT did not understand my prompt correctly") were excluded, as they did not represent objectives for using ChatGPT. The initial coding scheme was reviewed a week later to confirm consistency, and no additional concepts were identified. The coding scheme developed for this study is detailed in Appendix 1 and Appendix 2. Second, using the developed coding scheme, the researcher counted the frequency of learners' objectives for using ChatGPT and their prompt approaches. To ensure consistency in calculating the frequency, the analyst conducted another analysis at three-week intervals. The test–retest reliability was assessed using Kappa, which resulted in a perfect score of 1.0.

Second, to explore the congruence between objectives and prompt approaches, we created a 2-mode network mapping objective types to approach categories. Unlike 1-mode networks that focus on one type of relationship, 2-mode networks encompass two distinct entity types and their interconnections (Kwak, 2017). Furthermore, degree centrality was computed to highlight the significance of nodes,

emphasizing their crucial roles due to extensive connections. The network visualization displayed connection strength through edge thickness and node importance through size, with larger nodes being more significant. Network analysis and visualization were performed using Ucinet 6.759 and Netdraw 2.179.

Third, we computed the mean, median, and standard deviation to evaluate the distribution of participants' pre- and post-writing scores in each scoring domain. Due to the small sample size, the nonparametric Wilcoxon signed-rank test was used to explore mean differences between pre-test and post-test scores. The matched pairs rank biserial correlation calculated the effect size to assess the treatment's impact. All statistical analyses were conducted using SPSS version 23.0.

Meanwhile, the rationale behind using network analysis in this study is as follows: First, by scrutinizing the relationship between learners' objectives for using ChatGPT and their prompt approaches through network analysis, one can determine whether the connection between these two entities is balanced or unbalanced. This relationship can be verified through the degree centrality measure. An imbalance in a 2-mode objective-approach network would indicate that learners rely on a limited number of prompt approaches to address multiple objectives, revealing how they use ChatGPT. Second, network analysis also identifies the most relied-upon objectives and prompt approaches using centrality measures, with node sizes indicating key elements. Third, network analysis provides an intuitive method to evaluate the alignment between learners' objectives and prompt approaches. That is, constructing a 2-mode network makes relationships visible that are not apparent through simple frequency analysis, with the strength of relationships visually represented by the thickness of the edges.

Results

Descriptive Analysis of Objectives and Prompt Type

As displayed in Table 1 , the analysis revealed fifteen distinct objectives for which participants used ChatGPT during the revision of their English writing. The most common objective was improving grammatical accuracy ($N=11$), followed by generating ideas to enrich content ($N=8$). Participants frequently sought assistance with vocabulary adjustments: finding context-appropriate words ($N=7$), finding synonyms ($N=2$), checking word usage ($N=2$), checking word spelling ($N=2$), and finding advanced vocabulary ($N=1$). They aimed to enhance sentences by making them sound more natural ($N=5$) or translating from their native language ($N=1$). Other objectives included checking essay organization ($N=3$), generating titles ($N=2$), checking tone ($N=2$), rephrasing texts ($N=1$), and checking information ($N=1$).

In addition, through the analysis of participants' prompt approaches, ten distinct approaches were observed during their interactions with ChatGPT. The most common was requesting a rewrite of their entire essay (Rewriting Essay, $N=16$). Other frequent approaches included requesting ideas relevant to the topic (Requesting Ideas, $N=8$), performing a general grammar check (General Grammar Check,

Table 1 Results of analyzing degree centrality

NO	Objectives for Using ChatGPT	Freq	Degree Centrality	Type of Prompt Approach	Freq	Degree Centrality
1	Improving Grammar Accuracy	11	0.50	Rewriting Essay (RE)	16	0.60
2	Generating Ideas	8	0.40	Requesting Ideas (RI)	8	0.27
3	Finding Words	7	0.50	General Grammar Check (GGC)	6	0.27
4	Making Sentences Natural	5	0.10	Simplification (SF)	6	0.27
5	Checking Organization	3	0.30	Requesting General Feedback (RGF)	6	0.27
6	Generating Title	2	0.20	General Error Check (GEC)	3	0.20
7	Checking Tone	2	0.10	Rewriting Essay Partly (REP)	2	0.13
8	Finding Synonym	2	0.20	Searching Synonym (SS)	1	0.07
9	Checking Word Usage	2	0.20	Checking General Organization (CGO)	1	0.07
10	Checking Word Spelling	2	0.20	Requesting Samples (RS)	1	0.27
11	Seeking General Feedback	2	0.10			
12	Translating Sentences	1	0.10			
13	Finding Advanced Word	1	0.20			
14	Rephrasing Text	1	0.10			
15	Checking Information	1	0.10			
	Total	50			50	

