



From Errors to Improvement: Leveraging ChatGPT and Peer Feedback in L2 Writing

Mart Christine Johnston

Takushoku University, Japan

Abstract

This research discusses the integration of ChatGPT and peer noticing to enhance writing skills. Students wrote four essays and were tasked with utilizing ChatGPT-3 to correct any errors in their writing. Additionally, they engaged in peer noticing of errors to help them remember more of their errors. To determine whether learning was acquired or missed, individualized tests were administered, requiring students to correct at least five errors flagged by ChatGPT in all of their essays. Another method used was to count all correctly written sentences by those who completed all writing activities. Results show that while students demonstrated confidence in forming verbs in the past tense, they struggled with remembering multi-word items, understanding comma rules—especially for compound sentences—and constructing complex sentences. There was no significant difference in writing improvement, likely due to the missed learning opportunities provided by ChatGPT. This research underscores the critical need for instruction from language teachers, as second/foreign language students may not be ready to fully rely on generative intelligence technologies.

Keywords: ChatGPT-3, second language writing, peer noticing, writing errors, multi-word items

1. Introduction

“There is no substitute for hard work” is a famous quote attributed to Thomas Edison, who lived in the 19th and early 20th centuries. However, in 2022, ChatGPT (Chat Generative Pre-trained Transformer) was launched by OpenAI, and many academic problems seemed answerable with this AI software (Barbas et al., 2023), though it was not without its flaws (Kosinski, 2023). ChatGPT continues to develop and is predicted to achieve PhD-level intelligence by late 2025 or early 2026, according to OpenAI’s Chief Technology Officer, Mira Murati (Dartmouth Engineering, 2024).

However, many educators remain hesitant and pessimistic about introducing it in the classroom due to concerns about students committing plagiarism or cheating (Jimenez, 2023). In the context of second/foreign language (L2) classrooms, students could submit lexically rich essays, free of grammar errors through ChatGPT or translators. This paper describes a method to avoid such issues, that centers around conducting all second language writing assignments in the classroom using pen and paper, without access to electronic devices.

Rather than viewing AI as a hindrance to learning, this paper will examine the extent of learning gained and missed by university students in L2 classrooms using ChatGPT to improve their writing skills. ChatGPT was used by the students only as a post-writing tool in this research, and peer noticing was incorporated into their post-writing activities. The following section will discuss the methods used, followed by an analysis of the corrections flagged by ChatGPT that students learned from and missed.

2. Method

Forty-five Japanese students participated in this research. Generally, they had passed EIKEN Grade 2 or Pre-2, a popular English language proficiency test in Japan. Students completed four writing assignments throughout the semester, based on topics discussed in their textbook. The writing assignment questions were as follows:

- W1: Which season most reminds you of your childhood? Why do you feel this way?
- W2: When you were younger, did you enjoy learning English? Why or why not? Provide three reasons.
- W3: What is the best place you have ever traveled to, and why?



- W4: Describe a memorable online shopping experience you have had. What did you purchase, and why did you choose to buy it online rather than in a physical store?

Students were given a scaffolding pattern used across all four writing activities, which was: The _____ was _____ because _____ (supporting sentence). Another reason was that _____ (supporting sentence). Finally, _____ (supporting sentence). This approach aimed to prioritize grammar focus and assess students' knowledge of past verb forms and dependent clauses.

For W1, students wrote essays of 50 to 70 words within 12 minutes. Subsequent writing activities required essays of 70 to 90 words, with 20 minutes allotted for each essay. After collecting the papers, I noted the errors and categorized them grammatically, recording them in an Excel sheet. The maximum score students could receive was eight points. The following week, their essays were returned with only their preliminary scores indicated, and without any comments on their work.

Students were required to input their essays into ChatGPT-3 using this prompt: "Can you correct the grammar errors in this essay: _____?" The prompt was specific because ChatGPT tends to alter wording, words, and ideas if other prompts, such as "Can you rewrite this essay?" are used. The prompt for post-writing activities was designed to focus on correcting grammar and awkward phrases while preserving the students' ideas and vocabulary. For completing this part of the exercise, students were awarded an additional two points.

After students noted their errors, they discussed with their peers what they learned from ChatGPT. This method was implemented from W2 to W4, following a suggestion by Jim Ronald during a research presentation at the JALTCALL 2024 conference. Peer discussion was believed to help students remember their errors better. Each student also shared what they had learned from ChatGPT in class, and I took note of these learning items.

All errors and comments for each student's essay were inputted into AntConc 3.5.8 (Anthony, 2019) to identify the top grammar categories that needed improvement. Learning items mentioned by students were also analyzed with the same software. The top six grammar categories, representing the most common errors and learning items, were used as a guide for the final test (these errors are listed in the Results and Discussion section below). This 5-item individualized test assessed whether students could remember the errors flagged by ChatGPT, with a focus on the top six grammar categories. If a student did not commit any of the errors, their other errors were included in their individualized test.

