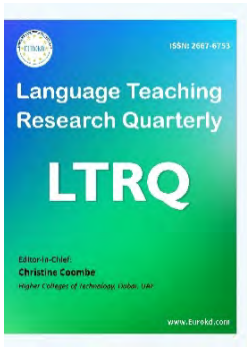

Language Teaching Research Quarterly

2024, Vol. 43, 1–21

Thematic Collection[1]

Exploring the Educational Potential of ChatGPT: AI-Assisted Narrative Writing for EFL College Students

Ji-Young Seo

Kookmin University, Republic of Korea

Received *04 May 2024* ***Accepted*** *01 August 2024*

Abstract

Although ChatGPT provides excellent features as a writing assistant tool, few empirical studies have been conducted on its integration into writing education investigating students' usage pattern and impact on the writing skills of students. This exploratory research aims to fill this research gap by analyzing prompts initiated by students and examining the effects of ChatGPT-assisted narrative writing to explore the educational potential of ChatGPT in college-level L2 writing. Toward this end, the study recruited 44 university students in South Korea. The study explored their patterns of use of ChatGPT and the effects of the narrative writing intervention assisted by ChatGPT. The major findings were as follows. The top three request prompts were: Request for language use, Request for revision, Request for information. The most frequent requests were related to linguistic aspects, and ChatGPT demonstrated successful surface-level error detection. Next, the results of the paired sample *t*-test and Wilcoxon signed rank test to determine the effect of ChatGPT-assisted narrative writing demonstrated high post-test scores in writing fluency and overall performance, and this difference was statistically significant. On the other hand, the post-test scores for syntactic complexity were lower than those for the pre-test, and this difference was also significant. Regarding clause complexity, clausal complements per clause exhibited a significant increase in the post-test. Based on these findings, the pedagogical implications are suggested.

Keywords: *Artificial Intelligence, Generative AI, ChatGPT, Chatbot Integrated Language Learning, Narrative Writing, Syntactic Complexity*

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

Seo, J. Y. (2024). The intersection of AI and language assessment: A study on the reliability of ChatGPT in grading IELTS writing task 2. *Language Teaching Research Quarterly*, *43*, 1-21. https://doi.org/10.32038/ltrq.2024.43.01

[1] This paper is part of a thematic collection (2024, 43) entitled: Artificial Intelligence and ChatGPT for Language Education: A Research Agenda.

* Corresponding author.
E-mail address: sjy@kookmin.ac.kr
https://doi.org/10.32038/ltrq.2024.43.01

Introduction

Since its launch in November 2022, ChatGPT, a large language model (LLM) developed by OpenAI, has had a marked impact on English writing education. In contrast to previous writing assistance tools (e.g., Grammarly), ChatGPT generates human-like natural texts within seconds in response to prompts. For individuals who spend a substantial amount of time writing and editing by considering rhetoric convention, style, and register, it can be considered innovative, because it significantly reduces the time required. Teachers also benefit in many ways. In April 2023, Intelligent (2023) conducted a survey on 1,000 high school teachers and university professors who said they knew about ChatGPT. The results indicated that more than 98% of teachers said that ChatGPT helped them organize lessons, give feedback, and write emails. Alternatively, 93% responded that they were saving time by using it for time-consuming tasks such as grading assignments or providing feedback for the written work of students. However, only 79% of teachers responded that they approved of the use of ChatGPT among students. The reason was that students may extremely rely too much on ChatGPT instead of critically thinking and independently writing on their own. They also expressed concerns about students acquiring inaccurate information.

From the perspective of students, they want to leverage ChatGPT to enhance the speed and quality of their work (Hart-Davis, 2023). Korean college students also welcome the advent of ChatGPT. By entering prompts, such as the requirements and evaluation criteria of an assignment, they can quickly achieve excellent outcomes without writing. It also benefits students from underprivileged backgrounds with limited access to native speakers of English. Even before ChatGPT, many Korean university students had been using AI-based grammar checkers despite the many errors and awkward expressions in the results. In an academic context, writing is primarily intended to demonstrate a student's knowledge and they are evaluated based on their written works. Many Korean students believe that AI-generated outputs surpassed their writing. Part of the reason can be that many students lacked experience in English writing before entering university. University entrance examinations in Korea include only English listening and reading, such that they do not invest much time in acquiring productive skills, particularly in English writing. Briggs (2018) also points out that within the context of competition-oriented education in Korea, college students are forced to use a translator to compensate for their lack of skills and to receive higher grades in the evaluation, although using a translator is not desirable for learning. Without institutional restriction, Korean students are predicted to continue to use ChatGPT for English writing. In addition, given the reality in which generative AI is ubiquitous and its impact is increasing, outright prohibiting students from using ChatGPT will be unwise (Sharples, 2022). The role of ChatGPT in writing tasks is a complete package from idea generation to final editing. Evidently, students will readily use it in their writing assignments. However, recognizing that obtaining instant AI-generated writing output without effort and excessively relying on ChatGPT can hinder the development of writing skills is essential. Thus, equipping students with the skills required to effectively utilize tools is crucial.

There is an ongoing debate about using ChatGPT in writing education, but scholars widely acknowledge that educators do need to explore innovative ways to integrate AI into writing education and develop new evaluation methods for writing (Baskara, 2023). Writing is not a skill acquired automatically and cannot be developed naturally. The process-oriented writing

instruction is time-consuming, and teachers may feel the burden of providing personalized feedback at each step. Particularly, in large class sizes (50 or more students) in Korea, offering real-time instant feedback is challenging. ChatGPT has potential to provide solutions to such issues. As such, educators should consider the integration of AI support into writing classes to guide learners in the effective use of AI during the writing process.

ChatGPT is still in the early stages of development and as it was not initially made for educational purposes, there are currently no established guidelines for its use in education. Therefore, the existing literature mainly focuses on suggestions and recommendations for teaching methods along with insights into user reactions and perception. To explore the educational potential of ChatGPT in second language (L2) writing, it is essential to apply it to actual learners and closely examine any changes in writing performance. To this end, closely examining the use of ChatGPT in the writing process of students and providing insight through analysis of writing results before and after ChatGPT intervention are necessary steps. Therefore, this analyzes the patterns of college students' ChatGPT use and its effect on writing with the assistance of ChatGPT. In terms of writing genre, the study selected narrative writing, which involves personal experience, and it is suitable regardless of the learner's level, even Korean students with little experience in English writing. Kormos (2011) mentioned that narrative writing is frequently taught in general foreign language courses starting from the beginning level. Knapp and Watkins (2005) also stated that narrative writing is suitable for addressing EFL students' lack of experience in English writing. They suggested that one of the effective ways to develop narrative skills is by having students retell stories they have read in class. Moreover, ChatGPT continues to face a critical issue of generating false information as if it were a fact, which is called hallucination. Lingard (2023) argued that assessing the accuracy of the responses of ChatGPT requires domain-specific knowledge. Therefore, instead of students engaging in free writing on various topics, the study adopts a story retelling format with the objective of minimizing the risk of hallucination. Inspired by this notion, the present study engaged students in narrative writing by watching movie scenes together during class and then retelling the events from the perspective of the characters.