Freq stands for Frequency in the table

$N=6$), asking for simplification or summaries (Simplification, $N=6$), and requesting general feedback on writing quality (Requesting General Feedback, $N=6$). Less common prompts involved general error checks (General Error Check, $N=3$), partial essay rewrites (Rewriting Essay Partly, $N=2$), searching for synonyms (Searching Synonym, $N=1$), checking general organization (Checking General Organization, $N=1$), and requesting sample essays (Requesting Samples, $N=1$).

Exploration of Learner Prompting Behaviors

Participants exhibited several prompting behaviors in their use of ChatGPT during the revision phase. First, they predominantly focused on surface-level features of writing, such as grammar, vocabulary, and sentence structure, rather than higher-order aspects like organization, style, tone, and logical coherence. This focus suggests that learners prioritized immediate linguistic accuracy over deeper structural and rhetorical improvements.

In addition, there was a greater number of objectives for using ChatGPT compared to the number of approaches employed, as exhibited in Table 1. While fifteen distinct objectives were identified, only ten types of prompt approaches were used. This discrepancy indicates that learners had a wide range of specific objectives but utilized a limited set of approaches to interact with ChatGPT. The constrained variety of approaches may have limited the effectiveness of addressing their diverse revision needs.

Regarding the alignment between objectives and approaches, some participants made direct and specific requests that closely matched their objectives. For example, they used the "Requesting Ideas" approach when aiming to generate ideas or the "General Grammar Check" to improve grammatical accuracy. This direct alignment is predicted to facilitate efficient interactions with ChatGPT, allowing learners to effectively address their revision goals.

However, further misalignment between learners' objectives and their prompt approaches were also observed as exhibited in Fig. 1. Participants often utilized general prompts to achieve specific objectives, which may have hindered the tool's effectiveness. For instance, instead of explicitly requesting assistance with particular grammar issues or vocabulary enhancements, learners frequently used general approaches with broad terms such as "General Grammar Check" and "General Error Check." These prompts lacked specificity and did not adequately convey the learners' precise needs, potentially limiting the quality of feedback provided by ChatGPT.

Moreover, there was a notable over-reliance on the "Rewriting Essay" approach. Participants frequently requested ChatGPT to rewrite their entire essays as an