To determine whether students had gained knowledge or missed learning items flagged by ChatGPT, correctly written sentences were counted. While all 45 students participated in the final test, only 33 students who completed all four essays and participated in the post-writing activity (using ChatGPT as a proofreader) were included in the second analysis method. The scores were analyzed using ANOVA to determine the F-value or p-value and assess whether the differences among students' scores were statistically significant.

3. Results and Discussion

This section provides the final list of the most common errors in the students' essays. It includes mistakes I noticed as well as those identified by the students while using ChatGPT. First, here are the key errors I observed in their essays:

1. Preposition
2. Past tense
3. Pluralization
4. Dependent clause
5. Comma
6. Spelling
7. Article
8. Fragments
9. Chunks/Collocation

Student feedback indicated that the top grammar items learned from ChatGPT were:

1. Preposition



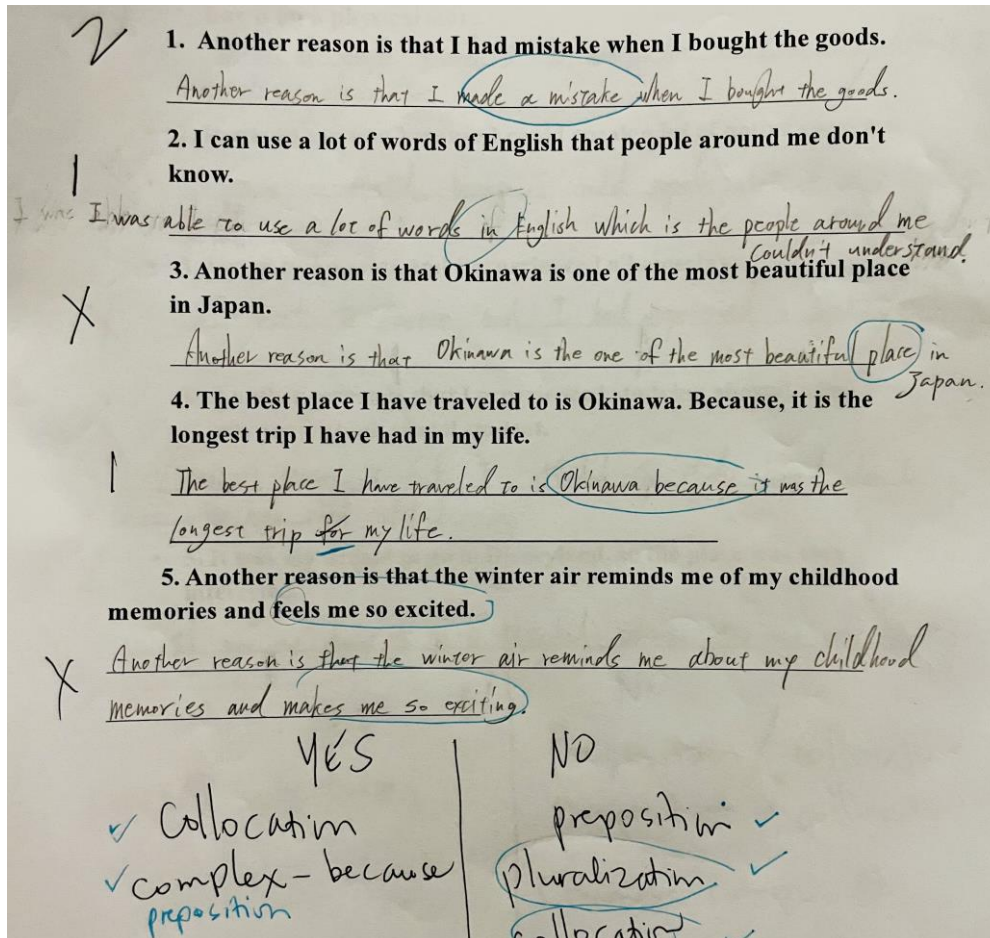
2. Past tense
3. Comma
4. Spelling
5. Pronoun
6. Chunks/Collocation
7. Adjective
8. Article

Therefore, the top six grammar items used to develop their individualized tests were:

1. Past tense
2. Preposition
3. Pluralization or generalization in the plural
4. Dependent clause using "because" as the connector
5. Comma rules in compound sentences and with multiple items
6. Collocations/chunks as replacement for awkward phrases

Figure 1 shows a sample individualized test. The student was required to correct only five sentences, and each item was worth two points. Below the test items, errors the student corrected and errors they missed were recorded. For example, if the student missed two preposition errors, these were tallied as two preposition errors in the Excel sheet. In Table 1, column A displays the errors that students corrected, while column B shows the errors they did not correct. The students were expected to correct a total of 235 errors cumulatively. They successfully corrected 97 errors, which constitutes 41%, while 138 errors, or 59%, were not corrected.

Fig. 1. A sample individualized test



2 ✓ **1. Another reason is that I had mistake when I bought the goods.**
Another reason is that I made a mistake when I bought the goods.

1 | **2. I can use a lot of words of English that people around me don't know.**
I was able to use a lot of words in English which is the people around me couldn't understand.

X | **3. Another reason is that Okinawa is one of the most beautiful place in Japan.**
Another reason is that Okinawa is the one of the most beautiful place in Japan.