A better understanding of AI's potential applications is necessary to empowers instructors to offer valuable guidance and facilitates research on improving AI utilization for writing based on user experiences. Scholars widely acknowledge that ChatGPT contains excellent features as a writing assistant tool, but empirical studies that investigate its integration into writing education, the use pattern of students, and its impact on writing skills are few (Han et al., 2023). The current study intends to fill this research gap by exploring prompts initiated by students to provide a guide of AI-assisted writing education. Therefore, the major contribution is the insight gained into the interaction of students with ChatGPT and its impact on their narrative writing. The results may be applicable to other L2 writing education. The research questions are as follows:

RQ_1: What are the patterns of college students' use of ChatGPT?

RQ_2: What is the effect of ChatGPT assistance on the narrative writing of EFL learners?

Literature Review

English Writing Education with ChatGPT

In a recent study using generative AI, Hwang (2023) highlighted the potential benefits of ChatGPT for Korean students' English writing skills. He asserted that ChatGPT can provide much more abundant linguistic feedback in real time than conventional machine translation tools in which students can improve their writing abilities with the help of feedback. He also noted that ChatGPT can function as a supplementary tool for personalized learning. ChatGPT can provide customized materials, because it can easily convert text based on the level of difficulty and genre, which fosters genre-based instruction. He also described the role of teachers, that is, guide students in taking advantage of the latest AI-based programs in a meaningful and productive manner.

Su et al. (2023) noted that ChatGPT differs from conventional chatbots that only provide preset responses. Instead, it generates human-like responses. It can also sustain conversations and utilize prior exchanges for subsequent interaction (OpenAI, 2022). They highlighted that ChatGPT can process longer input and provide personalized feedback on writing samples; thus, it can serve as a virtual learning partner. They proposed strategies for collaborating with ChatGPT across the stages of the argumentative writing process. However, they cautioned that ChatGPT may produce inaccurate information and inconsistent feedback. The outcomes can vary according to the prompts used, which could confuse content feedback. Despite this aspect, its function as a proofreading aid is outstanding. It provides valuable assistance in grammatical accuracy and meaning elaboration, suggests alternatives, refines lexical choice and syntactic structure, and improves academic tone.

With the help of the integration of ChatGPT into writing classes, teachers are encouraged to assign higher-order tasks that emphasize creative ideas, critical thinking, and logical reasoning (Stokel-Walker, 2022). Rudolph et el. (2023) addressed the expected major changes in the procedures and assessments of traditional higher education (e.g., essays). They also recommended including AI-powered writing assistants, such as Grammarly, Wordtune, and Elicit, as part of the curriculum.

In the short term, teachers should incorporate methods in writing tasks that leverage areas where ChatGPT has limitations for now. Mills (2023) and Nowik (2022) proposed the design of writing assignments that utilize areas outside the scope of ChatGPT such as analyzing images and videos and incorporating students' personal experiences or viewpoints, which are difficult for AI systems to emulate. Nowik (2022) recommended evaluating students based on their skills to integrate multiple sources and present original arguments. Jairoun et al. (2024) presented several methods to detect AI-generated texts, including the use of plagiarism detection tools, requesting raw data, and closely examining references and citations. However, the areas in which ChatGPT currently has limitations are being updated rapidly; thus, teachers are urged to leverage ChatGPT to innovate teaching strategies and to alter the writing evaluation method from a long-term perspective.

Narrative Writing

Abbott (2014) described the narrative genre as the art of storytelling; narratives can be observed in every moment of life, and humans consistently engage in narration. Narratives can be retelling of actual events (recounts) or imagined events told in chronological order. Scholars

have asserted that narrative writing is a fundamental and pedagogical genre for the teaching of ESL/EFL writing (Knapp & Watkins, 2005; Kormos, 2011).

Different genres exhibit various structural, lexical, and syntactic features. Holloway and Freshwater (2007) stated that narrative writing necessitates critical thinking and logical reasoning, because writers must carefully consider elements such as plot, pacing, and character development. Ellis and Yuan (2004) confirmed that narrative writing requires language proficiency and the use of descriptive and engaging language. Moreover, Norrick (2016) indicated that narratives frequently employ indirect reports and quotations. In writing a narrative, the narrator must closely observe situations and understand the psychology and relationships of characters to be able to convey them. In this process, direct and indirect reports are employed as a means for exhibiting the stance of the narrator. Syntactically, narrative writing tends to contain less complex sentence structures than do other genres. In this regard, Lu (2011) conducted a study to compare the syntactic complexity of argumentative and narrative writing among college students and found that argumentative texts exhibited higher syntactic complexity. Zhang and Liu (2021) conducted a similar research on Chinese learners and confirmed that genre type significantly impacts L2 syntactic complexity. They revealed that although a positive correlation exists between syntactic complexity and overall writing scores in argumentative writing, no such correlation was found in narrative writing.

Research on the development of assessment tools, such as rubrics for narrative writing, has also been conducted due to diverse requirements and characteristics across genres. For example, Pourdana and Asghari (2021) demonstrated that the Smarter Balanced Narrative Writing Rubric (Smarter Balanced Assessment Consortium, 2012) has been widely employed to assess written narratives. This rubric includes five components: 1) narrative focus, 2) organization, 3) elaboration of narrative, 4) language and vocabulary, and 5) conventions.

Macken-Horarik and Sandiford (2016) pointed out that many educators tend to rely on ambiguous rubrics or mainly focus on linguistic aspects, which, according to them, hinder proper evaluation by failing to reflect salient language choices specific to a particular genre. The authors developed a framework capable of multidimensional analysis of the narrative writing of students. Their framework describes the three key levels (i.e., genre, phase, and sentence) of narrative writing in detail, categorized into five distinct levels. For instance, in terms of phase level, under the voicing criterion, scores could range from 0 to 4 based on the extent of a vivid description of the character's identity created through dialogue. Unlike argumentative writing where citation and quotation play roles in providing evidence, narrative writing emphasizes characters' speech and verbal projection through dialogue. By incorporating these elements, narratives become increasingly vivid, which makes them feel more realistic to the readers.

In recent research on narrative writing using ChatGPT, Zhou et al. (2023) compared the writing performance of ChatGPT and Chinese intermediate English learners on a narrative topic. After analysis using Coh-Metrix, the results revealed that ChatGPT performed better than human writers in areas such as narrativity, word concreteness, referential cohesion, and syntactic simplicity. However, it lagged far behind Chinese intermediate English learners' writing in terms of deep cohesion. Deep cohesion refers to the useof connecting words in texts, which are crucial for narrative writing. For example, ChatGPT's output often lacked causal

connectives (e.g., therefore, even though) but tended to use the coordinating conjunction "*and*" even when connecting two cause-and-effect sentences.

