RECIPE: How to Integrate ChatGPT into EFL Writing Education

Jieun Han*
KAIST
Daejeon, South Korea
jieun_han@kaist.ac.kr

Haneul Yoo*
KAIST
Daejeon, South Korea
haneul.yoo@kaist.ac.kr

Yoonsu Kim
KAIST
Daejeon, South Korea
yoonsu16@kaist.ac.kr

Junho Myung
KAIST
Daejeon, South Korea
junho00211@kaist.ac.kr

Minsun Kim
KAIST
Daejeon, South Korea
9909cindy@kaist.ac.kr

Hyunseung Lim
KAIST
Daejeon, South Korea
charlie9807@kaist.ac.kr

Juho Kim
KAIST
Daejeon, South Korea
juhokim@kaist.ac.kr

Tak Yeon Lee
KAIST
Daejeon, South Korea
takyeonlee@kaist.ac.kr

Hwajung Hong
KAIST
Daejeon, South Korea
hwajung@kaist.ac.kr

So-Yeon Ahn†
KAIST
Daejeon, South Korea
ahnsoyeon@kaist.ac.kr

Alice Oh†
KAIST
Daejeon, South Korea
alice.oh@kaist.edu

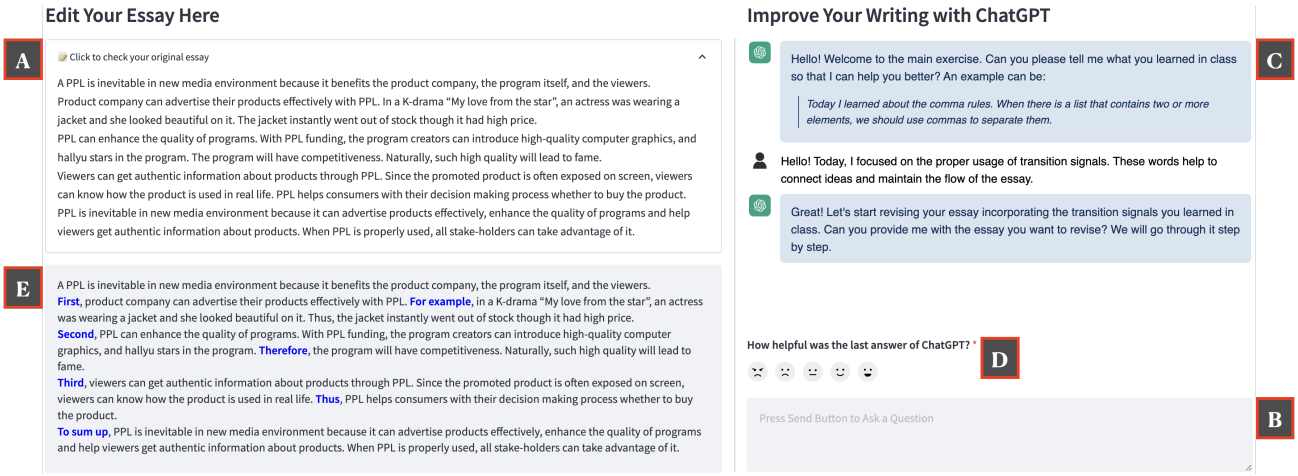

Figure 1: A screenshot of the writing exercise in our platform. A student's original essay is initially provided at (A). A student starts a conversation with ChatGPT at (B) by summarizing what they learned in the previous lecture as we instructed. Conversation history is logged at (C), and a student is prompted to assess the performance of ChatGPT's every response in a 5-point Likert scale at (D). Then, a student can revise the essay at (E) and repeat the process from (B) to (E).

ABSTRACT

The integration of generative AI in the field of education is actively being explored. In particular, ChatGPT has garnered significant interest, offering an opportunity to examine its effectiveness in English as a foreign language (EFL) education. To address this need, we present a novel learning platform called RECIPE (Revising an Essay with ChatGPT on an Interactive Platform for EFL learners). Our platform features two types of prompts that facilitate conversations between ChatGPT and students: (1) a hidden prompt for ChatGPT to take an EFL teacher role and (2) an open prompt for students to initiate a dialogue with a self-written summary of what they have learned. We deployed this platform for 213 undergraduate and graduate students enrolled in EFL writing courses and seven instructors. For this study, we collect students' interaction data from RECIPE, including students' perceptions and usage of the platform, and user scenarios are examined with the data. We also conduct a focus group interview with six students and an individual interview with one EFL instructor to explore design opportunities for leveraging generative AI models in the field of EFL education.