1 | **4. The best place I have traveled to is Okinawa. Because, it is the longest trip I have had in my life.**
The best place I have traveled to is Okinawa because it was the longest trip for my life.

X | **5. Another reason is that the winter air reminds me of my childhood memories and feels me so excited.**
Another reason is that the winter air reminds me about my childhood memories and makes me so exciting.

YES	NO
<ul style="list-style-type: none"> ✓ Collocation ✓ complex - because preposition 	<ul style="list-style-type: none"> preposition ✓ pluralization ✓ collocation ✓



Table 1. Individualized test results

Items learned	Missed learning points
Double verb recognition-1	Subject-verb agreement-1
<i>For example</i> changed to <i>such as</i> -1	Verb phrase-1
fragment/ clause completion-1	Present perfect-1
gerund-1	missing verb-1
comma-1	Modal - <i>could do</i> -1
comma-complex-1	Modal verb rule-1
Because of complex-1	fragment-1
Adverb-1	complex-1
Present perfect - 2	compound-1
article- 2	double verb-2
comma-multiple-2	Fragment-starting with <i>especially</i> -2
pronoun-3	adjective clause-2
comma-compound-3	article-2
Correct vocabulary-5	spelling-3
because-complex-7	Fragment starting with <i>for example</i> -3
collocation-preposition-9	Correct vocabulary-3
collocation-9	comma-multiple-3
preposition-14	past (can- could)-6
pluralization-16	Past tense-7
Past tense-18	collocation-preposition-10
Total 97 = 41%	pluralization-10
	preposition-14
	because-complex-19
	comma-compound-19
	collocation-24
	Total 138 = 59%

The results show that students had a good grasp on the use of past tense if the context is in the past. Students also learned to pluralize or generalize items in the plural form and use appropriate prepositions for certain nouns. For clarification, the difference between “collocation” and “collocation-preposition” is that the former represents students remembering proper chunks to be replaced for awkward phrases they initially wrote while the latter means that the students were able to supply appropriate prepositions that go along with the given verb. There were seven corrections made that represented their knowledge on writing complex sentences beginning with “because.” Only three corrections were made for the use of commas in compound sentences.

Conversely, the top writing error that was missed by many students were collocations which suggests there is a need for students to learn multi-word items instead of single vocabulary words. Collocation-preposition error means that the students were not able to supply the appropriate prepositions for either nouns or verbs. In this case, it can also be beneficial for students to learn phrasal verbs. There were several unnoticed errors in terms of complex and compound sentences. This means that AI is



not sufficient for them to understand how to construct complex and compound sentences. The use of prepositions in prepositional phrases was also deemed challenging for the students. In terms of verb conjugations, it is worth noting that the past tense for modal/irregular verb *can* was not known to six students. There were also students that wrote fragments beginning with “especially” and “for example.” Such an issue calls for a need for practice in identifying complete and incomplete sentences.

Regarding the counting of correctly written sentences by 33 students, the results returned an F-value of 1.606, indicating a slight difference in their writing abilities, but not a substantial one. The p-value was 0.193, which suggests that the differences in their writing scores are not statistically significant. The lack of significant evidence for writing improvement may be attributed to the students’ insufficient knowledge of multi-word items and other learning items from ChatGPT that they missed mentioned in previous discussions.

4. Conclusion

This research explored whether L2 students can learn from ChatGPT and improve their writing through peer noticing of errors. All four writing activities focused on the past tense, making this grammar category the primary learning item for the students. However, students showed inconsistent performance in pluralization and the use of prepositions. Significant errors persisted in the areas of multi-word items, comma rules for compound sentences, and complex sentence constructions. Statistically, there was no significant evidence of writing improvement among the students, which relates back to the learning items they missed. The findings highlight the need for intensive instruction on comma rules, various sentence types (including complex and compound), and a greater emphasis on teaching multi-word items rather than focusing solely on single-word vocabulary. Maintaining the use of ChatGPT as a post-writing tool, along with peer noticing in writing activities, proves beneficial in promoting independent learning and teaching students to leverage AI for their own educational growth. However, this approach requires further scaffolding from language teachers to maximize its effectiveness.

REFERENCES

- [1] Anthony, L. (2019). AntConc (3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- [2] Barbas, M. P., Vieira, A. T., & Branco, P. D. (2023, October). The importance of Chat GPT training for Higher Education: case study. In *International Conference on Design and Digital Communication* (pp. 695-705). Cham: Springer Nature Switzerland.
- [3] Jimenez, K. (2023, February 2). 'This shouldn't be a surprise' the education community shares mixed reactions to ChatGPT. USA Today. Retrieved February 27, 2023, from <https://www.usatoday.com/story/news/education/2023/01/30/chatgpt-going-banned-teachers-sound-alarm-new-ai-tech/11069593002/>
- [4] Kosinski, M. (2023). Evaluating large language models in theory of mind tasks. *arXiv e-prints*, arXiv-2302.
- [5] Dartmouth Engineering. (2024, June 19). AI everywhere: Transforming our world, empowering humanity [Video]. YouTube. <https://www.youtube.com/watch?v=yUoj9B8OpR8>