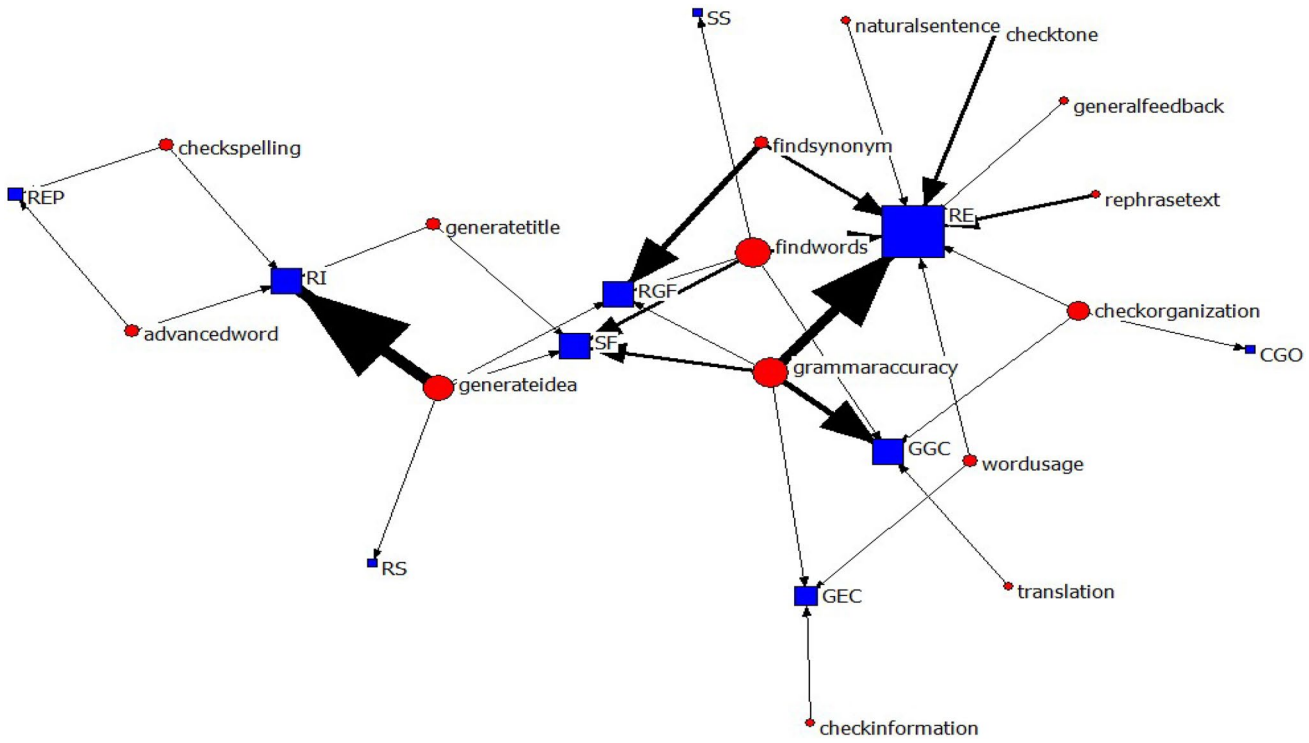

Fig. 1 Objective-Approach 2-mode network. In the Objective-Approach 2-mode network, node names have been shortened not only to differentiate clearly between objectives and approaches but also to streamline the visualization process

indirect and roundabout method to address specific objectives. This approach was employed for a variety of purposes, including improving grammatical accuracy, enhancing vocabulary, adjusting tone, and making sentences sound more natural. By depending heavily on essay rewriting, learners may not have fully leveraged ChatGPT's capabilities to provide targeted assistance for their specific revision goals. This over-reliance suggests a tendency to use a one-size-fits-all solution, which may reflect a lack of prompt literacy among learners when it comes to addressing different writing challenges.

Meanwhile, a complete discrepancy was also observed between learners' objectives and the prompt approaches in certain cases. The misalignment between what learners aimed to accomplish and the approaches they used reveals yet another example of their struggle to create effective prompts. For example, participants aiming to generate new ideas or gather information on a topic sometimes employed the "Requesting General Feedback" approach, which does not directly solicit idea generation. Similarly, those seeking to check the organization or logical flow of their essays used prompts unrelated to structural analysis, such as general grammar checks or essay rewriting. This misalignment could lead to suboptimal support from ChatGPT, as the tool may generate responses that do not align with the learners' intentions.

These findings emphasize the difficulties learners encounter in expressing their needs when interacting with ChatGPT. Moreover, the results suggest that intervention and guidance to enhance learners' prompt literacy would be essential for effectively incorporating ChatGPT into English writing instruction.

Impact of Learner Prompting Patterns on Revised Writing Quality

Network analysis revealed that the way learners approach their prompting behaviors can significantly influence the quality of their revised writings. As noted previously, learners primarily concentrated on surface-level features such as grammar, vocabulary, and sentence structure, with their prompts being more direct and specific for these elements. In contrast, for higher-level writing qualities like content, organization, and logical coherence, learners employed more general and indirect prompts. This pattern suggests that learners communicated more effectively regarding surface-level issues, may have received clearer feedback on them, and had more opportunities to enhance these areas of their writing, leading to improvements in surface-level features.

This hypothesis was confirmed by an analysis of learners' pre- and post-revision results, as shown in Table 2. Statistical tests indicated significant improvements in grammar ($z = -2.21$, $p < 0.05$, $r = 0.92$) and spelling ($z = -2.46$, $p < 0.05$, $r = 0.90$). However, improvements in higher-order aspects such as content, organization, and cohesion were not statistically significant.

These results indicate that the quality of learners' revisions may vary depending on how they engage in prompting behavior. They also highlight the importance of aligning learners' goals with their prompts by focusing on both superficial and deeper aspects of writing, crafting specific prompts, and explicitly requesting feedback on higher-order writing features. This alignment is crucial for effective engagement with ChatGPT and for achieving comprehensive improvements in writing quality through prompt-driven generative AI models like ChatGPT.

