

In Search of the Prompt that Produces Useful Written Corrective Feedback for L2 Composition Classes

¹James R. Brawn

¹Department of English Education, Graduate school of Education,
Hankuk University of Foreign Studies, Seoul, South Korea.

Abstract

The use of artificial intelligence (AI) in language education may be in its infancy, but technological advances, especially natural language processing, will lead to its widespread adoption far sooner than many may think. For example, large language models (LLMs) like ChatGPT are often used when individuals utilize AI systems. This means that researchers in second language learning must begin evaluating the utility of AI-based tools for second language instruction. This study describes the importance of prompt engineering in designing effective prompts for second-language writing feedback. This action research (AR) study revealed that prompts could constrain the usefulness of AI-generated feedback and suggests that, like LLMs, users are few-shot learners. Adapting the prompts and understanding the limitations and constraints that these prompts produce will allow instructors to design prompts to make ChatGPT and other AI-based applications more helpful to learners in second-language composition classes.

Keywords: Prompt Engineering, Written Corrective Feedback, AI, ChatGPT, L2 Composition.

1. Introduction

Action research (AR) is a reflective, systematic approach to investigate and improve teaching practices and students' learning outcomes. It is usually collaborative because it involves both the teachers and the students. The aim of AR is to identify issues and challenges in the language learning environment. Once these issues and challenges are identified, the next step is not only to understand the phenomenon but also to take action based on the findings, thus improving both pedagogical strategies and student performance. Therefore, AR is especially helpful in the second language (L2) writing classroom. Instructors can systematically investigate the issues and challenges that students face when writing in another language. to improve student writing outcomes.

One challenging issue is written corrective feedback (WCF) in L2 writing classes. It has been an area of significant research, and it continues to present ongoing challenges for both teachers and learners. For example, Ferris found that students who received detailed corrective feedback made fewer grammatical errors in subsequent drafts, but the feedback needed to be clear and targeted to be effective. Truscott, on the other hand, claimed that grammar correction does not lead to long-term improvements and can negatively affect motivation. Consequently, he recommended that teachers avoid the time-consuming process of providing detailed corrective feedback since there was no clear evidence of significant benefits. More recently, Hyland & Hyland published a study that looked at both explicit corrective feedback and content-based feedback. In that study, they suggested that combining feedback types, that is, providing both form-focused feedback and content-focused feedback, was superior to just providing corrective feedback on form alone. They also suggested

tailoring feedback to individual students' needs was a more effective strategy for enhancing student motivation and writing outcomes. Two meta-analyses of WCF were conducted in 2015, one by Liu and Brown and the other by Kang and Han. Both studies suggested that, in general, WCF helps learners improve their writing, but they identified vital factors that can make WCF more effective. For example, Liu and Brown noted that feedback needs to be clear and consistent so learners can notice, understand, and internalize corrective patterns. Kang and Han [6] found that focused feedback was more effective than unfocused feedback, and indirect feedback, which encourages self-correction, is better for higher-proficiency learners, while direct feedback is more suitable for lower-level learners.

To summarize the importance of WCF, the studies above collectively suggest that WCF is necessary and beneficial for learners. Although a debate continues regarding the value of explicit grammar correction, key factors for effective feedback have been identified. These include feedback that is clear, consistent, and suited to individual student needs. Moreover, the research suggests that balancing feedback between content and grammar and combining different kinds of feedback, such as direct, indirect, and metalinguistic feedback, is more effective than limiting the feedback to just one area or type. Therefore, WCF is an essential part of L2 writing instruction because it helps learners improve not only their accuracy but also facilitates the internalization of complex language structures. The downside of providing WCF to learners is that it is a time-consuming, labor-intensive endeavor. This raises the question: Is there a way to automate this process?

LLMs like ChatGPT have been incorporated into L2 composition classes to provide learners with WCF. This is due to their ability to generate natural language responses quickly and tailor feedback to specific errors. For example, it has been found that ChatGPT can provide feedback that goes "beyond one-by-one correcting by changing surface expressions and sentence structure while maintaining grammatical correctness". Moreover, LLMs like ChatGPT can offer corrective feedback on grammar, vocabulary, coherence, and style. However, providing this feedback in a manner that the learner can use and benefit from linguistically is an issue.