Related Work: Students' Usage of ChatGPT

The integration of ChatGPT into writing is yet to be fully explored (Su et al., 2023). Jung (2023) noted that despite the growing interest in ChatGPT, empirical studies both in Korea and internationally remain notably scarce. Indeed, although writing instructors generally acknowledge that generative AI has become more prevalent and have begun incorporating it into curricula, its effects remain unknown (Bedington et al., 2024). As ChatGPT itself is still in its early developmental stages, and no unified guide for its utilization in the educational context currently exists. Scholars such as Chan (2023) and Floridi and Cowls (2021) advocate for the necessity of a comprehensive AI education policy and try to establish guidance for the responsible use of AI in university teaching and learning. Universities around the world have also begun providing their own guidelines (e.g., Garber et al., 2023). In the field of L2 education, the current research primarily focuses on classroom methods for teachers and provides suggestions, limitations, and recommendations (Jeon & Lee, 2023). A number of studies with actual users investigate changes in their perception and responses through questionnaires (Albayati, 2024; Billingsley & Gardner, 2024). However, empirical studies that examine the integration of ChatGPT into writing education, the use patterns of students and its impact on the writing skills of students are few. To gain an in-depth understanding of how to effectively implement ChatGPT in classrooms, it is essential to comprehend students' actual usage patterns.

In a previous study on student usage patterns, Joshi et al. (2023) conducted a research on students majoring in Computer Science, which is a field familiar with ChatGPT. The study conducted a survey and interviewed 480 undergraduate students in India to examine the real-world use of ChatGPT. A total of 75.4% of students perceived ChatGPT as a learning and education tool. The most frequent cases of the use of ChatGPT included collecting information, summarizing learning content, providing assistance in coding and written work for the course (e.g., writing of emails and essays). The authors found a recurring usage pattern, that is, students obtained answers from ChatGPT, double-checked with Google, corrected errors using ChatGPT, and retrieved accurate responses.

Liu and Biebricher (2024) investigated the cognitive processes and behavioral patterns of Chinese undergraduate students during digital multimodal composing with generative AI. The study found that students actively used various prompts to obtain appropriate images and created bridge slides to ensure smooth text flow. The authors argued that these patterns positively affect writing skill development.

Sun and Deng (2024) investigated how students use ChatGPT to enhance experiential learning. Toward this end, they designed a ChatGPT activity for 39 college students in America. The findings revealed that the students tend use ChatGPT to understand conceptual knowledge (e.g., knowledge of theories), while instructors focused on the creation of metacognitive knowledge (e.g., the ability to understand and apply strategies for learning and problem solving). They also found that students' prompts are affected by the first prompt provided by their instructors. This result highlighted the significance of the instructor's role as a facilitator in enhancing AI-based experiential learning.

Han et al. (2023) recruited 213 college students in Korea and explored the interaction patterns between the students and ChatGPT based on the edit history of their argumentative essays. After analyzing essay edit patterns during interaction with ChatGPT, the study found that the top three request prompts were: Request for language use, Request for revision, Request for information. They also reported that learners were inclined to be receptive to feedback in terms of language use. Interestingly, a few students who did not embrace feedback from ChatGPT disregarded suggestions such as trivial feedback (e.g., modify "largely" into "broadly") and hallucination. They added that the students achieved notable improvement. In addition, they noted significant differences between the first and final scores of the draft essay specifically in terms of content, organization, language, sentence count, and perplexity.

Based on the previous research, it is clear that ChatGPT is opening up new possibilities for education, particularly as a writing assistant tool widely used by college students. Notably, a distinct trend exists in which students frequently request feedback on language use. However, few studies specifically examined which type of prompts they used and how students interact with AI during the writing process. There is a need for empirical studies on the actual integration of AI into writing classes. Thus, the current study aims to fill these research gaps by conducting an in-depth investigation on the use patterns of ChatGPT by students during writing classes as well as the impact of ChatGPT-assisted narrative writing. Analysis of students' prompts can provide a more detailed view of the challenges they encounter during writing, and the findings are applicable for future research on AI-assisted writing education.

Methodology

Participants

The study was conducted in South Korea from March to June 2023 during the early stages of ChatGPT's release. The free version of ChatGPT 3.5 was used. A priori power analysis was conducted using G*Power version 3.1.9.7 to determine the minimum sample size required for testing the hypothesis. Accordingly, out of 60 enrolled students, 44 students who agreed to participate in the research were chosen. After reviewing the study's objectives, procedures, and their own rights, all participants provided informed consent. Moreover, participants knew that their participation was voluntary and they could withdraw from the study at any time. The course was designed to encourage students to connect their own experiences to an American movie's themes, practice using the language in context, and develop their narrative writing skills. Throughout the course, students engaged by watching a movie to learn authentic English expressions and increase their cultural awareness. They discussed the film, characters, and cultural insights and wrote a narrative to reflect on their personal experiences that were similar to the themes presented in the film. The class was conducted once per week for 3 hours over 15 weeks, including an orientation session, a pre-test, and a post-test.

The students were composed of 23 (52%) were female and 21 (48%) were male. Their scores for the Test of English for International Communication (TOEIC) test ranged from 400 to 890 points with an average score of 630, which classifies them at the intermediate level (B1) according to the Common European Framework of Reference (CEFR) standards. Table 1 summarizes the participants' demographic data.

Material

The course materials were the American film, *I Don't Know How She Does It* (McGrath, 2011), and its screenplay (Yoon, 2013) from Screen English was used. The movie portrays the life of a working mom trying to balance between work and home responsibilities. It is based on a New York Times bestselling novel of the same title, such that students can explore the background. Furthermore, students can benefit from comparing the conversational language used in the film with the written language in the novel, which assists in learning various English expressions. Moreover, the film covers everyday events, such as milestones in life, holidays, and family conflicts, thereby providing topics for students to compose narrative writing based on their personal experiences.