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments**; • **Human-centered computing** → *User centered design.*

*Authors contributed equally.
†Corresponding authors

L@S '23, July 20–22, 2023, Copenhagen, Denmark.

ACM ISBN 979-8-4007-0025-5/23/07…$15.00
https://doi.org/10.1145/3573051.3596200

KEYWORDS

Generative AI, ChatGPT, Learner-ChatGPT Interaction, Essay Writing, EFL Learners

ACM Reference Format:

Jieun Han, Haneul Yoo, Yoonsu Kim, Junho Myung, Minsun Kim, Hyunseung Lim, Juho Kim, Tak Yeon Lee, Hwajung Hong, So-Yeon Ahn, and Alice Oh. 2023. RECIPE: How to Integrate ChatGPT into EFL Writing Education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23), July 20–22, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3573051.3596200

1 INTRODUCTION

In the context of English as a foreign language (EFL) education, the integration of artificial intelligence (AI) technology has been shown to enhance students' learning experience. The use of AI-based tools, including Grammarly and Quillbots, has resulted in significant improvement in EFL learners' writing abilities. Moreover, learners have exhibited positive perceptions towards the use of these tools in their writing class [1, 4].

ChatGPT [1], a large language model (LLM)-driven chatbot by OpenAI, has made a significant breakthrough in the domain of language learning. While earlier chatbots were unnatural and incapable of engaging language learners [5], ChatGPT can generate natural and personalized responses, which make students' learning experience more interactive and engaging [8, 9]. Recent studies have suggested the potential educational benefits of incorporating generative AI into education [6]. Hence online educational platforms such as Khan Academy and Duolingo have already started integrating ChatGPT to their functionalities.

Despite such suggestions and attempts to incorporate ChatGPT into language education, only episodic and anecdotal knowledge has been shared rather than systematic investigation [2, 7]. Therefore, we need to examine the effective use of ChatGPT in higher education and identify its learning effect. In order to effectively integrate ChatGPT into EFL writing courses, it is crucial to investigate the specific design of the integration, considering learners' perceptions and usage of ChatGPT in higher education. In this regard, we introduce an interactive learning platform, RECIPE (Revising an Essay with ChatGPT on an Interactive Platform with EFL learners), aiming to leverage the data collected from our platform to guide learners towards a more effective and engaging learning experience. The main contributions of this work are as follows:

(1) We analyze the perception and usage of ChatGPT in the context of EFL among learners and instructors.
(2) We introduce RECIPE (Revising an Essay with ChatGPT on an Interactive Platform with EFL learners), a learning platform designed to integrate ChatGPT with underlying two types of prompts for students' better learning experience.
(3) We collect students' interaction data on their perception and ChatGPT usage through RECIPE to investigate further development of our platform.

2 PRELIMINARY QUESTIONNAIRE

To investigate students' attitudes, usage, and expectations of ChatGPT, especially in the context of EFL, we conducted a preliminary questionnaire with 213 (91 undergraduate and 122 graduate) college students in South Korea. They were enrolled in three types of English writing courses for the spring semester of 2023. In specific, 91 undergraduate students were enrolled either the Intermediate English Reading & Writing (IRW; 46 enrolled) or the Advanced English Writing (AW; 45 enrolled) course, based on their TOEFL writing scores (15-18 for IRW, and 19-21 for AW). 122 graduate students were enrolled in Scientific Writing (SW). All studies in this research were performed under KAIST Institutional Review Board (IRB) approval.

The questionnaire consists of 5-point Likert scale questions about participant's prior experience and evaluation of ChatGPT, with regard to PERCEIVED HELPFULNESS, TRUSTWORTHINESS, CREDIBILITY, APPROPRIATENESS OF STYLE/TONE, PERFORMANCE, OVERALL SATISFACTION, and REFERRAL INTENTION [2] of ChatGPT in the context of college education. Results from this survey provide visions for developing an educational platform with ChatGPT.

2.1 Perception and Usage of ChatGPT

The majority of students reported a positive user experience with ChatGPT across all seven factors. Specifically, 85% of students with ChatGPT exposure used the tool in their academic work, indicating a potential for leveraging and developing the technology for educational purposes. Moreover, less than half of the students used ChatGPT to improve their English writing skills, highlighting an opportunity and a need for research on integrating ChatGPT into EFL education. These students also relied on other widely available AI tools including Grammarly, Turnitin, and Google Translate, Papago, Wordtune, ExplainPaper, and Elicit for help with English writing (90.4%) most, followed by reading (61.6%), grammar (57.1%), speaking (27.8%), and listening (12.6%) skills.