Although LLMs can quickly proofread and correct drafts, designing prompts that will not only help L2 learners make more informed revisions but can also facilitate language development is a challenge. Currently, there are varying opinions on the effectiveness of LLMs for WCF. For example, Fathi and Rahimi report that ChatGPT effectively enhanced L2 learners' writing abilities through interactive feedback tailored to learners' needs, which allowed for gradual improvement in areas like grammatical accuracy and vocabulary. However, they also noted a risk of learners becoming overly dependent on AI-generated suggestions. This reliance could hinder the development of learners' critical thinking and self-editing abilities if not managed carefully. The authors recommend balancing AI use with human instruction to ensure students continue developing these essential skills. A second study by Hou, He, and Cui [9] found that AI-generated WCF helped learners make notable improvements in grammar, vocabulary, and coherence. However, these authors observed that learners often struggle to craft effective prompts to obtain relevant feedback from the AI. Moreover, some learners needed help to interpret and use the feedback provided. The authors conclude that this challenge suggests learners need training in using AI tools effectively to maximize the usefulness of the feedback. Another interpretation would be for the instructor to provide the prompts and provide instructions on how to use the output.

As Hou, He, and Cui [9] noted, prompt engineering is a task that learners often struggle with. One solution to this problem would be for the instructor to provide prompts that maximize

the WCF for their composition students. Thus, the purpose of this AR study is to find a prompt that can maximize the effectiveness of WCF provided by ChatGPT.

Research Question

How does the prompt affect the quality of ChatGPT's written feedback, and to what extent does that written feedback facilitate the writing development of L2 learners in a composition class?

Context of the Study

This study looks at the integration of ChatGPT into an undergraduate second language composition class at a major university in Seoul, South Korea. Approximately twenty-five students are enrolled in the course, and their English proficiency ranges from IELTS 5.0 to 7.0. Over a sixteen-week semester, the students turn in four final papers. This action research reflects the initial attempt to use ChatGPT to give WCF on the students' first essay assignment. The first assignment is a self-introduction essay based on their Life Map, an icebreaking activity learners make on the first day of class [10]. In the next class, they used the Life Map to organize their self-introduction essay, and they did an in-class writing assignment. In week three, they do a peer editing activity in groups. They try to figure out the indirect corrective feedback that their instructor has given them and make suggestions about ways to improve their writing. In week four, they need to use the feedback and the advice from their peer editing group to finalize their essay. For this research, they were also instructed to submit their final draft to ChatGPT, and they used the prompt that they had been given. Learners were to send their instructor the output ChatGPT produced and the corrected finalized essay. The underlying goal of this integration is to demonstrate to students how AI and LLMs like ChatGPT can be ethically used to assist in the writing process; however, the challenge for the instructor was creating a prompt that would be both useful and effective for the learners.

Prompt Iterations and Results

Before sending the prompt to his students, the instructor tested each prompt for the usefulness and effectiveness of WCF. The first iteration of the prompt submitted to ChatPT was as follows: "Please proofread this draft and correct my writing." The usefulness of Prompt #1 as a learning tool was extremely limited (see Figure 1). Although the LLM corrected the essay in terms of clarity, tone, and readability, the output didn't help the learner notice the errors they made. Noticing is an essential step in the developmental process of language learning because it facilitates the internalization of language structures and forms. Noticing involves a learner's ability to recognize specific aspects of the language, such as vocabulary, grammar structures, or pronunciation, in spoken or written input. This does not involve incidental and passive exposure; instead, it requires focused attention on language features. For instance, when learners read a text in their target language and consciously recognize the use of a particular grammatical structure, they are engaging in noticing. The first prompt did not help the language learners notice their errors; therefore, the output was not an effective learning tool.

The output from Prompt #1 lacked explicit feedback. Nothing was in the output to draw learners' attention to problematic areas. To overcome these limitations, the instructor attempted a second iteration. In Prompt #2, the following was submitted: "I am a second-language learner; please proofread my writing and consider grammar, punctuation, formatting, and readability. Provide a summary of the errors that were made." This prompt provides more information about the nature of the task and who is submitting it. It outlines what aspect of language should be corrected, explains who is submitting the essay, and

summarizes errors at the end. The initial output of this prompt was precisely the same as in Prompt #1. The LLM corrected the essay regarding the features specified by Prompt 2: “grammar, punctuation, formatting, and readability,” and summarized those errors at the end (see Figure 2). Even though this was an improvement, the output still failed to help learners notice the problematic areas in their writing. The main failing was that it again didn’t promote noticing, which is essential to second language acquisition. The summary codified the errors, but only the most dedicated learners would return to the original text to find them. A better prompt would need to produce output that included visual cues like bolding, underlining, or coloring text in which errors occurred.