Table 1

Participants

Categories	Items	*n*	%
Gender	Male	21	48%
	Female	23	52%
Year	1	7	16%
	2	16	36%
	3	15	34%
	4	6	14%
TOEIC score	400–545 (A2, Pre-intermediate)	16	36%
	550–780 (B1, Intermediate)	16	36%
	785–890 (B2, Upper-Intermediate)	12	28%
Major	Humanities and Social Sciences	13	30%
	Business and Management	10	23%
	Natural Sciences and Engineering	18	41%
	Arts and Sports	3	6%
ChatGPT Editing Experience	Frequently	3	6.8%
	Occasionally	7	15.9%
	Rarely	25	56.8%
	Never	9	20.5%
Location of English Learning	Home	2	4.5%
	School	32	72.8%
	Private academy	2	4.5%
	None	8	18.2%

Procedure

In the first week, a researcher/instructor provided an orientation session and explained the course structure, narrative writing, and course materials. After watching the first five minutes of the film, the students were asked to write a narrative paragraph as a pre-test, which lasted for 15 minutes. They wrote a narrative from the perspective of the protagonist of the movie, Kate, as she prepares a fake pie for her daughter's bake sale. Instead of writing from their personal experiences, the students were instructed to narrate the incidents in the movie with a narrative tone for the pre- and post-tests to control for content and focus on specific aspects of narrative writing. This process helped students to understand the starting point and track their progress. For the pre-test, the students wrote a narrative from the perspective of Kate about making a fake pie for her daughter's bake sale. In the second week, the researcher provided

instructions on writing a piece of narrative writing. The students also learned various ChatGPT prompts for asking for suggestions and editing to improve the clarity, grammar, and overall coherence of the written work. From weeks 3 to 14, the students engaged in narrative writing instruction, and they completed three narrative writing assignments during class under the researcher's supervision. After writing, the students used ChatGPT for revision. The process was conducted in class because of plagiarism and ethical issues generated by ChatGPT (Cotton et al., 2024). Although plagiarism detection tools are being developed, discerning between a student's genuine writing and the responses produced by a chatbot application can still pose challenges. Cotton et al. (2024) proposed a strategy to address these issues in higher education, suggesting close monitoring of students' work. Specifically, they recommend asking students to present their work in class, submit drafts for review, and include in-text citations as well as a list of references. Taking0 these points into consideration, this study had students complete writing assignments during class time.

In the final week (week 15), students submitted a ChatGPT editing feedback analysis, summarizing the revisions and interactions conducted with ChatGPT. For the post-test, students were asked to write a narrative from the perspective of Kate about the email mistakes to her boss, Jack. Table 2 shows a summary of the study procedures.

Table 2

Procedures

Week No.	Description
Week 1	-Course orientation -Pretest (15 min.)
Week 2	-Introduction to narrative writing -Explain ChatGPT prompts for editing
Week 3-14	-Narrative writing instruction -Students' in-class narrative paragraph writing three times: -Student's in-class ChatGPT revision after writing
Week 15	-Students submit ChatGPT editing feedback analysis -Post-test (15 min.)

Data Collection Instrument and Method of Analysis

Narrative writing scoring rubric for overall writing performance

The researcher developed a modified analytic scoring rubric based on previous studies (Macken-Horarik & Sandiford, 2016; Pourdana & Asghari, 2021). The narrative writings were scored based on five criteria: (1) narrative focus, (2) organization, (3) elaboration (4) language and vocabulary, (5) conventions. These criteria are further elaborated as follows:

A. Narrative focus: a setting, characters, and point of view

B. Organization: a plot, sequence of events from beginning to end, opening (hook) and closure

(reflection), cohesive devices

C. Elaboration: details, dialogue, and description

D. Language and vocabulary: using various vivid, descriptive words to describe characters and feelings

E. Conventions: grammar, length, mechanics (punctuation, capitalization, and spelling)

Narrative writing analysis for syntactic complexity

The study also analyzed the students' writings in terms of fluency and syntactic complexity. Fluency was evaluated based on the average number of words and T-units per text. Syntactic complexity was gauged through the mean length of the T-unit (MLT) and number of clauses per T-unit (C/T). To investigate syntactic complexity, the study used two different automated tools, namely, the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010) and the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016). The latter is used to obtain a more detailed picture of noun phrase (NP) complexity. Specifically, the MLT score gives a general indication of how elaborated a particular main clause is but says nothing about the types of elaboration found. (Kyle & Crossley, 2018). By adding the TAASSC tool, the study further examines phrase and clause complexity.

The data were analyzed using SPSS for frequency, and a paired sample *t*-test was run to investigate differences in scores between the pre-test and post-test. Students' writing and analysis were rated by two assessors. For inter-rater reliability, students' written works were rated by two assessors (the researcher and the teaching assistant) who scored the written samples blindly using a rating scale, examined the written works of the students. Preliminary disagreements between the assessors were resolved by consensus discussion until an acceptable level of agreement was reached. Pearson's correlation coefficient between the two assessors was calculated, and interrater reliability was found to be high (r = .84).

ChatGPT editing feedback analysis

Students were required to complete three pieces of narrative writing in class followed by the use of ChatGPT for editing and feedback. After receiving the edited version from ChatGPT, students selected 15 sentences for further discussion. They engaged in an in-depth analysis of the feedback provided and examined the suggested changes. The study considered the underlying rationale, and evaluated how these revisions contributed to enhancing their writing. This analytical process facilitated the development of an in-depth understanding of language usage, grammar rules, and effective writing techniques. Referring to the suggestions of Sovrano et al. (2023), the students requested feedback from ChatGPT by specifically focusing on the use of only the sentences they provided to minimize the potential risk of hallucination.

The submitted reports included the original sentences of the students as inputs, prompts they asked of ChatGPT, outputs received from ChatGPT, and analysis on the suggestions and revised versions of ChatGPT. Students were allowed to write in either Korean or English. Table 3 presents a sample given to the students.

Students' prompts in their reports were classified based on Han's (2023) research that investigated college students' interaction with ChatGPT during essay writing. The interrater reliability test between two raters resulted in a Kappa value of 0.80, which indicates substantial agreement beyond chance.

Table 3

Sample of ChatGPT Editing Feedback Analysis

Prompts + Input	ChatGPT Output	Analysis & Revised version
Prompt: Could you please review and edit the following sentence?	I had trouble sleeping, despite feeling exhausted from a long business trip.	Analysis: When the subject of the main clause and the subordinate clause is the same, omitting the repetitive subject ("I") and using a participial phrase can make the sentence more concise. In doing so, the conjunction "even though" should be replaced with "despite" to introduce a noun phrase. Observing the usage of participial phrases, in addition to learning the grammar rules, helps in better understanding. I believe that I will pay more attention to it when refining my writing in the future.
Input: "I had trouble sleeping even though I was exhausted from a long business trip."		Revised version: I had trouble sleeping, despite feeling exhausted from a long business trip.

Results

Patterns of Students' Use of ChatGPT

The students selected 15 sentences from their narrative writing pieces and interacted with ChatGPT. At the end of the semester, they submitted ChatGPT editing feedback analysis. Based on their reports, Table 4 lists the ChatGPT prompts that they frequently used during the editing stage, which are presented in order of frequency.