This questionnaire also underscores the need for comprehensive and well-guided instructions for effectively and efficiently integrating ChatGPT into EFL writing courses, rather than implementing an LLM-agnostic class policy and allowing students to use ChatGPT on their own discretion.

Students with prior experience and knowledge of LLMs expressed higher levels of satisfaction and expectations of using ChatGPT in the courses they enrolled. In particular, those who had low LLM understanding and enrolled in IRW and AW, expected that ChatGPT would be useful for asking questions about the lecture and finding sources to support their writing, which are notable limitations of ChatGPT with hallucination [3]. Whereas students with a high level of LLM understanding exhibited greater satisfaction with their previous ChatGPT experience regarding PERCEIVED HELPFULNESS, APPROPRIATENESS OF STYLE/TONE, PERFORMANCE, OVERALL SATISFACTION, and ENCOURAGEMENT, compared to those with low LLM understanding. However, we did not find a statistically significant difference in terms of TRUSTWORTHINESS and CREDIBILITY. Students with a high level of LLM understanding and those with ChatGPT exposure also exhibited significantly positive expectations regarding PERFORMANCE, CREDIBILITY, and OVERALL SATISFACTION towards the use of ChatGPT in academic courses. Based on these findings, we posit our proposed platform RECIPE can effectively and

[1] https://chat.openai.com/

[2] to what extent they would recommend others to adopt ChatGPT

efficiently guide students towards obtaining satisfactory responses from ChatGPT, ultimately enhancing English writing skills.

3 PLATFORM DESIGN

This section outlines the design of RECIPE, geared towards the development of English writing skills, utilizing ChatGPT's capabilities in a targeted and effective manner. We deployed RECIPE for 213 students who enrolled to EFL courses (IRW, AW, and SW) in the spring semester of 2023. Students' interaction data is gathered throughout the semester from undergraduate and graduate students using our platform. RECIPE comprises three main components: a pre-survey (§3.1), the writing exercise (§3.2), and a post-survey (§3.3). Table 1 displays the data collected at each phase of the platform.

3.1 Pre-survey

We designed a pre-survey to collect (1) students' expectations for the upcoming exercise and (2) their understanding of the topics that they will discuss during the exercise. The pre-survey aims to compare students' expected and actual assistance and to examine changes in their understanding before and after the exercise.

Students are asked to indicate all applicable purposes of using ChatGPT during the exercise from a list of nine options. Additionally, students are required to rate their understanding of the topics to be addressed during the exercise, using a 5-point Likert scale.

3.2 Writing Exercise

Figure 1 displays a screenshot of our platform, which has a writing exercise interface divided into two sections: the left side is for editing an essay, while the right side is for interacting with ChatGPT. The students' essay from the previous session is available on the left side, enabling them to revise their work with the help of ChatGPT.

On the right side, students can consult ChatGPT about their essay. Unlike the existing ChatGPT interface from OpenAI, our platform provides two initial prompts based on empirical prompt engineering, as shown in Table 2: (a) a hidden prompt for the model to set a persona for ChatGPT, acting as a personalized English writing teacher, and (b) an open prompt for students to start a dialogue efficiently. ChatGPT instructs students step by step to revise their essay based on the content they learned during the class, and students are asked to summarize what they learned during the corresponding week or previous classes as the first dialogue. We advise both ChatGPT and students not to provide or request a revised version of the entire essay without any explanation. These instructions serve two purposes: first, to remind students of the lecture content and enhance their learning, and second, to help students receive a more class-relevant response from ChatGPT. We believe this suggested interface can guide EFL learners to write a more specific opening prompt for ChatGPT.

After each turn of the conversation, the edited version of the students' essay is saved to analyze how they revised it with ChatGPT's response. In addition, students are asked to rate the helpfulness of ChatGPT-generated response using a 5-scale Likert scale. Students can continue to converse with ChatGPT and revise their essay.