Discussion

This study aimed to identify the types of learners' prompting behaviors observed during ChatGPT-assisted English writing revision. Specifically, we examined learners' objectives for using ChatGPT, their prompt-writing approaches, and the alignment between these objectives and approaches. Furthermore, we investigated how these prompting behaviors affected the quality of the learners' revised English writings.

Content analysis of learners' ChatGPT usage objectives and prompt approaches, along with network analysis of the alignment between these objectives and approaches, indicated that learners, with a restricted pool of prompt approaches, predominantly focused on surface-level features

Table 2 Comparative Analysis of Pre-Writing and Post-Writing Scores

	N	Pre-writing			Post-writing			*Z*	*p*	Effect Size
		Md	*M*	*SD*	*Md*	*M*	*SD*			
Content	11	2.90	2.86	.32	2.75	2.41	1.02	− 1.63	.102	.16
Organization	11	2.00	1.95	.72	2.19	2.10	.80	− .76	.450	.76
Cohesion	11	1.88	1.86	.78	2.36	2.41	.49	− 1.86	.063	.97
Vocabulary	11	2.17	2.14	.71	2.75	2.68	.46	− 1.71	.088	.72
Grammar	11	2.00	2.04	.72	2.83	2.77	.41	− 2.21	.027	.92
Punctuation	11	2.64	2.59	.49	2.91	2.91	.30	− 1.89	.059	.98
Spelling	11	2.17	2.23	.41	2.83	2.77	.41	− 2.46	.014	.90

of writing, such as grammar and punctuation, rather than higher-order writing aspects like content, organization and cohesion. In addition, there was a significant misalignment between learners' objectives and their prompt approaches, with many learners heavily relying on general, one-size-fits-all prompting approaches, and occasionally displaying a complete discrepancy between their intended objectives and the prompts they employed.

These findings suggest a lack of prompt literacy among learners and, more specifically, a deficiency in prompt-engineering skills—the skill of crafting prompts that accurately reflect one's intentions, aligning with previous studies reporting that learners often experience difficulties when interacting with LLM-based generative AI tools such as ChatGPT. More specifically, prior research has found that learners are uncertain about how to ask questions and what to inquire about when using these technologies during writing revisions (Bok & Cho, 2023; Lee, 2024). The prompting behaviors observed in this study also resemble the non-specific prompting strategies identified among non-expert programmers in earlier studies (Zamfirescu-Pereira, Wong, Hartmann, & Yang, 2023).

Such absence of prompt-engineering skills can lead to several issues in AI-assisted writing revision. First, learners may receive inaccurate or irrelevant responses that do not align with their objectives. Second, even when responses are relevant, the lack of specificity can render the feedback unproductive. Third, overly general prompts may generate an overwhelming amount of information, imposing a significant cognitive load on learners as they attempt to process and apply the feedback, as supported by several previous research reporting that ChatGPT's responses can be more verbose than expected, increasing the cognitive burden on learners and hindering effective writing improvement (Koltovskaia et al., 2024).

Although this study did not delve deeply into the causes behind learners' ineffective prompting behaviors, several potential causes may explain why learners exhibit such prompting behaviors. Fundamentally, prompt literacy is an ability acquired through experience and training. Given the relatively recent popularization of ChatGPT, learners may not have had sufficient opportunities to develop effective prompt design skills. Indeed, participants in this study reported using ChatGPT for personal or academic purposes but had not received formal education on prompt-engineering strategies beyond the workshop provided in this research. Moreover, learners might lack knowledge about which aspects to focus on to improve writing quality and how to formulate their queries using appropriate keywords. That is, a lack of meta-knowledge regarding prompt construction and writing revision may contribute to these issues. As a further point, learners may overestimate the capabilities of generative AI, expecting its human-like communication ability without recognizing that AI systems like ChatGPT do not process information in the same way humans do.