Table 4

Students' Common ChatGPT Editing Prompts

No.	Intent	Category	Examples	*n* (%)
1	Request for Language Use	Revision	- Can you help me edit this text? - Proofread for grammar and punctuation errors.	309 (46.8%)
2		Native-like expressions and word choice	- Paraphrase my text to sound like a native speaker. - Edit text to sound more natural and fluent. - Why did you change A into B? What's the difference between A and B?	106 (16.1%)
3		Level of difficulty	- Can you convert it into plain, simple English? - Change it to A1 level of English.	18 (2.7%)
4	Request for Revision (other than language use)	Style & tone	- Include a narrative style, such as sharing a tale with my friend. - Can you edit it using a narrative tone?	36 (5.5%)
5		Register (formality)	- Can you make this text sound less formal and more casual?	42 (6.4%)
6	Request for Translation	Translation	- Can you translate my text into English? - Explain your answer in Korean.	68 (10.3%)
7	Request for Information	Information beyond language use	- Can you show me how to write a good narrative paragraph? - Tell me about English articles.	54 (8.2%)
8	Request for Evaluation	Evaluation	- How would you rate my writing on a scale of 1 to 10? - What do you think about my sentence?	27 (4.1%)
			Total	660 (100%)

The classified prompts are presented in order of frequency: Request for Language Use (n = 433, 65.6%), Request for Revision other than language use (n = 78, 11.9%), Request for Translation (n = 68, 10.3%), Request for Information (n = 54, 8.2%), Request for Evaluation (n = 27, 4.1%). Students most commonly asked ChatGPT about linguistic aspects such as grammar errors or vocabulary choices. Moreover, the majority of students requested feedback on whether or not their writing was accurate using "revise" and "proofread" prompts. They were curious about expressions commonly used by native speakers. Additionally, they inquired about the appropriateness of their original words in terms of nuances and differences (e.g., What's the difference between overlook and ignore?). When the revised version of ChatGPT included complex sentence structures and advanced vocabulary, the students asked to simplify the response into plain and casual English. They thought that ChatGPT tended to respond with lengthy, formal, and complex sentence structures.

The next most commonly used prompt was also related to revision. Requests for revision beyond language use, such as style, tone, and formality, ranked second in frequency after those related to language aspects such as vocabulary and grammar. The writings of a number of students tended to be relatively dry, which focused on conveying factual information instead of outlining a lively and engaging narrative style. The study inferred that the students were influenced by the language classroom materials of Korean students. They studied English using English textbooks or exams. English textbooks contain many fact-based stories, and English exams mainly consist of formal, objective, and structured non-fictional writing.

The third frequent request was the Request for Translation. Students chose to write sentences in Korean when faced with difficulties in finding vocabulary or constructing sentences within the given time. They then requested ChatGPT to translate the sentences that they could not express in English. Request for Information refers to cases in which the students asked for additional explanation regarding the responses of ChatGPT. For example, after receiving feedback on errors in the use of the past progressive tense, they requested explanations with example sentences. This aspect can be viewed as an interaction for additional learning that goes beyond sentence modification. Request for Evaluation refers to requests for the scoring or evaluation of sentences. A few students asked for evaluation from ChatGPT before submission to gauge their performance.

Errors corrected by ChatGPT include tense, articles, verb variety, coherence, wordiness, and word choice and usage. The corrected errors involve surface-level issues, such as tense and word choice, as well as items absent in the native language of Korean such as articles. In terms of verb variety, ChatGPT frequently offers unsolicited feedback when Korean students tend to use simple copular verbs. Students could learn how to express the same ideas in a diverse manner.

ChatGPT also detected errors regarding lack of coherence as in 4). ChatGPT can identify surface-level errors such as tense, punctuation, spelling, and grammatical errors as well as clarity and coherence. As in 5), ChatGPT breaks down lengthy sentences into short ones, such that they sound similar to narrative sentences, and enhance readability. As shown in 6), students made errors interfered by their mother tongue. ChatGPT identifies errors, such as awkward sentences, by directly translating Korean sentences into English. Notably, however, ChatGPT does not address all issues.

Table 5

Common Errors Received by Students from ChatGPT

	Errors	Example
1	Tense	My mom said that bird droppings fall. → My mom said that bird poop had fallen.
2	Article	I participated in national cheerleading competition. → I participated in a national cheerleading competition
3	Lack of verb variety, monotonous verbs (copular verbs/do/get)	I couldn't go to sleep because I thought about things I made on the list. → I couldn't fall asleep as my mind was occupied with thinking about the list of things I needed to do.
4	Lack of coherence	The Bronx accent is a little rough, and by using it, he showed himself as a quiet personality. → By adopting a Bronx accent, he was able to come out of his shell and shed his previous image of being reserved.
5	Verbosity and wordiness	He said to her that just gave Chips Ahoy. → He suggested giving Chips Ahoy.
6	Word choice and usage	As the saying goes "attract to the opposite," → As the saying goes, "opposites attract,"

Effects of ChatGPT-assisted Narrative Writing

Normality tests were conducted using skewness and kurtosis values and the Shapiro-Wilk test. All skewness and kurtosis values were within the range of -1 to +1. The Shapiro-Wilk test results showed that the obtained p-values were greater than .05, indicating that the normality assumption was met. Data were analyzed for differences in scores using the paired sample *t*-test. However, the Shapiro-Wilk test for overall performance showed that it did not meet normal distribution. Therefore, a non-parametric alternative, the Wilcoxon signed-rank test, was employed to compare the scores. In addition, to control for Type-1 error, the study used Bonferroni's correction method and adjusted for the significance level by dividing it by the number of hypotheses, which resulted in a significance level of 0.008.

Table 6 presents the pre-test and post-test results. To evaluate writing fluency and syntactic complexity, the study employed the L2SCA. The two raters evaluated overall performance on the basis of the narrative writing scoring rubric.

The results of the paired sample *t*-test indicate that there were significant differences between students' pre-test and post-test mean scores on writing fluency (Number of Words and Number of T-units). The students produced more than double the amount of text and their T-units also increased by more than two times. However, there was a decrease in the average scores of syntactic complexities. Specifically, the MLT decreased from 13.84 to 11.66, showing a significant difference, while the C/T decreased from 1.55 to 1.48, but the difference was not statistically significant. Two raters evaluated overall performance, and the Wilcoxon signed rank test was used to compare the scores. The result revealed a significant improvement in overall performance compared with that at pre-test.

However, a decrease was observed in the scores in terms of syntactic complexity, which contradicts the results of previous research. Previous studies (Lu, 2011; Ortega, 2003) argued that language learners tend to generate longer T-units as their proficiency increases. This discrepancy may highlight the limitations of machine scoring.

