

Language Assessment and Learning through AI Technology: An Exploratory Study on Using GPT for Young EFL Learners' Writing

Mikyung Kim Wolf*

ETS, USA

Michael Suhan

ETS, USA

Correspondence

Email: mkwolf@ets.org

Abstract

With the rapid advancement of generative AI technology using large language models, various AI tools have emerged for educational use. Among these tools, ChatGPT has gained traction in second language (L2) writing due to its significant potential as a writing assistant for ESL/EFL learners. This study explored how young EFL students interact with and utilize feedback generated by GPT-4 for their writing. Conducted within the context of developing an automated writing feedback tool to support potential TOEFL Junior Writing test users, this article reports on our small-scale prototyping study, detailing the development of the prototype tool. Building on the extensive work by Carol Chapelle on the use of technology in L2 learning and assessment, our study offers new insights into young EFL students' use of generative AI technology in their L2 writing development. The implications of our findings for further research and practice for using generative AI technology in language assessment and learning for young learners are discussed.

ARTICLE HISTORY

Received: 06 August 2024

Revised: 23 April 2025

Accepted: 12 December 2025

KEYWORDS

Automated Writing Feedback, Generative Artificial Intelligence (AI), Young EFL Learners, Technology, Writing Assessment

How to cite this article (APA 7th Edition):

Wolf, M. K., & Suhan, M. (2025). Language assessment and learning through AI technology: An exploratory study on using GPT for young EFL learners' writing. *Language Teaching Research Quarterly*, 50, 77–100. <https://doi.org/10.32038/ltrq.2025.50.07>

¹Introduction

Undoubtedly, Carol Chapelle is one of the most influential scholars in the field of applied linguistics, particularly noted for her expertise in language assessment and technology

¹ This paper is part of a special issue (2025, 50-51) entitled: In honour of Carol A. Chapelle's contributions to language assessment and learning (edited by Christine Coombe, Tony Clark, and Hassan Mohebbi).

applications in language education. Since the 1990s, Chapelle has provided both theoretical underpinnings and methodological guidance on the use of technology and its evaluation in second language (L2) education and research. Her seminal work, *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research* (Chapelle, 2001), is one such example that comprehensively accounts for the application and evaluation of early technology in L2 development, assessment, and research. Chapelle's consistent and pervasive contributions to advance this area are not only through her numerous articles but also through her leadership in convening the annual Technology for Second Language Learning (TSSL) conferences over the past two decades. The TSSL conferences have been an unparalleled avenue for accessing the latest research and stimulating ideas among researchers on the application of technology in L2 education around the world.

In a recent TSSL proceeding, Chapelle and her colleagues asserted, "This is the time for us as language professionals to increase our own critical engagement with language and language technologies in a world never so fascinated by language" (Chapelle et al., 2024, p. 535). This assertion is partly due to the groundbreaking emergence of generative artificial intelligence (AI) using large language models. The introduction of ChatGPT by OpenAI in November 2022 has ushered in a new era, expanding the use cases of AI technology into daily life with human-like interactions. While AI technology involves interdisciplinary expertise from fields such as computer science, data science, and linguistics, its appropriate and effective application in language education inevitably requires applied linguistics expertise, encompassing both theoretical and empirical understanding of L2 acquisition, teaching, and assessment. Chapelle's commentary, *Research Questions for a CALL Research Agenda: A Reply to Rafael Salaberry*, published in the journal *Language Learning & Technology* (Chapelle, 1999), remains relevant today. Her insight and guidance on formulating research questions for language researchers are still applicable for exploring current generative AI technology. For instance, grounded in the principles of interactionist theory, Chapelle has emphasized research questions focusing on L2 learners and their interrelationship with contextual aspects (e.g., computer-assisted language learning software, activities, teaching). Amid the rapidly growing applications of generative AI in L2 education, her research frameworks, methodologies, example questions, and directions offer valuable guidance for language researchers to tackle current technological issues in language education.

In this article, we present our recent study that has examined young L2 learners' use of a writing feedback tool utilizing GPT-4. One of the compelling use cases of generative AI models like GPT-4 for L2 learners lies in the automated evaluation and feedback provision for writing skills. While there is growing research on the potential of ChatGPT to provide feedback for L2 learners' writing, there has been little research examining young English-as-a-foreign-language (EFL) learners' use of writing feedback generated by GPT-4. As Chapelle et al. (2024) note, it is essential to first examine how L2 learners

use technology to better understand its effects on L2 development. In what follows, we first provide a review of previous studies on the use of ChatGPT for feedback provision in the context of L2 writing. Then, we describe our study context involving young EFL students in developing and utilizing a writing feedback prototype tool with features utilizing GPT-4. Based on our prototyping study, we discuss our findings and their implications for further research and practice in the use of generative AI technology to support young EFL students' writing skills.

Previous Research on the Use of ChatGPT for L2 Writing Feedback

Although research on the use of ChatGPT in L2 education is expanding, it remains in its early stages due to the tool's recent introduction. There are only a few studies exploring ChatGPT's potential for providing L2 writing feedback at the time of this writing (Allen & Mizumoto, 2024; Banihashem et al., 2024; Guo & Wang, 2024; Koltovskaia et al., 2024; Lu et al., 2024; Teng, 2024). These studies primarily focus on the quality of feedback generated by ChatGPT and students' perceptions of it. Generally, researchers have found that ChatGPT provides more detailed and coherent feedback than human teachers. Students were also found to generally prefer the ChatGPT-generated feedback because of its corrective feedback, constructive suggestions, and explanations for the feedback given. Furthermore, researchers have highlighted the benefits of ChatGPT feedback in enhancing student engagement in the writing process, including revision and editing, as well as in promoting personalized learning. However, they also caution about the need for careful prompting to elicit appropriate feedback and the importance of students' critical thinking in evaluating the adequacy of the feedback and applying it to improve their writing skills. Notably, these studies primarily involved L2 learners at the undergraduate or graduate levels. Given the increasing use of technology in language learning among young L2 learners, it is vital to explore how they interact with AI technologies like ChatGPT.

The Context of the Study

This study was conducted within the context of developing an automated writing feedback tool intended to support potential TOEFL Junior Writing test users. The TOEFL Junior Writing test, created by ETS, was recently relaunched with modifications in 2023 (the previous version was in use from 2013 to 2017). The test is designed to assess students' computer-based English writing skills in social and academic contexts. The target test-takers are students who learn English as an additional language, either as a second or a foreign language, targeted at students aged 11 and older (ETS, 2019). One of the main modifications in the current TOEFL Junior Writing test is the employment of ETS's automated scoring model, which was developed based on a large sample of human-scored writing responses. The adoption of automated scoring has enabled more consistent and faster turnaround times for scoring.