3.3 Post-survey

First, to gain insights into students' decision-making regarding essay revision, we inquire about their reasons for not making an edit. As the essay editing history is collected during the writing exercise, we can compare the original version of the student's essay with the submitted version after the conversation with ChatGPT. To further analyze prompt engineering, we ask students about the main topic of their conversation using multi-select questions, including whether it was about the current week's lecture, previous lectures, and content not covered in the lecture. To compare students' expectations of the exercise with their actual usage of ChatGPT, we provide the same nine options mentioned in the pre-survey (§3.1). Additionally, we collect data on students' confidence in their essay and overall satisfaction with ChatGPT's responses to analyze their experiences throughout the semester. Finally, students share how they utilized ChatGPT's responses to modify their essay and their experiences regarding it in a free-text form.

3.4 User Scenario

Undergraduate courses (IRW and AW) and graduate courses (SW) have different requirements, as SW is focused on writing scientific research papers while IRW and AW are intended for essay writing. To align with varying needs, we designed separate user scenarios.

For IRW and AW, we collect in-class timed essay data from students before they use our platform. Students are provided with a prompt by their instructors and asked to write an argumentative essay. The essay is written in class within 40 minutes with pens and pencils, ensuring that they cannot rely on online tools for assistance. The essays are digitized and stored in our database. Throughout the course, students revise their essays utilizing our platform through multiple iterations. The conversation and revision are based on what they learned from the course, as students give the first prompt indicating the contents covered in the class. Students also rate the helpfulness of every response from ChatGPT. After each session, their interim essays are saved for use in subsequent iterations.

Unlike IRW and AW courses, SW students submit a paragraph of their choice to get help with while using the platform. This is because the course is designed to write a complete research paper over the semester by adding sections, and it cannot be completed during an in-class timed setting like essay writing. Therefore, data shown in Table 1a is an optional element in the platform for SW. Students with a paragraph can revise the essay and converse with ChatGPT, similar to IRW and AW user flow. Whereas students who do not submit any paragraph to edit can still review the course content and have conversations with ChatGPT, a personalized English writing teacher, to deepen their understanding.

We illustrate actual interaction data collected from RECIPE with students taking IRW, AW, and SW. According to the collected data, the student taking SW followed the user scenario as we intended. The student described the content of the class in the first prompt, which is focused on methodology, and submitted the methodology section of his or her own paper. Furthermore, after receiving ChatGPT's response suggesting three possible improvements to the paper, the student incorporated one of the suggestions into his or her own writing. In addition, a student taking IRW asked, "Is 7 and 8 grammatical error?" after ChatGPT listed up a revised version with ten grammatical errors. The student continues to question the response from ChatGPT by switching to Korean from English. Both students did not simply accept all the suggestions but sought clarification on suggested revisions that they did not understand.

Table 1: Data collected during each phase of the platform.

	Pre-survey	Writing Exercise	Post-survey
(a) Essay	• Students' original essay	• Student's edited essay and its revision history	• Reason not to revise essay • Confidence on students' essay
(b) ChatGPT		• Conversation log data • Helpfulness of each ChatGPT response	• Topics discussed with ChatGPT • Overall satisfaction towards ChatGPT responses
(c) Student	• Expected help for the session • Comprehension level about course topics	• Timestamp of each sent message	• Actual help through the session • Improvements in comprehension level after writing exercise

Table 2: Initial prompts at writing exercise

(a) Hidden prompt for ChatGPT	(b) Open prompt for students
Act as an English writing class teacher and instruct a student to revise an essay based on the content learned during class. The student will provide you with a summary of what they learned. Please guide them step by step and do not revise or rewrite the essay at once. If you understand, please say yes only.	Hello! Welcome to the main exercise. Can you please tell me what you learned in class so that I can help you in a better way? An example can be: "Today I learned about comma rules. When there is a list that contains two or more elements, we should use commas to separate them."

Therefore, this suggests that students are engaged in using RECIPE, and perform critical thinking, indicating the potential benefit of RECIPE for learning effect.

4 INTERVIEW FOR FURTHER DEVELOPMENT

We conducted two interviews to further scrutinize the needs for future work on RECIPE: (1) a focus group interview among six students who had already taken at least one EFL writing course and (2) a 1:1 in-depth interview with an instructor who had taught EFL courses for 6 years, including IRW and AW. Three students (S1, S2, S3) had taken undergraduate courses (IRW and AW), and the other three students (S4, S5, S6) had taken a graduate course (SW). We asked both students and instructor to use RECIPE and suggest future directions to ChatGPT-integrated educational platforms. In general, both of them acknowledged the usefulness of ChatGPT in EFL writing and expressed the need for a specialized platform integrated with ChatGPT in the EFL writing courses.