Table 6

*Results of the Paired Sample T-*test for *Pre- and Post-test Narrative Writing*

Category	Component	Test	M	SD	t/Z	p	Effect Size (d)
Fluency	Number of words	Pre	80.16	28.331	-9.441†	0.000	1.71
		Post	163.68	63.007			
	Number of T-units	Pre	6.05	2.332	-10.461†	0.000	2.10
		Post	14.00	4.808			
Syntactic complexity	Mean length of T-unit (MLT)	Pre	13.84	3.901	3.617†	0.001	0.67
		Post	11.66	2.315			
	Number of clauses per T-unit (C/T)	Pre	1.55	0.327	1.437	0.158	0.30
		Post	1.48	0.199			
	Overall performance	Pre	6.18	1.386	-5.726‡	0.000	1.48
		Post	8.57	1.810			

†statistical significance at the level p < .008, ‡Wilcoxon signed rank test

Taking the writing excerpts in Table 7 as an example, the average T-unit was 15.8 in the pre-test, which decreased to 12.2 in the post-test. This result indicates a decrease in syntactic complexity. However, the post-test writings can be considered more vivid in narrative style in terms of content and organization. They start from the topic sentence, describe the setting, organize the sequence of events in a more systematic manner, and use narrative elements, such as direct quotations and emotional expressions, to elaborate.

Table 7

Excerpts from a Student's Writing: Mean Length of T-unit

Test	Example	Length of T-unit
Pre-test	As a working mom, I returned from a business trip	10
	and I wanted to send homemade pies to my daughter's bake sale because of my bad memories as a child and	21
	I didn't want to make her feel different with others and	11
	I didn't want to disappoint her because I run out of time due to a business trip.	17
	So, I made a fake homemade pie by spraying sugar powder on the pie that I bought from the bakery.	20
		Mean = 15.8
Post-test	I made the worst mistake in my whole life.	9
	One day of the meeting, I pitched my fund proposal to Harcourt.	12
	And finally, my fund item was chosen.	7
	I was thrilled about it!	5
	When I got back to my office, I received an email from Allison who is my best friend saying "Let's blow off the task and go to bar to drink."	30
	And moments later I also received an email from Jack saying "I'm looking forward to meeting you."	17
	I replied to Allison saying "would love to get blotto, but sadly can't.	13
	I have to go New York and blow somebody"	9
	and I also replied to Jack, "me too."	8
		Mean = 12.2

To more precisely investigate clause complexity, the study used another tool, the TAASSC, which offers two broader indices of clausal complexity. The first index measures the average number of dependents per clause, while the second measures the standard deviation of the number of dependents per clause, capturing syntactic variation (Kyle, 2016). For 31 clause complexity indices, the study conducted a paired sample *t*-test and applied Bonferroni's correction with an adjusted significance level of 0.008. Only clausal complements per clause displayed a significant difference between the pre- and post-test mean scores of the students (Table 8).

Table 8

Differences in Clausal Complexity

Category	Test	*M*	*SD*	*t*	*p*	*Effect size (d)*
Clausal complements per clause	Pre	.068	.037	-3.722†	.001	0.85
	Post	.133	.060			

†statistical significance at the level p < .008

Increases in the clausal complement can be interpreted as genre-specific aspects of the writers' craft. Macken-Horarik and Sandiford (2016) explained that "whilst verbal projection serves to insert moments of character speech and dialogue into a narrative, it is used to cite authorities in expository texts to quote from primary texts in response genres" (p. 80). With the inclusion of dialogue, quotation, or reported speech, the students described the experiences as well as various aspects of the characters. Examples of clausal complements are as follows:

-Allison called me and said that my email response didn't correspond to her question.

-My boss told me that my proposal has been accepted.

-I kept mumbling and said "No, no, I didn't send you that email. I sent you a different one!"

-I received an email from Jack saying that "I'm looking forward to meeting you."

-Feeling depressed, I replied to Allison's email that "Sadly I can't. I have to go to New York."

Discussion

The study examined the common ChatGPT editing prompts of the students through their ChatGPT editing feedback analysis. The results demonstrated that the most common requests were related to linguistic aspects. ChatGPT demonstrated successful surface-level error detection. The most commonly asked prompts, which are presented in order of frequency, were as follows: Request for Language Use, Request for Revision (other than language use), Request for Translation, Request for Information, Request for Evaluation. Students asked for feedback on correcting errors related to grammar and vocabulary, which aligns with the findings of Han et al. (2023), that is, EFL students primarily use ChatGPT for language use. The students actively engaged with ChatGPT with requests for style and formality adjustment, further information, and evaluation of their writing. Unlike traditional rule-based chatbots that depend on a predefined set of guidelines derived from external sources, ChatGPT offers a more dynamic interaction. (Kohnke et al, 2023), ChatGPT enables students to engage in two-way

communication. This aspect features the ability of ChatGPT as a deep learning model to flexibly respond to various requests.

However, it should be noted that nearly half of the students' requests (46.8%) were simply for editing and proofreading. To get better outcomes with generative AI, it is important to provide specific prompts. However, students often use prompts similar to those they use with Google Translate or Grammarly. For example, the result of a prompt such as "Can you help me edit this text?" is very different from "Can you help me edit my narrative writing to an A2 level of English? I want it to sound informal, as if I were talking to a friend. Include narrative style writing with vivid expressions and direct quotation." This aspect highlights the importance of prompt literacy skill. It refers to knowing how to input appropriate prompts into AI tools to get the desired response. Gattupalli et al. (2023) asserted that AI tools are prompt-dependent, and their value is determined by users who interact with them through well-crafted prompts. Thus, implementing prompt literacy education is seemingly necessary. Toward this end, teachers should understand the basics of crafting effective prompts. And they should test various prompts created by students and compare the results before an AI-assisted writing class. This iterative process not only helps students develop agency but also deepens their understanding of the capabilities and limitations of AI tools. Additionally, it is crucial to note that ChatGPT does not cover all error detections and is mainly successful in providing feedback on surface-level issues, such as clarity, tense, punctuation, spelling, and grammatical errors (Algaraady & Mahyoob, 2023). Instruction related to surface-level editing prompts can be beneficial for EFL writing. Students can note their writing habits, become proficient in editing their work, and become less reliant on teachers while effectively managing time.

Meanwhile, the study recommends that instructors redirect their attention toward content and advanced writing skills. According to Pourdana and Asghari (2021), when students and teachers tend to focus on form and mechanics regardless of proficiency when giving writing feedback. Given that ChatGPT can now partially address mechanics and local and surface errors, the study suggests that instructors should shift their focus toward content. This includes detecting of hallucination and discussing the appropriateness and accuracy of AI-generated feedback. Additionally, instructors should focus on higher-order writing skills such as logical connections and key aspects of specific genres. Specifically, teaching the literary effects inherent in narratives, such as rhythm, alliteration, and language play (Knapp & Watkins, 2005), would be valuable. The present study rarely identified such literary techniques in the writings of the students.