With advancements in AI technology for writing evaluation, a team of ETS researchers and assessment specialists initiated a research project to explore how technology can provide useful feedback for students preparing for the test. This project aimed to develop an AI-assisted writing feedback tool to help students practice writing tasks and improve their writing skills. As an initial step, we developed a prototype tool using GPT-4 and ChatGPT features. We then investigated how young EFL students interacted with and utilized the feedback it provided. It was crucial to engage target users early in the development process to ensure the tool's usability and usefulness. In this article, we describe the development of the prototype tool and report on our small-scale prototyping study.

Prototype Development

Overview of the Writing Tasks

Although the TOEFL Junior Writing test includes four task types (Edit, Email, Opinion, Listen-Write), the scope of the present study included only the Opinion task. The Opinion task requires students to write a paragraph to express their opinion and provide supporting details in response to a statement or a question in the writing prompt. Task scores are based on a holistic five-point rubric, ranging from 0-4, with descriptors covering content, organization, and language use. The descriptors of the highest and lowest score points of the Opinion task rubric are provided in Appendix A.

For the prototype tool, we created three prompts, following the same format for the Opinion task. The topics were related to computer games, homework, and pets, providing students with options to choose from for their practice writing.

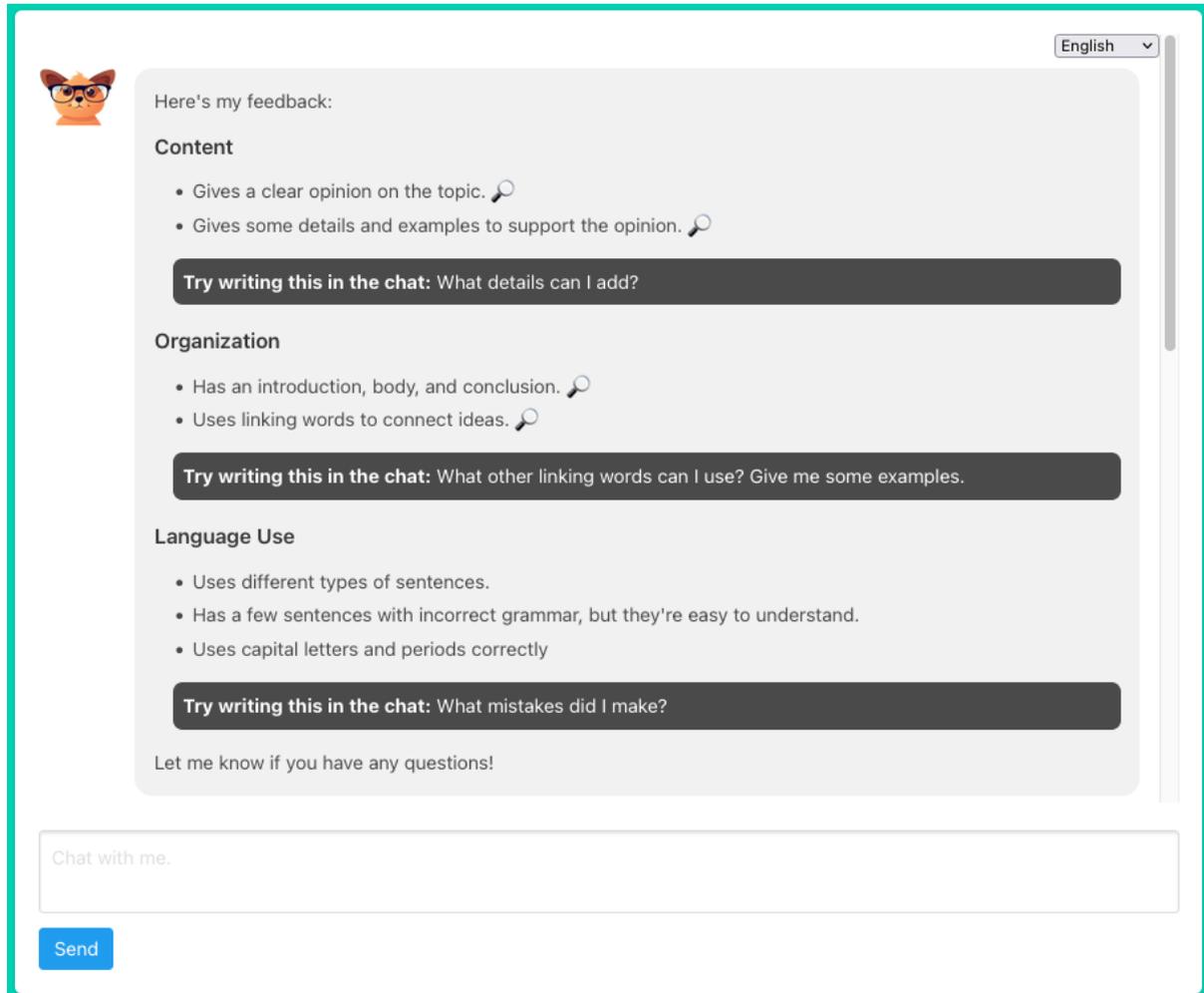
Prompt Engineering to Generate Rubric-Based Feedback

The prototype tool was designed to offer personalized feedback tailored to individual students' needs without providing a score for their writing. We utilized the scoring rubric of the Opinion task in our prompt engineering to ensure the feedback aligns with the constructs and rubric criteria of the test. Additionally, we implemented several key principles in prompt engineering to provide appropriate feedback language: being age-appropriate, constructive, positive, and using simple language.

All feedback was generated using OpenAI's GPT-4 model via the ChatCompletion API, but different approaches to prompt engineering were developed for each type of feedback. Prompts were improved iteratively through several rounds of review by a team of four subject matter experts. Each round, the team used the prototype's user interface to review sample feedback generated by the prototype for accuracy and appropriateness to advise on how to improve the feedback. Sample feedback was generated using responses to the Opinion task in Wolf et al. (2024) and from the TOEFL Junior Writing test. In the prototype, feedback responses generated by the model were validated to ensure that they were in the correct data structure and regenerated if not.

We designed the tool in a way where students receive initial feedback about their writing based on the rubric organized by the dimensions of content, organization, and language use. Figure 1 displays a screenshot of a sample of the initial rubric-based feedback generated by the prototype tool.