Meanwhile, the analysis of the current study also provided evidence that the identified prompting behaviors could negatively impact the quality of learners' revised writings. Evaluations showed that while surface-level writing features improved, higher-level aspects did not show significant enhancement. Considering the nature of ChatGPT's response, this outcome can be attributed to learners' focused requests for feedback on surface-level features and their reliance on general or one-size-fits-all prompts, which are less likely to elicit specific feedback on the detailed aspects they intended to improve. ChatGPT's language is influenced by an attention mechanism that identifies and prioritizes important keywords and their relevant terms from the input (Soydaner, 2022; Vaswani et al., 2017). The Transformer model, pre-trained on extensive datasets, predicts the most statistically appropriate and relevant words to follow the given input (Brown et al., 2020; Fezari & Ali-Al-Dahoud, 2023). Consequently, unless keywords or specific content related to higher-level writing aspects are explicitly included in the prompt, ChatGPT may not offer feedback on these areas. However, learners can still receive surface-level feedback relatively easily and more frequently, even when using general or less specific prompts. Therefore, the lack of specific prompts directly requesting assistance with higher-order writing skills limits the potential for improvement in these domains.

In summary, this study highlights the critical importance of developing prompt literacy among learners to enhance the effectiveness of AI-assisted writing revision. The findings suggest that without the ability to construct specific, goal-aligned prompts, learners may struggle to obtain useful feedback that addresses higher-order writing skills. This underscores the necessity of incorporating prompt-writing instruction into educational programs that utilize generative AI tools. By equipping learners with the skills needed to interact effectively with such technologies, educators can ensure that students fully benefit from the capabilities of AI and achieve meaningful improvements in their writing.

Conclusion and Limitations

The significance of this research lies in highlighting the critical role of prompt literacy in effectively utilizing AI-assisted writing tools like ChatGPT. By illuminating the gap between learners' objectives and their prompting approaches, the study underscores the necessity of incorporating prompt-writing instruction into English writing programs. Enhancing learners' ability to craft specific, goal-aligned prompts can lead to more productive feedback from AI tools, thereby facilitating meaningful improvements in writing skills.

However, despite the significance of the current study, the study has limitations. We did not consider learners' responses to ChatGPT's feedback—such as accepting or rejecting it and how they incorporated it into revisions. In addition, we failed to explore the root causes that make it difficult for learners to create effective prompts. Future research, therefore, should address these limitations by comprehensively examining the integration of user objectives, prompt approaches, and learner responses to gain a holistic understanding of the prompting process and its impact on writing quality. Investigating the underlying factors contributing to difficulties in prompt-writing is also necessary to develop targeted strategies for enhancing learners' prompt literacy. Finally, learners' prompt literacy may be influenced by factors like AI literacy. Additionally, their prompting behaviors could be shaped by their English proficiency depending on the task type. Consequently, future studies need to examine the role of various learner factors in learner-ChatGPT interactions and prompt literacy growth.

In conclusion, this study emphasizes the importance of developing prompt literacy among learners to maximize the benefits of AI-assisted writing revision tools. By equipping learners with the skills to construct specific, goal-aligned prompts, educators can help students engage more effectively with generative AI technologies like ChatGPT, leading to substantial improvements in both surface-level and higher-order writing skills.

Acknowledgements This study underwent review by the Institutional Review Board (IRB) of Yonsei University and obtained approval.

Author Contributions Myunghwan, Hwang: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Roles/Writing—original draft; and Writing—review & editing. Robert, Jeens: Conceptualization, Data curation, Methodology. Hee-Kyung, Lee: Conceptualization; Data curation, Investigation, Methodology, Project administration, Supervision, Validation, Roles/Writing—review & editing.