The results showed a significant decrease in syntactic complexity, as measured by MLT, compared to the pre-test. However, it is essential to note that the decrease in MLT should not be solely interpreted as a decline in language proficiency. A closer examination of students' writing and the increase in clausal complement per clause (see Table 7) revealed that the writing exhibited more of a narrative style. It is worth noting that some previous studies (Lu, 2011; Ortega, 2003) indicated that language learners tend to write longer T-units as they become increasingly proficient. However, Ortega (2015) argued that the textual genre in use exerts an impact on cognitive complexity and should be considered. Ryshina-Pankova (2015) emphasized that the analysis of complexity should begin with the analysis of genre, as narrative and argumentative writing display distinct features. As such, making hasty judgments based on syntactic complexity, especially the decrease in MLT in the post-test, would be unwise. Fact-

based writing (e.g., argumentative) requires concise and logical expressions using technical vocabulary, while narrative writing typically involves emotional expressions and differentiated sentence structures. Zhou et al. (2023) explained that short sentences can enhance readability in narrative writing, while Myhill (2008) cited that sentence complexity alone does not indicate sophisticated writing. In narrative writing, interactions between characters through dialogue and the use of short rhetorical questions are frequently utilized to effectively build up events and reactions. Similarly, the students in the current study seemingly became accustomed to the narrative writing genre; therefore, they used shorter, simpler, rhythmical units instead of longer MLT as shown in Table 7. This issue underscores the limitations of machine grading and the necessity of human instructors. Human discretion and judgment cannot be replaced by AI tools. Currently, there is a lack of writing assessment tools that reflect the characteristics of different genres. In this study, the students tended to produce simpler sentences in the post-test. From the syntactic complexity perspective, this tendency could be interpreted as a decrease in sentence maturity. However, this aspect was not a regression; instead, it was a progression toward a narrative genre and away from the initial summary-like writing they demonstrated in the pre-test. This could be detected through careful observation by human instructors. In this regard, the study is hopeful that analytical assessment tools that incorporate genre-specific elements will be developed to help students gauge their progress in self-directed learning. For teachers, these tools can provide insights into the proficiency levels and developmental stages of their students, which enables teachers to offer guidance on which instruction is needed.

Conclusion

Since the emergence of ChatGPT, there have been diverse efforts to integrate it into education. Although ChatGPT can be used as an excellent writing assistant tool, empirical studies that examine its integration into writing education, the use pattern of students, and its impact on the writing skills of students remain few. Thus, the present study aimed to provide insights into the interaction of students with ChatGPT and its impact on their narrative writing. Toward this end, the study design incorporated ChatGPT into an EFL narrative writing class to explore two major aspects: 1) the patterns of college students' use of ChatGPT at the revision stage and 2) the effect of ChatGPT-assisted narrative writing intervention.

Based on the findings, the study provides the following suggestions. First, it confirmed the potential of ChatGPT to serve as a substitute language tutor and a call for a change in the roles of teachers. Students primarily asked ChatGPT about language use, style, and tone, and ChatGPT demonstrated successful surface-level error detection. With the help of ChatGPT, students can receive personalized feedback in real time. Meanwhile, the study suggests that instructors should shift their focus toward content and higher-order writing skills.

Second, prompt literacy education is required. Prompt literacy refers to the competency of interacting with AI, such as providing specific prompts to generative AI, discerning the accuracy of outcomes, and performing a series of actions to obtain the desired outcome (Hwang, 2023). This ability can empower learners to effectively control AI and harness it as a learning mate and, eventually, foster self-directed learning. The results also reveal that advanced English learners critically evaluated outputs from ChatGPT and actively requested clarification and elaboration while comparing these elements to their existing knowledge. In other words, they used ChatGPT to connect the structure of their existing knowledge and

meaningfully incorporate new information into their long-term memory. Conversely, beginners tended to simply accept ChatGPT outputs, even when the sentence structures were extremely complex and contained unfamiliar vocabulary. To achieve this aspect, language teachers also need to be digitally competent (Kohnke et al., 2023). In terms of the pedagogical implications in EFL writing classes, the study suggests the organization of heterogeneous groups for collaborative activities, instead of teacher-dominant prompt instruction. In this scenario, students can collectively discuss potential prompts and brainstorm follow-up questions. By sharing ideas and generating creative prompts together, students can foster innovative thinking, explore diverse perspectives, and, ultimately, equip themselves with the skills required to effectively leverage their writing.

Lastly, there is a need exists for the development of AI tools for writing assessment that reflects the characteristics of different genres. The current results imply that an increase occurs in the use of simple and rhythmical phrases as students become increasingly accustomed to the narrative writing genre. However, machine grading interpreted this aspect as a decrease in syntactic complexity. This result highlights the continued need for careful observation conducted by human instructors. If AI assessment tools that incorporate genre-specific elements are developed, then students can assess their writing progress, which promotes self-directed learning, and educators can provide tailored instructions.

The current study has several limitations that should be acknowledged and addressed by future studies. First, it is conducted as exploratory research, which aims to observe natural phenomena without experimental manipulation and to gain insights. The absence of a control group represents a limitation. The inclusion of a none-AI feedback group is necessary to track improvement over time. This will enable teachers to thoroughly investigate how students optimally utilize generative AI to enhance writing skills (Rastgou, 2023). In addition, this result may be inapplicable to other educational contexts, because a relatively small number of prompts from participants was sampled from one English course in South Korea. Thus, a comprehensive examination of all prompts may produce different results. Next, analysis of ChatGPT editing feedback was derived from self-reported data, which may include potential limitations in terms of objectivity. Moreover, although the results displayed a wide range of patterns that can serve as a foundation for future research, a detailed examination based on students' proficiency levels was not conducted. As such, further investigation into the relationship between students' proficiency levels and synthetic complexity in AI-assisted narrative writing is necessary to verify these results.

ORCID

https://orcid.org/0000-0002-0940-7654

Acknowledgements

Not applicable.