Figure 1
Screenshot of Rubric-Based Feedback



The prompt for the GPT-4 model used to generate this initial rubric-based feedback began with a brief description of the purpose for the feedback and the learning context. This was followed by instructions for how to write feedback based on the task rubric. The prompt also included instructions for how to structure the output. Few-shot prompting was implemented by providing nine examples of input and output. Unlike zero-shot prompting, which provides no examples to the model, few-shot prompting includes a small number of input-output examples, formatted to help the model infer a pattern and guide its response to real input. Few-shot examples were included in the prompt “in the form of example user/assistant interactions after the initial system message” (Microsoft, n.d.), which contained previously described prompt components. Each example input included an essay prompt and an Opinion task response from Wolf et al. (2024) or the

TOEFL Junior Writing test; writing prompts were included in the few-shot examples so that the same model prompt would work with different writing task prompts. Example inputs were selected to represent a range of proficiency levels as well as response characteristics such as severe incoherence, going off-topic, and writing parts of the response in a language other than English. At the end of the sequence of example interactions, the student's response was appended, which cued the model to generate feedback structured like the example outputs in the sequence. Table 1 provides an overview of the prompt using terminology from Microsoft's Azure OpenAI Service Documentation (Microsoft, n.d.).

Table 1*Overview of Rubric-Based Feedback Prompt*

| Data | Prompt Component | Content | |
|--------|----------------------|---|-------------------------------------|
| Input | Supporting content | Description of purpose and learning context | |
| | Complex instructions | Instructions for generating feedback | |
| | Supporting content | Rubric | |
| | Complex instructions | Instructions on data structure | |
| | Few-shot examples | | Example essay 1 with writing prompt |
| | | | Example feedback on essay 1 |
| | | | ... |
| | | | Example essay 9 with writing prompt |
| | | | Example feedback on essay 9 |
| | Primary Content | Essay written by student | |
| Output | | Rubric-based feedback | |

Note. Supporting content and instructions were included in the system message sent to the ChatCompletion API, followed by few-shot examples and primary content included as user/assistant interactions.

In the first iterations of prompt development, the holistic Opinion task rubric was included unedited as freeform text with numeric scores included. However, the model frequently miscategorized descriptors, generated feedback using difficult language, or included task scores in the feedback despite being instructed not to. To mitigate these issues, we created an analytic rubric using simplified language to include in the prompt. The new rubric was included in XML format, following the recommendation in the Azure OpenAI Service Documentation to use XML as a means of providing "clear syntax" (Microsoft, n.d.). Task scores in the rubric were replaced with descriptions of writing ability (starting, growing, good, and very good), while the qualities associated with scores of zero were described in the instructions instead of the rubric (e.g., "Check if the essay is off-topic or not, and leave feedback about it being off-topic if it is.").

Prompt Engineering for the Chat Agent's Feedback

To generate feedback responses from a chat agent similar to ChatGPT, a process combining few-shot prompting and prompt chaining was used. The prompt first included instructions to provide "overall feedback" and to then respond to questions about the feedback. Thereafter, it was explained that two essays (the first essay and the second essay) would be the subject of evaluation and questioning and that the content and feedback from the first essay must be kept secret from the student receiving feedback on

the second essay. Few-shot prompting was then implemented by providing the following input/output examples:

- The writing prompt and content of an example essay (i.e., essay 1)
- Rubric-based feedback on essay 1 (i.e., the “overall feedback”)
- A sequence of 6 questions and answers about essay 1
- The writing prompt and content of the latest revision of the student essay (i.e., essay 2)
- Rubric-based feedback on the student essay (described previously)

When a student asks the chat agent a question, the question is appended to the end of the prompt to cue an answer to the question. In later questions asked by the student, earlier questions and answers are included in the prompt to create context for the chat agent, and the student essay (the second essay) is updated with the latest revision. The final questions about the first and second essay are prefixed with “Question:” to support cueing answers to questions. Without this prefix, the response would at times respond with outputs resembling the rubric-based feedback provided as few-shot examples.

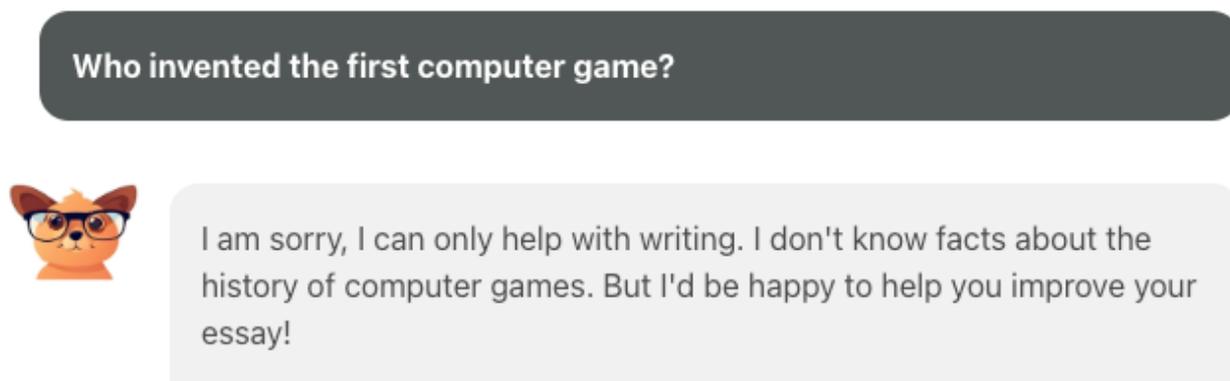
In the first part of the prompt, prior to the few-shot examples, contextual information along with specifications for how to appropriately respond to and format answers to questions are included, with instructions on how to use HTML tags to format the text. In line with evidence that using a personable, conversational style in instructional content can promote learning (Clark & Mayer, 2023; Moreno & Mayer, 2004), instructions state: “Be friendly and personable. Use emojis when appropriate.” The following instructions for how to respond to questions are included as guardrails to minimize the possibility of generating responses that do the work of writing for the student, contain hallucinations, or enable counterproductive behavior:

- Do not engage in off-topic discussion.
- Do not rewrite the entire essay for students. If asked for an example, only assist with providing ideas; do not provide entire sentences or paragraphs that incorporate the new ideas. If students ask you to write parts of the essay for them, offer sentence starters or suggestions for ideas as support.
- Do not translate the entire essay from a different language into English.
- Do not provide facts to support the content of the essay; inform the students that you can only give feedback on writing and cannot tell them facts about science, history, and so on.
- If students write anything offensive or behave badly, remind them that their teacher or another adult may read the chat.

Members of the project team tested the efficacy of the guardrails by attempting to elicit undesirable responses from the chat agent. Figure 2 shows an example message from a project team member requesting factual information and the chat agent’s response.

Figure 2

Screenshot of Request for Factual Information during the Testing Phase



Additional Feedback Features

In addition to the TOEFL Junior Writing rubric-based feedback and interactive feedback features using a chat function, our prototype tool contained features of (a) providing support in the student's first language (L1, in our study, Japanese and Korean), (b) highlighting areas in students' writing for feedback, and (c) allowing for interactive revisions based on continuous feedback from the chat agent.