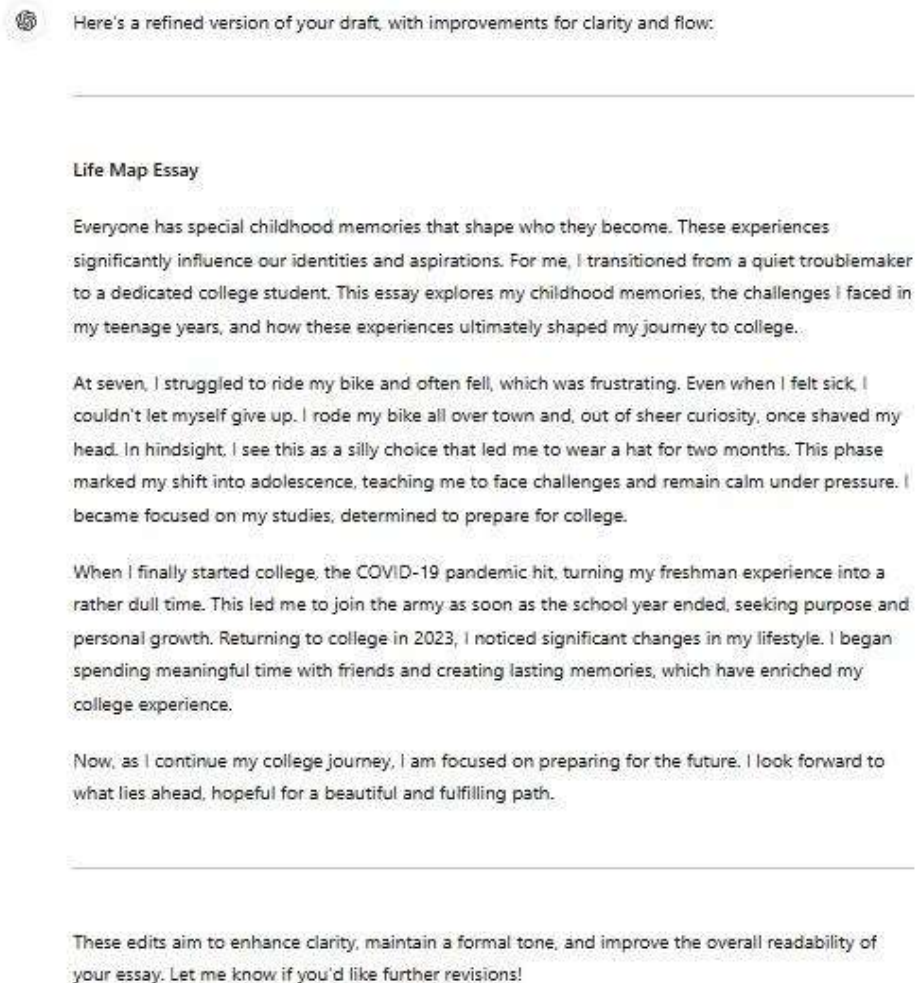


Figure 1. ChatGPT's output of prompt #1

Providing visual cues like bolding and underlining is a technique known as input enhancement. It is used in second language learning to make sure language features are more noticeable to learners. Typically, it involves underlining linguistic features such as grammar or vocabulary to increase their salience [12]. To improve the output of AI-produced corrective feedback, the prompt must describe to the LLM how input enhancement could signal problematic areas in the text. Prompt #3 tries to rectify that problem. Prompt #3 used the following text: “I am a second-language learner, and you are my composition teacher. Please give feedback on my essay. Consider grammar, punctuation, formatting, and readability. Show the results in a table format with the original paragraph on the left and the

suggested changes on the right. Underline all the proposed changes and summarize these actions to improve my writing.”

Summary of Errors

1. Grammar and Word Choice:

- Changed “pretty dull time” to “rather dull time” to maintain a slightly more formal tone.

2. Punctuation:

- Corrected minor comma placements for readability.
- Added or adjusted commas for clarity in compound sentences (e.g., after phrases like “Even when I felt sick” and “Because of this”).

3. Clarity and Flow:

- Minor rephrasing to avoid repetitive language and improve the flow, such as removing “for me” after “frustrating” and replacing “the challenges I faced during my teenage years” with “the challenges I faced in my teenage years.”