Table 3 Coding Scheme for Learners' Objectives for Using ChatGPT

No	Objectives for Using ChatGPT	Definition	Examples for Learners' answers to Q1
1	Improving Grammar Accuracy	Improve sentence grammar for accuracy	Improve my essay by reducing grammar mistakes. (Learner 4)
2	Generating Ideas	Ask for ideas to enrich the text	I used ChatGPT to make 10 facts about the most important aspects of the topic. (Learner 13)
3	Finding Words	Find and select context-appropriate vocabulary	I couldn't find the proper word that clearly expressed my thinking at times. (Learner 9)
4	Making Sentences Natural	Make sentences feel more native to English speakers	As English is not my mother tongue, ChatGPT was used to make sentences better, smoother. (Learner 7)
5	Checking Organization	Refine the structure for better clarity	I wanted to come up with a more organized essay. (Learner 9)
6	Generating Title	Generate a suitable title for the work	When asking for titles, I had to ask a few times with adjectives such as simpler, interesting for more suitable options and answers. (Learner 7)
7	Checking Tone	Tailor the tone to meet academic standards	I wanted to come up with a more eloquent essay that articulated my thoughts clearly. (Learner 9)
8	Finding Synonym	Search for synonyms	I experienced lack of synonyms. (Learner 11)
9	Checking Word Usage	Determine if a word is appropriate for the context	To get some new vocabulary, to ensure that the information I am using to write essay is correct (Learner 12)
10	Checking Word Spelling	Check for correct word spelling	I confused the usage of spelling (Learner 14)
11	Seeking General Feedback	Request general feedback or advice	Correct mistakes and ask feedback to improve essay (Learner 5)
12	Translating Sentences	Translate text from a native language into English	Help me to improve my writing sentences and translate my language into English (Learner 10)
13	Finding Advanced Word	Search for advanced vocabulary	I had weak difficult vocabulary (Learner1) ChatGPT was used for more formal word choices for writing (Learner 7)
14	Rephrasing Text	Ask for rewriting	I wanted to polish the English texts (Learner 10)
15	Checking Information	Inquire about the accuracy of known information	I used ChatGPT to ensure that the information I am using to write essay is correct (Learner 12)

Learners' answers to Q1 represent examples of their responses to question 1 in the reflection log, which asked, *"What difficulties did you experience and why did you use ChatGPT during your writing revisions?"*

Table 4 Coding Scheme for Learners' Prompt Approaches

No	Type of Prompt Approach	Definition	Examples for Prompts Written by Leaners
1	Rewriting Essay	Rewrite the essay	Help me revise the article. Revise this article on the basis of the sentence. (Learner 10)
2	Requesting Ideas	Request for writing ideas	Analyze the difference between living in a city and living in a rural area. Tell me about benefits and drawbacks of living in a city. (Learner 12)
3	General Grammar Check	Check for grammatical accuracy	Correct the grammar in this paragraph (Learner 14)
4	Simplification	Summarize feedback content and the essay	Summarize the text (Learner 1) Make them (feedback) easier for understanding. (Learner 7)
5	Requesting General Feedback	Request overall feedback on the text	Analyze this paragraph and give feedback. (Learner 5)
6	General Error Check	Check for errors throughout the text	Checked the final essay for mistakes. (Learner 12)
7	Rewriting Essay Partly	Partially revise the essay	Revise the paragraph, only change the words (Learner 14)
8	Searching Synonym	Search for synonyms	What is the academic version of find out? (Learner 1)
9	Checking General Organization	Check if the text is organized properly	Make my essay more organized. (Learner 9)
10	Requesting Samples	Request a sample essay	Find a good conclusion for this kind of essay. (Learner 13)

Examples for Prompts Written by Learners refers to actual examples of prompts created by learners to achieve their objectives using ChatGPT. These examples were taken from learners' ChatGPT log histories and their responses to Q2 in the reflection log, which asked, *"Record the specific prompt you used to overcome these difficulties."*

Funding This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea [NRF-2023S1A5B5A16079327].

Data availability As requested, the corresponding author is able to supply it.

Appendix

See Tables 3, 4

References

Arvidsson, S., & Axell, J. (2023). Prompt engineering guidelines for LLMs in requirements engineering. Retrieved from https://hdl.handle.net/2077/77967.

Bok, E., & Cho, Y. (2023). Examining Korean EFL college students' experiences and perceptions of using ChatGPT as a writing revision tool. *Jounral of English Teaching through Movies and Media, 24*(4), 15–27.

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., Remesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Proceedings of Advances in Neural Information Processing Systems, 33*, 1–25.

Butler, J. A., & Britt, M. A. (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication, 28*(1), 70–96.

Council of Europe. (2024, June). Common european framework of reference for languages (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages/table-2-cefr-3.3-common-reference-levels-self-assessment-grid

Ekin, S. (2023). Prompt engineering for ChatGPT: A quick guide to techniques, tips, and best practices. *TechRxiv*. https://doi.org/10.36227/techrxiv.22683919.v2

Fezari, M., & Ali-Al-Dahoud, A. A. D. (2023). *From GPT to AutoGPT: A brief attention in NLP processing using DL*. ResearchGate. https://www.researchgate.net/publication/370107141_From_GPT_to_AutoGPT_a_Brief_Attention_in_NLP_Processing_using_DL.