Translation of the rubric descriptors was also generated using GPT-4. This approach to translation was used given the breadth of languages that GPT-4 can potentially translate text into. The translation prompting included the following instructions, with the name of the language being translated dynamically added to the prompt:

You are a translation API. Translate the text into {language}. Do NOT change the data structure. Do NOT translate text inside quotes. Quotes should be in English. Return only the translation.

One-shot prompting was implemented by including a single input and output example after the instructions. The input example was structured in the same format as the output example included in Appendix B, which includes dimension names and feedback. The translation output example was a Japanese translation. The rubric-based feedback output was appended at the end of the input/output example to cue the generation of translated feedback. Students were also allowed to use their L1 to ask questions within the chat agent.

When the tool provided the initial rubric-based feedback, it displayed magnifying glass icons (see Figure 1). When students selected an icon, specific text was highlighted to indicate the areas for corresponding feedback. This feature was possible by generating an annotated version of students' writings. For prompt engineering, a brief description of the process was described, followed by instructions on how to appropriately level the

language and structure the output. Few-shot prompting was used again, providing the model with three examples of input and output.

With respect to the revision feature, students could edit their first draft after receiving the initial rubric-based feedback and while interacting with the chat agent in the tool. The chat agent enabled students to ask for clarification on the initial rubric-based feedback or to seek feedback on areas not addressed by the initial feedback. Feedback provided by the chat agent was based on the most up to date version of the student's draft.

User Interface

The user interface consists of four pages: an index page where tasks are selected, a page for writing the first draft, a page for revising the draft with the support of AI feedback, and a page displaying completed tasks. The development of the user interface incorporated research-informed best practices following the principles for designing multimedia learning content prescribed by Clark and Mayer (2023). From the index page, students are presented with general instructions and a list of writing tasks which they may begin, resume, or view if completed. Each available task includes a relevant image and the text of the writing prompt. The completed task page presents the first draft side-by-side with the final draft, abiding by the spatial contiguity principle of instructional design by ensuring that both texts will be on screen simultaneously (Clark & Mayer, 2023).

After submitting the first draft, students are taken to the revision page. As shown in Appendix B, this page includes two columns. The writing prompt, a text area where students may edit their draft, and an information box are in the left column, whereas the rubric-based feedback, which is integrated with the chat agent, is in the right column (Appendix B). When the page first loads, to reduce cognitive load by directing attention to key instructional elements (Noetel et al., 2022), a callout pointing to the text area signals to students that they should edit their writing.

In the chat area of the right column, students may interact with a chat agent to ask questions and receive feedback on revising their writing (see Appendix B). Options for rating the chat agent's response as "Helpful" or "Not helpful" are included for the purpose of collecting students' perception during the prototyping study. Students may interact with the chat agent in any language, and likewise, responses may be in English, another language, or multiple languages, which is based on the input from the student message. To scaffold use of the chat agent, example questions written by the project team related to each dimension of feedback are provided below each dimension's descriptors.

An avatar of an anthropomorphized animal was created using the Stability AI SDXL model. The decision to use an avatar was informed by research evidence suggesting that pedagogical agents are generally preferred by learners (Schroeder & Adesope, 2014) and

may increase motivation by providing a social presence (Clark & Mayer, 2023; Schroeder & Adesope, 2014). Following the advice of Clark and Mayer (2023) to mitigate the risk of dividing learners' attention between the learning task and the avatar, a static two-dimensional image was used. The choice to use an animal of indeterminate speciation as the avatar, as opposed to a human, was made in light of evidence suggesting that non-human avatars do not diminish the efficacy of e-learning tasks (Fiorella & Mayer, 2022; Castro-Alonso et al., 2021).

Prototyping Study

We conducted a small-scale, exploratory study to gather target users' experiences with the prototype tool. Specifically, we addressed the following research questions (RQs):

RQ₁: What types of feedback do students seek in the tool?

RQ₂: How do students use feedback in their revisions?

RQ₃: What are students' perceptions of the feedback tool?

The findings of the study are expected to inform areas of further development of the tool. Additionally, the empirical data produced from this study will help us better understand how young EFL students use AI technology to improve their writing skills.

Participants

Two extracurricular English programs in Japan and South Korea were selected as data collection sites mainly due to convenience sampling. The program in Japan was a summer camp that included English instruction, while the program in South Korea was a private English language academy offering after school classes to students. Teachers of students within the intended age range of the TOEFL Junior tests were asked to invite their students to voluntarily participate in the study. This process led to the recruitment of 4 students from Japan and 10 students from Korea.

A background information questionnaire was included at the beginning of a survey administered to students who had completed the writing activity. The mean age of the participating student was 12.8 years (mostly 11-14). No students in the Japanese sample had taken any of TOEFL Junior tests before, but seven students from the Korean sample had taken one of the TOEFL Junior tests before. Students reported that they had been studying English for 4-10 years, on average for 6 years ($SD = 1.6$) and that they had been studying English writing from 1-8 years, on average for 4.2 years ($SD = 2.2$). Table 2 lists background information for individual students.

Table 2*Student Background Information*

| ID | Country | Age | TOEFL Junior Test Previously Taken |
|------|-------------|-----|---------------------------------------|
| JS01 | Japan | 14 | No |
| JS02 | Japan | 17 | No |
| JS03 | Japan | 14 | No |
| JS04 | Japan | 12 | No |
| KS01 | South Korea | 13 | Yes |
| KS02 | South Korea | 12 | No |
| KS03 | South Korea | 13 | No |
| KS04 | South Korea | 11 | Yes |
| KS05 | South Korea | 12 | Yes |
| KS06 | South Korea | 12 | Yes |
| KS07 | South Korea | 14 | No |
| KS08 | South Korea | 13 | Yes |
| KS09 | South Korea | 11 | Yes |
| KS10 | South Korea | 11 | Yes |

Procedure

The teachers first explained the tool and the writing task to the students. The students wrote a first draft upon choosing one of the writing prompts and then received the initial rubric-based feedback. The three prompts were as follows:

- Do you agree or disagree with the following statement? Computer games help students develop useful skills.
- Some students think that homework is important. Others think that it is not necessary. Which do you agree with?
- Some people think that having pets teaches responsibility. Others think that pets are too much work. What do you think?

All prompts included the same set of instructions stating to write a paragraph expressing an opinion on the topic supported by reasons and examples. No time limit was imposed during any stage of the task. As described earlier, example questions were provided for students to use. Students were also instructed to enter their own questions in the chat. They were advised that they would receive feedback on their response and that they should use the feedback to revise their text. After completing the revision, students were directed to a screen showing a side-by-side comparison of their first and final draft. Students were asked to take a survey upon completion of the writing activity.