4. Verb Tense Consistency:

- Ensured consistency of verb tenses to keep the narrative clear.

These changes help improve readability and ensure that grammar, punctuation, and flow are smooth throughout your essay. Let me know if you'd like any further assistance!

Figure 2. Summary of errors produced by prompt #2

Prompt #3 produced a table (see Figure 3) where the original text could be easily compared to the edited text. This makes the corrective feedback more accessible because the learner doesn't have to look at the original draft to find the errors physically. AI also provided input enhancement through the use of italics. These changes significantly improved the usefulness of the WCF; however, Prompt #3 still fell short of the ideals. Although the WCF promoted by prompt #3 was clear, consistent, and suited to individual student needs, the prompt was less effective in balancing WCF between content and grammar. The prompt also failed to instruct ChatGPT to combine different kinds of feedback, such as direct, indirect, and metalinguistic feedback. As was noted above, WCF is more effective when the feedback is not limited to just one area or kind. So, additional iterations of the prompt should be developed.

Here is a feedback table for your essay, including specific grammar, punctuation, and clarity adjustments. Each change is underlined in the revised column for your reference.

Original Paragraph	Revised Paragraph
<p>Paragraph 1</p> <p>Everyone has special childhood memories that shape who they become. These experiences can greatly influence our identities and dreams. In my case, I went from being a quiet troublemaker to a dedicated college student. This essay will discuss my childhood experiences, the challenges I faced during my teenage years, and how these events ultimately shaped my college journey.</p>	<p>Paragraph 1</p> <p>Everyone has special childhood memories that shape who they become. These experiences can <i>significantly</i> influence our identities and dreams. In my case, I went from being a quiet troublemaker to a dedicated college student. This essay will discuss my childhood experiences, the challenges I faced <i>in</i> my teenage years, and how these events ultimately shaped my college journey.</p>

Figure 3. Table produced by prompt #3

My composition class used prompt #3 to help them revise their self-introduction essay. To promote noticing and internalization, I asked students to print the AI-generated WCF and bring it to class. First, I asked students to highlight the changes made by ChatGPT in their original text. Next, I had the students look at the summary of errors at the end of the WCF (see Figure 4), and I asked them to find those errors in their original text. The purpose of this

activity was to encourage autonomous learning and self-editing skills. The activity asked students to monitor their original writing by highlighting the changes and identifying the errors. Ferris contends that these activities are particularly beneficial in fostering long-term writing development as learners build their capacity to produce accurate and coherent texts without constant external feedback.

Summary of Suggested Changes

1. Word Choice and Clarity:

- Replaced "pretty dull time" with "rather dull time" and "big change" with "significant changes" to improve formality and clarity.
- Changed "to me wearing a hat" to "required me to wear a hat" for clearer expression.

2. Grammar:

- Replaced contractions ("couldn't" to "didn't") to match formal tone.
- Adjusted verb phrases like "looking for purpose and growth" to "seeking purpose and growth" to improve readability.

3. Punctuation:

- Added or adjusted commas to enhance readability, especially in complex sentences.

4. Verb Tense Consistency:

- Checked consistency in verb tenses throughout to ensure smooth narrative flow.

Figure 4. Summary of errors produced by prompt #2

2. Discussion

Although using ChatGPT to provide WCF on L2 composition assignments offers significant benefits, there are several fundamental limitations. For example, Liu and Brown identified limitations when using WCF, such as inconsistencies in application and learners' ability to understand and apply feedback. Although LLMs are more consistent in applying particular techniques, they share these limitations since they provide feedback without considering individual learner differences or a comprehensive understanding of the methodological framework. Another limitation LLMs face is that, unlike humans, LLMs cannot incorporate reflective practice or long-term pedagogical goals, making their feedback more transactional and less developmental. Kang and Han also highlighted the importance of targeted feedback. They believed a differentiated approach based on learner proficiency was an essential feature of effective WCF. Although LLMs can provide differentiated feedback, the reasoning behind this differentiation is algorithmic and lacks the nuanced understanding of when to provide explicit or implicit feedback based on learner needs.

If we consider the research of Fathi and Rahimi, a clear limitation would be the over-reliance on AI tools. They noted that while LLMs foster learner autonomy, they can also reduce critical thinking and self-editing skills. As was stated above, the prompts for WCF need to promote engagement with errors and noticing. If LLMs fail to promote engagement with errors and noticing, this would be a crucial limitation of LLM-generated WCF because learners would then bypass deeper engagement with their errors in favor of simply accepting AI-generated corrections. Another observation was that learners might struggle with contextualizing feedback from LLMs, especially when the AI fails to address discourse-level issues like coherence and argumentation.