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering, 51*(12), 2629–2633.

Han, J., Yoo, H., Kim, Y., Myung, J., Kim, M., Lim, H., Kim, J., Lee, T., Hong, H., Ahn, S., & Oh, A. (2023). RECIPE: How to integrate ChatGPT into EFL writing education. *Proceedings of the Tenth ACM Conference on Learning @ Scale*, 416–420.

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S. Y., & Oh, A. (2024). RECIPE4U: Student-ChatGPT Interaction Dataset in EFL Writing Education. *arXiv preprint* arXiv:2403.08272.

Heston, T. F., & Khun, C. (2023). Prompt engineering in medical education. *International Medical Education, 2*(3), 198–205.

Hwang, Y. H., Lee, J. H., & Shin, D. K. (2023). What is prompt literacy? An exploratory study of language learners' development of new literacy skill using generative AI. *arXiv preprint* arXiv:2311.05373.

Hwang, Y. H. (2023). The emergence of generative AI and PROMPT literacy: Focusing on the use of ChatGPT and DALLE for English education. *Journal of the Korea English Education Society, 22*(2), 263–288.

Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive and affective engagement. *Journal of Second Language Writing, 65*, 101130. https://doi.org/10.1016/j.jslw.2024.101130

Kwahk, K. Y. (2017). *Social network analysis*. Chung Ram.

Kwon, D. H. (2024). Analysis of prompt elements and use cases in image-generating AI: Focusing on midjourney, stable diffusion, firefly DALLE. *Journal of Digital Contents Society, 25*(2), 341–354.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, Y. J. (2024). Can my writing be polished further? When ChatGPT meets human touch. *ELT Journal*. https://doi.org/10.1093/elt/ccae039

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students'academic writing? *Assessment & Evaluation in Higher Education, 49*(5), 616–633. https://doi.org/10.1080/02602938.2024.2301722

Luther, T., Kimmerle, J., & Cress, U. (2024, June). Co-writing with AI: How Do People Interact with ChatGPT in a Writing Scenario?. In International Conference on Human-Computer Interaction (pp. 198–207). Cham: Springer Nature Switzerland.

Maghamil, M., & Sieras, S. (2024). Impact of ChatGPT on the academic writing quality of senior high school students. *Journal of English Language Teaching and Applied Linguistics, 6*(2), 115–128.

Mollick, Ethan R. and Mollick, Lilach, Assigning AI: Seven Approaches for Students, with Prompts (September 23, 2023). The Wharton School Research Paper, Available at https://doi.org/10.2139/ssrn.4475995

Moon, S. B., & Byun, C. J. (2013). *Experimental design and analysis for the social science*. Hak Ji Sa.

OpenAI. (2024). Prompt engineering. *OpenAI*. Retrieved September 25, 2024, from https://platform.openai.com/docs/guides/prompt-engineering

Park, J. U. (2024, July 5–6). *Using ChatGPT to help with students' academic communication – wri*ting [Conference presentation]. 2024 The Korea Association of Teachers of English Conference (KATE), Yeosu, Jeollanam-do, South Korea.

Raheem, B. R., Anjum, F., & Ghafar, Z. N. (2023). Exploring the profound impact of artificial intelligence applications (Quillbot, Grammarly and ChatGPT) on English academic writing: A systematic review. *International Journal of Integrative Research (IJIR), 1*(10), 599–622.

Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Towards a catalog of prompt patterns to enhance the discipline of prompt engineering. Retrieved from chrome extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA-User-Journal.pdf. Accessed March 31, 2024.

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*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Soydaner, D. (2022). Attention mechanism in neural networks: Where it comes and where it goes. *Neural Computing and Applications, 34*(16), 13371–13385.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of Advances in Neural Information Processing Systems, 30*, 1–11.

Wang, Y. (2024). Reviewing the usage of ChatGPT on L2 students' English academic writing learning. *Journal of Education, Humanities and Social Sciences, 30*, 173–178.

Weir, C. J. (1990). *Communicative language testing*. Prentice Hall Rogents.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023, April). Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21).