Analysis

Quantitative and qualitative analyses of chat and process data were conducted to gain insights on what types of feedback students utilized (RQs 1 and 2). Chat messages were counted per student; each message was classified according to its intent and the language used (i.e., L1 or English). Similarly, process logs were analyzed to determine if students used the feature that translates rubric-based feedback and to what extent students used

the highlightable feedback associated with rubric descriptors. The number of descriptors varies, and not all may have associated highlightable feedback, so the analysis considers the percent viewed.

To determine how students use feedback in their revisions (RQ2), the research team examined students' keystroke stream data, which enabled the research team to replay the students' interactions with the prototype. An inventory of revision types was compiled, and each revision type was counted once per student; revision types were not counted multiple times per student because there was not always a clear way to delineate one revision episode from another. In addition, the algorithm described in Beigman Klebanov et al. (2024), which classifies token-level changes between texts based on Levenshtein distance, was adapted to code individual tokens as being kept, added, replaced, or deleted between drafts. Individual codes were counted per student to understand the extent and nature of revisions.

Regarding students' perception of the tool (RQ3), descriptive statistics and frequency distributions of students' survey responses about the understandability and usefulness of feedback were computed and considered together with responses to open-ended questions. Ratings of chat helpfulness reported by students while completing the activity were counted by student.

Results

RQ1. What Types of Feedback do Students Seek in the Tool?

All students sent at least one message to the chat agent regarding their writing, on average sending four messages ($SD=3.09$). Table 3 presents the types of messages students sent to the chat agent. The majority of messages (68%) were related to the dimensions covered in the Opinion task rubric, indicating that students were asking further probing questions about the initial rubric-based feedback. Most students ($n = 10$) asked a few questions about organization, typically asking for suggestions on what linking words to use or how to improve the introduction or conclusion (e.g., "How's my thesis?", "conclusion help!!"). One student asked if it was okay to write their essay using bullet points. Most students ($n = 8$) sought feedback on error correction as well as idea development. Of the students who asked about error correction, six students used the example question provided under the language dimension in the rubric-based feedback (i.e., "What mistakes did I make?"). One student asked specifically about grammar mistakes, and another asked why a specific sentence was incorrect. Of the nine messages about idea development, all but one asked the chat agent to provide more details. Some students specified which main ideas they wanted to add details about (e.g., "what details I can add about time-management?") and one student requested feedback on the quality of their ideas (i.e., "Are my reasons persuasive?"). Other students ($n = 2$) requested vocabulary suggestions, and one student asked for assistance with clarifying the meaning

in part of their essay. In addition, three students requested holistic feedback, and another asked for advice on how to write faster.

Messages sent by two different students reaffirm the need for guardrails built into chat agents designed for providing feedback to young learners. One student asked, in Japanese, how to sound more “ネイティブ [native]”. In response, the chat agent gave advice directed at improving cohesion without opining on the connotations implicit in the message. However, it may be prudent to add instructions to the prompt that, for example, acknowledge the many varieties of English or redirect the focus of the conversation towards intelligibility. The other student (KS10) asked the chat agent to write their essay for them and entered a series of nonsense messages (e.g., just the letter “a”), which were all met with general advice on starting or adding to the essay.

Table 3
Chat Messages Categorized by Intent

| Intent of Message | Number of Students | Number of Messages | Selected Examples with Student IDs |
|---|--------------------|--------------------|--|
| Help with organization | 10 | 18 | “僕の文章にお勧めなリンクワードは何ですか。[What linking words would you recommend for my writing?]” – JS03 |
| Give error correction | 8 | 9 | “what grammer i had mistake the most?” –KS05 |
| Help with idea development | 8 | 9 | “Where is the part where I should explain more about?” – JS04 |
| Give holistic feedback | 3 | 4 | “How was my essay” – KS09 |
| Translate something | 2 | 6 | “能力値 [ability level] in English” – JS03 |
| Suggest new vocabulary | 2 | 2 | “違う単語を使ったほうが良い部分がありますか。[Are there any parts where I could use different words]” – JS01 |
| Help clarify meaning | 1 | 1 | “how can i give clear opinion?” – KS07 |
| Help with writing like a “native” speaker | 1 | 1 | “ネイティブな文章にするにはどのようにすればよいですか。[How can I make my writing sound native?]” – JS01 |
| Self-clarification | 1 | 1 | “I mean introduction” – KS07 |
| Give advice on writing strategies | 1 | 1 | “How can I write essay faster?” – KS07 |
| Unintelligible | 1 | 4 | “a” – KS10 |
| Write the whole essay | 1 | 1 | “Give me an example of an essay” – KS10 |
| Total | 15 | 57 | |

Note. English translations are provided in brackets after the original text.

A closer examination of the process data (keystroke logs and engagement time) revealed a wide range of variation in students' engagement with the tool. As summarized in Table 4, there were 5 of the 14 students who viewed the highlightable feedback providing elaboration on rubric-based feedback. Only two students, both in the Japanese sample, used the feature that translate the rubric-based feedback from English. Three students also used their L1 to compose messages in the chat.

Table 4*Engagement with Feedback Types Summarized by Student*

| ID | Highlights (magnifying icon) Selected | Translated Rubric Feedback | Languages Used in Chat | Chat Messages | Engagement Time (Minutes) |
|------|---|----------------------------------|---------------------------|---------------|------------------------------|
| JS01 | 2 (50%) | Yes | JP, EN | 7 | 25.76 |
| JS02 | 2 (50%) | No | EN | 2 | 17.15 |
| JS03 | 0 (0%) | Yes | JP, EN | 7 | 25.91 |
| JS04 | 0 (0%) | No | EN | 2 | 19.63 |
| KS01 | 0 (0%) | No | EN | 2 | 4.58 |
| KS02 | 0 (0%) | No | EN | 3 | 1.84 |
| KS03 | 0 (0%) | No | EN | 2 | 3.08 |
| KS04 | 3 (100%) | No | EN | 1 | 3.80 |
| KS05 | 1 (25%) | No | EN | 8 | 24.52 |
| KS06 | 0 (0%) | No | EN | 1 | 1.08 |
| KS07 | 3 (100%) | No | EN | 7 | 12.41 |
| KS08 | 0 (0%) | No | KO | 1 | 1.33 |
| KS09 | 0 (0%) | No | EN | 3 | 2.17 |
| KS10 | 0 (0%) | No | EN | 10 | 22.16 |

Note. EN = English, JP = Japanese, KO = Korean

RQ2. How do Students Use Feedback in their Revisions?

Tables 5 and 6 summarize the revisions in tokens between the first and final drafts of each writing piece, accounting for changes in words, capitalization, and punctuation. Although there were five students who made no revisions, students added about 20 tokens, on average. Such revisions mainly involved adding elaboration to existing sentences, adding new sentences, adding entire paragraphs, and adding cohesive devices, such as linking words, though some tokens were also added through the course of correcting linguistic errors. No students deleted ideas while revising; instead, the few deletions that occurred were the result of correcting syntactic errors. Replacements were identified in half of the students' texts and typically indicate changes in word choice, morphology, spelling, or capitalization.