Funding

Not applicable.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

References

Abbott, H. P. (2014). *The Cambridge introduction to narrative* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511816932

Albayati, H. (2024). Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study. *Computers and Education: Artificial Intelligence, 6*, 100203. https://doi.org/10.1016/j.caeai.2024.100203

Algaraady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal, 9*, 3-17. https://doi.org/10.24093/awej/call9.1

Baskara, F. R. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning, 5*(1), 44-55. https://doi.org/10.31763/ijele.v5i1.858

Bedington, A., Halcomb, E. F., McKee, H. A., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers & Composition, 71*, 102833. https://doi.org/10.1016/j.compcom.2024.102833

Billingsley, F., & Gardner, S. (2024). Utilizing generative AI with second-language users and bilingual students. In Y. Poshan, M. James, A. S. S. Zain, & U. Laila. (Eds.), *Facilitating global collaboration and knowledge sharing in higher education with generative AI* (pp. 109-132). https://doi.org/10.4018/979-8-3693-0487-7.ch005

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *JALT CALL Journal, 14*(1), 3-24. https://doi.org/10.29140/jaltcall.v14n1.221

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20,* 38. https://doi.org/10.1186/s41239-023-00408-3

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143-188. https://doi.org/10.1017/S0272263102002024

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59-84. https://doi.org/10.1017/S0272263104026130

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. In Massimo, D. (Ed.), *Philosophical studies series* (Vol. 144, pp. 5-17). Springer. https://doi.org/10.1007/978-3-030-81907-1_2

Garber, A. M., Weenick, M., & Jelinkova, K. (2023). Guidelines for using ChatGPT and other generative AI tools at Harvard. *Office of the provost*. https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard

Gattupalli, S., Maloy, R. W., & Edwards, S. A. (2023). *Prompt literacy: A pivotal educational skill in the age of AI.* College of Education Working Papers and Reports Series. University of Massachusetts Amherst, MA. https://doi.org/10.7275/3498-wx48

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S. Y., & Alice, O. (2023). *Exploring student-ChatGPT dialogue in EFL writing education.* The 37th Conference on Generative AI for Education. Neural Information Processing Systems Foundation. https://gaied.org/neurips2023/files/19/19_paper.pdf

Hart-Davis, G. (2023). *Killer ChatGPT prompts: Harness the power of AI for success and profit*. Wiley.

Holloway, I., & Freshwater, D. (2007). Vulnerable story telling: Narrative research in nursing. *Journal of Research in Nursing, 12*(6), 703-711. https://doi.org/10.1177/1744987107084669

Hwang, Y. (2023). The emergence of generative AI and prompt literacy: Focusing on the use of ChatGPT and DALL-E for English education. *Journal of the Korea English Education Society, 22*(2), 263-288. https://doi.org/10.18649/jkees.2023.22.2.26

Intelligent. (2023, April 13). 8 in 10 teachers approve of student use of ChatGPT, nearly all use it themselves. https://www.intelligent.com/8-in-10-teachers-approve-of-student-use-of-chatgpt-nearly-all-use-it-themselves/

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*, 15873-15892. https://doi.org/10.1007/s10639-023-11834-1

Jairoun, A. A., El-Dahiyat, F., ElRefae, G. A., Al-Hemyari, S. S., Shahwan, M., Zyoud, S. H., Hammour, K. A., & Babar, Z. U. D. (2024). Detecting manuscripts written by generative AI and AI-assisted technologies in the field of pharmacy practice. *Journal of Pharmaceutical Policy and Practice, 17*(1), 2303759. https://doi.org/10.1080/20523211.2024.2303759

Joshi, I., Budhiraja, R., Akolekar, H. D., Challa, J. S., Kumar, D. (2023). "It's not like Jarvis, but it's pretty close!" Examining ChatGPT's usage among undergraduate students in computer science. *ArXiv*. https://doi.org/10.48550/arXiv.2311.09651

Jung, H. (2023). Navigating ChatGPT's challenges to English writing education: Finding a path forward. *The Journal of Modern British & American Language & Literature. 41*(3), 239-261. https://doi.org/10.21084/jmball.2023.08.41.3.239

Knapp, P., & Watkins, M. (2005). *Genre, text, grammar: Technologies for teaching and assessing writing*. UNSW Press.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*(2), 148-161. https://doi.org/10.1016/j.jslw.2011.02.001

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Doctoral dissertation]. Georgia State University. https://doi.org/10.57709/8501051

Kyle, K., & Crossley, S. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal, 102*(2), 333-349. https://doi.org/10.1111/modl.12468

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education, 12*(1), 261-270. https://doi.org/10.5334/pme.1072

Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education, 211*, 104977. https://doi.org/10.1016/j.compedu.2023.104977

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*, 36-62. https://doi.org/10.5054/tq.2011.240859

Macken-Horarik, M., & Sandiford, C. (2016). Diagnosing development: A grammatics for tracking student progress in narrative composition. *International Journal of Language Studies*, *10*(3), 61-94. http://www.ijls.net/pages/volume/vol10no3.html

McGrath, D. (Director). (2011). *I don't know how she does it* [Film]. United States: The Weinstein Company.

Mills, A. (2023). AI text generators: Sources to stimulate discussion among teachers. https://docs.google.com/document/d/1V1drRG1XlWTBrEwgGqd-cCySUB12JrcoamB5i16-Ezw/edit

Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education, 22*(5), 271-288. https://doi.org/10.1080/09500780802152655

Norrick, N., (2016). Indirect reports, quotation and narrative. In A. Capone, F. Kiefer, & F. L. Piparo (Eds.), *Indirect reports and pragmatics* (pp. 93-113). Springer. https://doi.org/10.1007/978-3-319-21395-8_6

Nowik, C. (2022, December 17). The robots are coming! The robots are coming! Nah, the robots are here. *Substack.* https://christinenowik.substack.com/p/the-robots-are-coming-the-robots#details

OpenAI. (2022, November 30). Introducing ChatGPT. OpenAI. https://openai.com/blog/chatgpt

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492-518. https://10.1093/applin/24.4.492

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing, 29*, 82-94. https://doi.org/10.1016/j.jslw.2015.06.008

Pourdana, N., & Asghari, S. (2021). Different dimensions of teacher and peer assessment of EFL learners' writing: Descriptive and narrative genres in focus. *Language Testing in Asia, 11*, 6. https://doi.org/10.1186/s40468-021-00122-9

Rastgou, A. (2023). Prospective methodological considerations in L2 written feedback research. *Feedback Research in Second Language, 1*, 175-196. https://doi.org/10.32038/frsl.2023.01.10

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing, 29*, 51-63. https://doi.org/10.1016/j.jslw.2015.06.005

Sharples, M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education, 32*(4), 1119-1126. https://doi.org/10.1007/s40593-022-00300-7

Sovrano, F., Ashley, K., & Bacchelli, A. (2023). *Toward eliminating hallucinations: GPT-based explanatory AI for intelligent textbooks and documentation.* The 24th International Conference on Artificial Intelligence in Education, Tokyo, Japan. https://ceur-ws.org/Vol-3444/itb23_s3p2.pdf

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays-should academics worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Sun, R., & Deng, X. (2024). Using ChatGPT to enhance experiential learning of college students. *Proceedings of the 57th Hawaii International Conference on System Sciences.* 64-73. https://hdl.handle.net/10125/106384

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Yoon, J. (2013). *I don't know how she does it*. Screen Play.

Zhang, L., & Liu, H. (2021). Genre Effect on L2 syntactic complexity and holistic rating for writing quality of intermediate EFL learners. *Chinese Journal of Applied Linguistics, 44*(4), 451-469. https://doi.org/10.1515/CJAL-2021-0029

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *ArXiv*. https://doi.org/10.48550/arXiv.2303.11812