This paper attempted to address the limitation that Hou and colleagues described; that is, learners often faced challenges in prompting LLMs effectively. This study attempted to avoid this by engineering a prompt that all the learners could use. From the beginning, the creators of ChatGPT at OpenAI suggested that prompt engineering would be a challenge because

language models are few-shot learners; that is, they learn through trial and error. As Brown noted, few-shot is the term used to describe one of the ways that LLMs are trained. In the few-shot approach, the model is given a few demonstrations of the task, and learning happens as the model adapts to the task. The corollary to this would be that users of LLMs are also few-shot learners; that is, to get the most out of the tool, our prompts must adapt to maximize output from the LLM. This means prompt writers must go through an iterative process of trial and error. This is unsurprising, as several researchers have pointed out that prompt writing is a challenging and complex task for those who are well-versed in the field of machine learning.

As the examples above show, several prompt iterations were necessary before the output provided suitable WCF for learners to improve their writing and develop their language proficiency. Still, even the final prompt needed to be improved as it did not combine different kinds of feedback, such as direct, indirect, and metalinguistic feedback. To improve the prompt, a “few more shots” are necessary to adapt it so that the LLM can maximize the effectiveness of its WCF.

3. Conclusion

Natural Language Processing will likely advance, allowing AI systems to better understand, interpret, generate, and provide written corrective feedback on human language. Action research should be conducted to tailor these tools to learners. This is especially true for prompt engineering, where specific prompts can maximize AI's usefulness and efficiency. Both LLM and its users learn through trial and error. As the examples above show, prompt engineering is an iterative process in which each iteration needs to be accessed for its effectiveness.

References

- Brawn, J. R. (2002). Making the Most out of Students' Lives: A Life Map Icebreaker for EFL Composition Classes. *KATE Forum* 26(3), 13-14. <https://www.tesol.brawnblog.com/HUFS-TESOL/MatDev/Ts/Archive/KateForum.pdf>
- Brown, T. B. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Burns, A. (2010). *Doing Action Research in English Language Teaching: A Guide for Practitioners*. Routledge.
- Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: A qualitative study. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2024.2374772>
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1-11. [https://doi.org/10.1016/S1060-3743\(99\)80110-6](https://doi.org/10.1016/S1060-3743(99)80110-6)
- Ferris, D. R. (2011). *Treatment of error in second language student writing*. University of Michigan Press.
- Fotaris, P., Mastoras, T., & Lameris, P. (2023). Designing educational escape rooms with generative AI: A framework and ChatGPT prompt engineering guide, in *Proceedings of the European Conference on Games-based Learning*.
- Gorer, B., & Aydemir, F.B. (2023). Generating requirements elicitation interview scripts with large language models, in *Proceedings - 31st IEEE International Requirements Engineering Conference Workshops*.

- Hou, X. L., He, S. Y., & Cui, G. R. X. (2024). Learner Use of AI-Generated Feedback for Written Corrective Feedback in L2 Writing: Usefulness, User Proficiency, and Attitude. Proceedings of the 8th International Conference on Education and Multimedia Technology (ICEMT 2024). <https://doi.org/10.1145/3678726.3678767>
- Hyland, F., & Hyland, K. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83-101. doi:10.1017/S0261444806003399
- Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, 99(1), 1-18. <https://doi.org/10.1111/modl.12189>
- Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30, 66-81. <https://doi.org/10.1016/j.jslw.2015.08.011>
- Mohammed, S. B., Khalifa, M. M., & Abubakar, A. (2024). Effect of petrol subsidy removal on government income, cost of living, consumption patterns, savings and investment, and SMEs performance. *Journal of Social Economics Research*, 11(3), 290-308. <https://ideas.repec.org/a/pkp/josere/v11y2024i3p290-308id3774.html>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165-179.
- Suleiman, M. (2011). Poverty reduction as a panacea for violence free elections in Nigeria. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7VZ4PGMAAAAJ&citationforview=7VZ4PGMAAAAJ:tzM49s52ZIMC
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
- Wu, H., Wang, W., Wan, Y., Jiao, Q., and Lyu, M.R. (2023) ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark. arXiv. <https://doi.org/10.48550/ARXIV.2303.13648>