Table 5*Summary of Student Revisions*

| ID | First Draft | Final Draft | Kept | Deleted | Added | Replaced |
|------|-------------|-------------|------|---------|-------|----------|
| JS01 | 36 | 43 | 34 | 2 | 11 | 7 |
| JS02 | 76 | 79 | 76 | 0 | 3 | 1 |
| JS03 | 35 | 49 | 35 | 0 | 14 | 4 |
| JS04 | 37 | 63 | 37 | 0 | 26 | 1 |
| KS01 | 120 | 179 | 120 | 0 | 59 | 0 |
| KS02 | 153 | 153 | 153 | 0 | 0 | 0 |
| KS03 | 216 | 216 | 216 | 0 | 0 | 0 |
| KS04 | 113 | 138 | 113 | 0 | 25 | 8 |
| KS05 | 102 | 207 | 102 | 0 | 105 | 6 |
| KS06 | 296 | 296 | 296 | 0 | 0 | 0 |
| KS07 | 167 | 188 | 166 | 1 | 23 | 3 |
| KS08 | 155 | 155 | 155 | 0 | 0 | 0 |
| KS09 | 94 | 94 | 94 | 0 | 0 | 0 |
| KS10 | 0 | 157 | 0 | 0 | 157 | 0 |

Table 6*Descriptive Statistics of Word Counts in Student Revisions*

| | First Draft | Final Draft | Kept | Deleted | Inserted | Replaced |
|-------------|-------------|-------------|--------|---------|----------|----------|
| <i>Mean</i> | 123.08 | 143.08 | 122.85 | 0.23 | 20.46 | 2.31 |
| <i>SD</i> | 75.69 | 75.36 | 75.84 | 0.60 | 30.58 | 2.98 |
| <i>Min</i> | 35 | 43 | 34 | 0 | 0 | 0 |
| <i>Max</i> | 296 | 296 | 296 | 2 | 105 | 8 |

Table 7 lists the different ways students engaged with feedback and quantifies how many students exhibited each behavior. The most common tendency, observed in six students, was the addition of ideas suggested by the chat agent. For instance, when a student (JS04) asked, "Where is the part where I should explain more about?", the chat agent responded with the following:

You could explain more about your reasons. For example:

- Why can't you do club activities after school if you have homework?
- How does homework stop you from reading books?

The student then elaborated on reasons in two sentences by adding causal adverbial clauses as in the following excerpt (the additions are noted in italics):

First, I cannot do club activities *because it takes time to finish homework*.

Second, I would not have time to read books *because I would be tired of homework*.

Table 7*Types of Revisions Made by Students*

| Dimension | Revision Type Category | Yes | No |
|--------------|--------------------------------------|-----|----|
| Content | Clarified meaning using highlighting | 1 | 13 |
| | Clarified meaning using chat | 1 | 13 |
| | Paraphrased information using chat | 3 | 11 |
| | Copied information from chat | 5 | 9 |
| | Added ideas elicited by chat | 6 | 8 |
| | Added unelicited ideas | 2 | 12 |
| Organization | Improved coherence using chat | 5 | 9 |
| Language Use | Corrected errors using chat | 5 | 9 |

The chat agent occasionally provided short examples, which some students attempted to copy and paste. In those instances, the tool displayed a message asking students not to copy and paste, but to type the examples instead. In total, four students attempted to copy and paste text generated by the chat agent. Among those students, one (KS05) chose not to include the copied text in their revision, one (KS10) paraphrased instead, and two (JS03 and KS07) typed a clause provided by the chat agent verbatim. Two other students typed text verbatim from the chat, while three others paraphrased examples provided by the chat agent without being prompted to do so after attempting to copy and paste.

Of the five students who utilized feedback to improve coherence, four were provided with linking words by the chat agent that they subsequently added to their texts. One student (JS01) used a suggestion provided by the chat agent to improve the lexical cohesion of their introduction by making an anaphoric reference (revision shown in italics): “I disagree with the statement. I have two reasons *why*.”

The five students who received corrective feedback used all of it. For instance, one student (KS07) asked, “What mistakes did I make?” and was presented with the following response by the chat agent:

Here are a few corrections:

- By solving other various questions, we can enhance ~~out~~ **our** problem solving skills
- In the second place, ~~to through~~ **through** one's homework
- It will help ~~to became~~ **to become** independent learners
- Like enhancing the knowledge and ~~improve~~ **improving** self-directed learning skill.

The student then corrected all errors. The other students who used corrective feedback to guide their revisions edited their texts in a similar manner.

RQ3. What are Students' Perceptions of the Feedback Tool?

In the survey, students generally reported that they liked the writing activity. All but three students reported that they “liked” or “really liked” the writing activity. On a scale from 0 to 4, with 0 being the least favorable rating, the mean rating for how much students liked the writing activity was 3.14. On the same scale, students reported a similar disposition towards typing on a computer, with a mean rating of 3.21. The mean response to “How easy or difficult was it to complete the writing activity” was 2.57, which is between “not easy or difficult” and “easy” on the same five-point scale. A student (KS01) who rated the writing activity as “very easy” commented that it was easy to provide reasons and examples because the writing prompts were about everyday topics. Another student (JS02), who rated the activity as “difficult”, commented that the reason for the difficulty was because they did “not have much vocabulary.”

Table 8 presents descriptive statistics and frequency distributions for the understandability of feedback. The mean rating of overall understandability was 3.43, indicating that most participants found the feedback either mostly understandable ($n = 8$) or completely understandable ($n = 6$). Specifically, rubric-based feedback on content had a mean score of 3.50, with the majority of participants ($n = 8$) reporting complete understanding, while feedback on organization had a mean score of 3.29, with the highest number of students ($n = 8$) indicating that they mostly understood the feedback. Feedback on language use had a slightly lower mean of 3.21, with a more varied distribution relative to other rubric-based feedback. Four students indicated they understood only “a little,” while seven students reported complete understanding. Although half of the students reported that they completely understood the highlightable feedback (i.e., feedback the magnifying glasses show), one reported that they did not understand it at all. Feedback from the chat agent was reported as the most understandable overall, with a mean of 3.79.