Recommendation for optimal prompts. Although we provided instructions on the initial prompt and an example for the writing exercise, S5 and S6 reported difficulties in initiating a dialogue with ChatGPT. S1 mentioned that sharing prompts that had effectively worked in the focus group was helpful. Both S1 and S4 recommended implementing a tool to share the learning history of each student, including successful prompts and dialogue pattern logs. We believe that the assessment data on each student's prompt and ChatGPT's response collected through this platform can be utilized to enhance RECIPE's prompt engineering techniques and build a recommendation model for optimal prompts.

Personalized persona setup. The instructor noted the need for personalized responses from ChatGPT based on students' situations, such as English proficiency level. Additionally, S1 observed that a detailed persona setup with diverse perspectives, such as EFL professors, domain experts, and slightly advanced level EFL learners, made ChatGPT generate insightful critiques. RECIPE's persona, acting as an instructor of EFL writing courses, can be broadened in a diverse way so that students may choose the persona that best suits their purpose.

Student-initiative platform design. The instructor remarked that continuous trials and errors in interactions with ChatGPT about revising their essays would be essential for their language learning. The instructor also commented that those interactions led by students' initiative would facilitate their writing skills. In particular, S1 suggested generating errorful essays relevant to what students learned through ChatGPT as learning materials and asking them to revise the essays by themselves step by step.

Automated essay scoring. The instructor emphasized the need for automatically-measured essay scores and explanations for evaluation based on their own rubrics or grading systems. While some state-of-the-art automated essay scoring (AES) systems for overall scores exist, it is nearly inaccessible for EFL instructors who don't have much background in AI techniques to incorporate and fine-tune them based on their own rubrics. We believe that instructor-friendly platforms based on interactions with ChatGPT may mitigate this issue.

5 CONCLUSION

In this paper, we investigated students' perception and usage of generative AI, including ChatGPT, in academic courses. Our results indicated that most students reported positive experiences with ChatGPT for general, academic, and essay-writing purposes. However, students with a limited understanding of LLMs faced challenges as expecting ChatGPT to find ground sources for their writings, and this hallucination is one of the major limitations of generative AI. To address this issue, we introduce RECIPE, a novel language learning platform, that leverages ChatGPT to cater to students' needs and enhance their English writing skills with two types of prompts. We aim to collect interaction data throughout the semester on students' perception and usage of ChatGPT in English writing through RECIPE. The interaction data we gather with RECIPE has the potential to serve as a baseline for developing an effective paradigm for integrating ChatGPT in academic settings.

ACKNOWLEDGMENTS

This work was supported by Elice.

REFERENCES

[1] Ariyanti Ariyanti et al. 2021. Technology-Enhanced Paraphrasing Tool to Improve EFL Students' Writing Achievement and Enjoyment. *Journal of English Language Teaching and Linguistics (JELTL)* 6, 3 (2021), 715–726. https://doi.org/10.21462/jeltl.v6i3.698

[2] David Baidoo-Anu and Leticia Ansah. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. https://doi.org/10.2139/ssrn.4337484

[3] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. arXiv:2302.04023 [cs.CL]

[4] Hui-Wen Huang, Zehui Li, and Linda Taylor. 2020. The Effectiveness of Using Grammarly to Improve Students' Writing Skills. In *Proceedings of the 5th International Conference on Distance Education and Learning* (Beijing, China) *(ICDEL 2020)*. Association for Computing Machinery, New York, NY, USA, 122–127. https://doi.org/10.1145/3402569.3402594

[5] Weijiao Huang, Khe Foon Hew, and Luke K. Fryer. 2022. Chatbots for language learning—Are they really useful?A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning* 38, 1 (2022), 237–257. https://doi.org/10.1111/jcal.12610

[6] Silvia Milano, Joshua A McGrane, and Sabina Leonelli. 2023. Large language models challenge the future of higher education. *Nature Machine Intelligence* (2023), 1–2. https://doi.org/10.1038/s42256-023-00644-2

[7] Junaid Qadir. 2022. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. https://doi.org/10.36227/techrxiv.21789434

[8] Jürgen Rudolph, Samson Tan, and Shannon Tan. 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching (JALT)* 6, 1 (2023). https://doi.org/10.37074/jalt.2023.6.1.17

[9] Miriam Sullivan, Andrew Kelly, and Paul McLaughlan. 2023. ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching (JALT)* 6, 1 (2023). https://doi.org/10.37074/jalt.2023.6.1.17