Table 8*Descriptive Statistics and Frequencies on the Understandability of Feedback (N = 14)*

| | <i>Mean</i> | <i>SD</i> | Did not understand at all | Had difficulty understanding | Understood a little | Mostly understood | Completely understood |
|--------------------------------------|-------------|-----------|---------------------------|------------------------------|---------------------|-------------------|-----------------------|
| Overall | 3.43 | 0.51 | 0 | 0 | 0 | 8 | 6 |
| Feedback on content | 3.50 | 0.65 | 0 | 0 | 1 | 5 | 8 |
| Feedback on organization | 3.29 | 0.61 | 0 | 0 | 1 | 8 | 5 |
| Feedback on language use | 3.21 | 0.89 | 0 | 0 | 4 | 3 | 7 |
| Feedback the magnifying glasses show | 3.14 | 1.17 | 1 | 0 | 2 | 4 | 7 |
| Feedback from the chat | 3.79 | 0.43 | 0 | 0 | 0 | 3 | 11 |

Descriptive statistics and frequency distributions for the usefulness of feedback are included in Table 9. On average, students found the feedback from the chat agent most useful, followed by feedback on content and language use. One student (KS09) commented, “It was very useful using the chat bot as it was good at finding my faults.” Another (KS01) added, “It was a great chance while using chatting AI. Due to the fact that those chatters helped me to understand better in my first draft. Hence, chatting with AI was an opportunity to learn and broaden my vocabulary and phrases skills.”

Table 9

Descriptive Statistics and Frequencies on the Usefulness of Feedback (N=14)

| | <i>Mean</i> | <i>SD</i> | Not useful at all | Not that useful | Somewhat useful | Useful | Very useful |
|---|-------------|-----------|----------------------|--------------------|--------------------|--------|----------------|
| Feedback on content | 3.50 | 0.65 | 0 | 0 | 1 | 5 | 8 |
| Feedback on organization | 3.29 | 0.61 | 0 | 0 | 1 | 8 | 5 |
| Feedback on language use | 3.36 | 0.50 | 0 | 0 | 0 | 9 | 5 |
| Feedback the magnifying glasses show | 2.71 | 1.27 | 1 | 1 | 4 | 3 | 5 |
| Feedback from the chat | 3.64 | 0.50 | 0 | 0 | 0 | 5 | 9 |

Table 10 summarizes student ratings of how helpful chat messages were. Most students rated all of their interactions with the chat agent as helpful. Some variations were also observed depending on the chat responses. Two students, who both engaged more frequently with the chat agent, rated one message from the chat agent as unhelpful and the rest as helpful. Only one student (KS10) rated all the responses as unhelpful.

A closer look at the chat interactions of this student (KS10) reveal that they did not engage with the chat agent as intended and may have experienced technical difficulties. Two of their messages included only the letter “a”, and another message asked the chat agent to write their essay for them. In regard to the two other messages rated as unhelpful, the one student’s (JS01) message was written in Japanese, requesting help with writing a conclusion. The chat agent’s response gave advice in Japanese on how to write conclusions and included an example conclusion that was above the writing level of the student. Other responses from the chat agent, which the student rated favorably, provided corrective feedback or support for idea development (in Japanese). The other student who rated a message as unhelpful (KS07) asked for advice on how to lengthen their introduction but received advice for how to write a conclusion. The student then clarified by replying, “I mean introduction”, and rated the next message as helpful.

Table 10*Student Ratings of Chat Message Helpfulness*

| ID | Messages | Ratings | Helpful | | Not Helpful | |
|------|----------|---------|---------|------|-------------|-----|
| | N | N | N | % | N | % |
| JS01 | 7 | 7 | 6 | 86% | 1 | 14% |
| JS02 | 2 | 2 | 2 | 100% | 0 | 0% |
| JS03 | 7 | 3 | 3 | 43% | 0 | 0% |
| JS04 | 2 | 1 | 1 | 50% | 0 | 0% |
| KS01 | 2 | 2 | 2 | 100% | 0 | 0% |
| KS02 | 3 | 3 | 3 | 100% | 0 | 0% |
| KS03 | 2 | 2 | 2 | 100% | 0 | 0% |
| KS04 | 1 | 1 | 1 | 100% | 0 | 0% |
| KS05 | 8 | 1 | 1 | 13% | 0 | 0% |
| KS06 | 1 | 1 | 1 | 100% | 0 | 0% |
| KS07 | 7 | 7 | 6 | 86% | 1 | 14% |
| KS08 | 1 | 1 | 1 | 100% | 0 | 0% |
| KS09 | 3 | 3 | 3 | 100% | 0 | 0% |
| KS10 | 10 | 8 | 0 | 0% | 8 | 80% |

Discussion

In this article, we presented a small-scale prototyping study conducted during the development of an AI-assisted feedback tool to support the writing skill development of young EFL students. Specifically, this tool was intended to serve as a useful resource for potential users of the TOEFL Junior Writing test. It is pivotal to engage target users early in the development of a technological tool, particularly when it involves young learners whose experiences with technology can vary widely. Before further developing our tool, we examined how young EFL students interacted with the prototype, focusing on their use of the interactive chat functions to receive and utilize AI-generated feedback.

Regarding the types of feedback students sought through the chat, the most frequent inquiries were related to organization, followed by corrective feedback and content. One possible explanation for this trend is that students were influenced by the initial rubric-based feedback, particularly in areas unfamiliar to them. Recall that the rubric-based feedback was organized by the dimension of content, organization, and language use. It seems that students were not used to receiving feedback on organization compared to the areas of content and language forms. Hence, they asked more follow-up questions about the organization dimension.

While it is not surprising that students inquired about their language use (e.g., mistakes, vocabulary) given their English proficiency levels, it is noteworthy that many chat questions concerned idea development. This capability, offering suggestions for ideas and details for content development, is a distinct advantage of generative AI technology using large language models like GPT-4. Yet, it is also noteworthy that there was a wide

range of variation in the use of the chat function, with students sending between one and ten chat messages each. This result suggests that students would need guidance and practice in prompting the chat to take advantage of this interactive feedback feature.

We found that the students who engaged most with the feedback features tended to revise their writing accordingly. Eight students (62%) were actively involved in acting upon the feedback provided by the tool. It is also interesting to note that students assessed the relevance of the feedback before deciding whether to integrate it. As reported in the results section, feedback that students did not integrate was sometimes seen as irrelevant, partly due to incomplete or inadequate prompting in the chat. This finding suggests that young L2 learners have the capacity to evaluate the relevance of feedback. Moreover, students' metacognitive skills in formulating chat questions to receive feedback and evaluating AI-generated feedback are critical in utilizing AI-assisted writing feedback tools. This finding corroborates with previous studies examining the effects of ChatGPT-generated feedback on adult L2 learners' writing and their critical thinking skills (e.g., Lu et al., 2024; Teng, 2024).

Despite such variations in using the interactive feedback feature among students, students' ratings for the feedback from the chat were high. This result is not surprising, given that the chat feedback was provided in response to students' specific questions posed in the chat. Students appeared to value personalized feedback in this study. In previous studies on young L2 learner's writing feedback use, researchers have reported a general preference among students for receiving more feedback, either from teachers or from AI-assisted tools, on their writing (e.g., Lee, 2008; Wang et al., 2020; Wolf & Oh, 2024). This preference aligns with our findings, indicating that both personalized feedback and more frequent interactions with feedback tools enhance engagement for young L2 learners.

This study was limited by its small convenience sample. The data collection period was also restricted, allowing for only a single writing task by students. With more comprehensive practice and greater familiarity with the tool, the study might have resulted in different usage patterns of students' interactive feedback features. However, our study was intended to implement an agile approach to integrate target users' feedback early in the tool development process. Despite these limitations, the study findings provided useful insights to guide the further development and refinement of the tool. This study also yielded empirical evidence supporting the potential of an AI-assisted tool for the development of young EFL students' writing skills. Features such as providing personalized feedback and enhancing students' metacognitive and self-regulation would be particularly beneficial. Additionally, by using a feedback tool linked to the writing assessment demonstrated in this study, both students and teachers can easily evaluate the tool's effectiveness through assessment results. The effectiveness of such AI-assisted

writing tools and systematic observations of students' use of the tools are certainly areas for further research.

In this article, we have also detailed the technical aspects of the prototype development, showcasing the involvement of cross-disciplinary expertise in creating an effective tool to assist L2 learners' writing skills. As Chapelle (2024) emphasizes, the role of language researchers is more critical than ever in ensuring the adequate use of generative AI technology in L2 education through interdisciplinary collaborative work. We hope that our study serves as a useful example for future endeavors in this field.

ORCID

 <https://orcid.org/0000-0002-0622-7754>

 <https://orcid.org/0000-0002-4734-3794>

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

The authors would like to acknowledge the valuable assistance of their colleague, Jean Alderman, in the development of the tool and the collection of data. They also wish to express their sincere appreciation to the teachers and students who participated in this study.

Funding

Not applicable.

CRedit Authorship Contribution Statement

Mikyung Kim Wolf: Conceptualization, Methodology, Investigation, Writing-Original Draft, Writing-Review & Editing, Supervision

Michael Suhan: Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, Project administration

Generative AI Use Disclosure Statement

No generative AI tools were used in the writing of this article.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Participants

Not applicable.

Competing Interests

There are no competing interests.

Data Availability

The data are available from the authors upon reasonable request.

References

- Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies. *RELC Journal*. <https://doi.org/10.1177/00336882241262533>
- Banihashem, S.K., Kerman, N.T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21, Article 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Beigman Klebanov, B., Suhan, M., O'Reilly, T., & Wang, Z. (2024). From miscue to evidence of difficulty: Analysis of automatically detected miscues in oral reading for feedback potential. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*. <https://aclanthology.org/2024.bea-1.38>
- Castro-Alonso, J., Wong, R.M., Adesope, O.O., & Paas, F. (2021) Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review*, 33, 989-1015. <https://doi.org/10.1007/s10648-020-09587-1>
- Chapelle, C. A. (2024). Open generative AI changes a lot, but not everything. *The Modern Language Journal*, 108, 534–540. <https://doi.org/10.1111/modl.12927>
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge University Press.
- Chapelle, C. A. (1999). Research questions for a CALL research agenda: A replay to Rafael Salaberry. *Language Learning & Technology*, 3(1), 96-104. <http://llt.msu.edu/vol3num1/reply/index.html>
- Chapelle, C. A., Beckett, G. H., & Ranalli, J. (2024). Paths for exploring AI in applied linguistics. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 262-274). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.15>
- Clark, R C., & Mayer, R.E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (5th ed.). John Wiley & Sons.
- ETS. (2019). TOEFL Junior® framework and test development. *TOEFL® Research Insight: Vol. 7*. <https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v7.pdf>
- Fiorella, L., & Mayer, R.E. (2022). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 277-285). Cambridge University Press.
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
- 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, Article 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17, 144-164. <https://doi.org/10.1016/j.jslw.2007.12.001>
- 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>
- Microsoft. (n.d.). *Prompt engineering: Chat* [Azure AI services]. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/prompt-engineering?tabs=chat>
- Moreno, R., & Mayer, R.E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, 96, 165-173. <https://doi.org/10.1037/0022-0663.96.1.165>
- Noetel, M., Griffith, S., Delaney, O., Harris, N R., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2022). Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Review of Educational Research*, 92(3), 413-454. <https://doi.org/10.3102/00346543211052329>

- Schroeder, N. L., & Adesope, O.O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3), 229-251. <https://doi.org/10.1080/15391523.2014.888265>
- Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, 44, Article 100449. <https://doi.org/10.1016/j.asw.2020.100449>
- Wolf, M. K., & Oh, S. (2024). Investigating the types and use of feedback in middle-school English language learners' academic writing. *Language Learning & Technology*, 28(1), 1-25. <https://hdl.handle.net/10125/73583>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). *Exploring the provision of AI-based feedback for TOEFL Junior Writing® practice tests* (Research Memorandum No. RM-24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>

Appendix A

Descriptors of the Highest and Lowest Score Points on the TOEFL Junior Writing Rubric for the Opinion Task

| Score | Development and Language Use Descriptors |
|-------|---|
| 4 | <p>A typical response at this level is characterized by the following:</p> <ul style="list-style-type: none"> • states a position on the topic • provides support for the position, with specific details and/or examples • is mostly well organized and coherent • shows lexical variation appropriate for the task • displays a varied sentence structure appropriate for the task • may contain minor errors but they do not interfere with meaning or clarity |
| 0 | <p>Only copies words from the prompt, rejects the prompt, is completely off topic, consists of keystroke characters, is written in a foreign language, or is blank.</p> |

Source. <https://www.ets.org/pdfs/toefl/toefl-junior-writing-scoring-guide.pdf>; Copyright © 2022 by ETS.

Appendix B

Screenshot of Feedback Interface

Question
Do you agree or disagree with the following statement? Computer games help students develop useful skills.

Your Writing Finish Writing

I disagree with the statement. I have two reasons why. First, we lose studying time. Second, many people don't apply the skills they learn from games in their future careers. Therefore, I think that Computer games don't help students develop useful skills.

Feedback:
ネイティブな文章にするにはどのようにすればよいですか。

ナチュラルな英語にするために、ここで少し言葉を調整しましょう。

- "I have two reasons."を"**I have two reasons why.**"に変える。
- "Second, many people don't use skills learned from games in the future."を"**Second, many people don't apply the skills they learn from games in their future careers.**"に変える。
- "so I think that Computer games don't help students develop useful skills."を"**Therefore, I think that computer games don't help students develop useful skills.**"に変える。

文章の流れも思い出させて、読み手に理由と結論をはっきりと理解させることが大切です。😊

Helpful Not helpful

Chat with me. Send

Notification:
You can edit your writing now. Let's work on a second draft together!
Select the magnifying glasses 🔍 to see more feedback.