

A Study on AI-Assisted Feedback in ESL Writing: A Case Study of ChatGPT

Jianmei YANG^{1,2}, Caisang HOU³, Xuanxiao LU⁴, Yong WANG⁵

¹ Department of General Education, Nanchang Medical College, Nanchang, 330052, Jiangxi, China

² College of International Education and Social Development, Zhejiang Normal University, Jinhua, 321004, Zhejiang, China

³ Faculty of Education, The University of Hong Kong, Hong Kong, 999077, China

⁴ Foreign Languages Teaching Department, Hainan Vocational University of Science and Technology, Haikou, 571126, Hainan, China

⁵ School of Foreign Languages, Wuchang Shouyi University, Wuhan, 430064, Hubei, China

Correspondence: Yong Wang, School of Foreign Languages, Wuchang Shouyi University, Wuhan, 430064, Hubei, China. Email: yangying520520@163.com

Received: January 22, 2025

Accepted: June 9, 2025

Online Published: October 10, 2025

doi:10.5430/wjel.v15n8p422

URL: <https://doi.org/10.5430/wjel.v15n8p422>

Abstract

To explore the application value of artificial intelligence in language teaching, this study used 277 ESL compositions by Chinese learners as its corpus. It invited three senior English teachers and ChatGPT-4 to grade the same set of compositions and systematically compared the similarities and differences in their feedback. The study found that ChatGPT provided significantly more feedback than the teachers. Teachers' feedback focused on grammatical and vocabulary errors, demonstrating the characteristic of "precise focus"; ChatGPT, by contrast, conducted comprehensive corrections, frequently replacing vocabulary and structures, thereby reflecting the feature of "comprehensive coverage." In terms of feedback focus, ChatGPT excelled at optimizing written language style and delivering comprehensive feedback, whereas teachers tended to employ strategies of indirect feedback and targeted local guidance. Regarding personalization, teachers offered positive encouragement tailored to students' backgrounds, rendering their feedback highly personalized; ChatGPT's feedback, in contrast, was standardized and lacked emotional care for learners. The study also identified limitations of ChatGPT: inaccurate or excessive corrections, coupled with weak detection of logical errors; feedback language often exceeding students' proficiency level, with extensive reformulations increasing cognitive load (in contrast to teachers' feedback, which aligns with the "i+1" principle); and inappropriate remarks or fabricated academic references, which undermine academic integrity, compromise research credibility, and obscure achievement attribution. Based on these findings, the study proposes optimization suggestions to provide useful insights for ESL writing teaching practice.

Keywords: ESL writing, written corrective feedback, ChatGPT, teaching implications

1. Introduction

Since the introduction of process writing into second language (L2) writing instruction in the 1960s and 1970s, feedback has become an integral component of writing pedagogy. Teacher-provided written corrective feedback (WCF) stands as one of the primary forms of writing feedback and plays a pivotal role in enhancing L2 learning efficiency (Mahfoodh, 2017). Effective corrective feedback helps L2 learners identify their own shortcomings. Long (Long, 1996) argues that teacher feedback is crucial for L2 learners, as it allows correct linguistic knowledge to replace their erroneous assumptions about the target language. Teacher-written corrective feedback can assist students in improving their writing proficiency by providing written responses to errors in their texts, thereby further facilitating their linguistic development (Hyland, 2013). However, teachers face difficulties related to the time spent on feedback provision, pressure from large class sizes, time constraints for returning feedback, and emotional fatigue (Yu, Zheng, Jiang, Liu, & Xu, 2021). Artificial intelligence (AI), though, has to some extent addressed the limitations of teachers' written feedback. ChatGPT, a chatbot driven by advancements in natural language generation (NLG), with its powerful computing and data-processing capabilities, has revolutionized the traditional teacher-centered model of writing feedback. Its robust functionality and user-friendly accessibility have exerted a profound impact on L2 writing instruction. Does the integration of generative AI technologies into L2 writing open an "Aladdin's lamp" or a "Pandora's box"? How will teachers' roles in L2 writing feedback evolve? Against the backdrop of these questions, this study uses college English essays as a corpus to compare feedback generated by ChatGPT with that provided by teachers. It aims to explore the characteristics of written corrective feedback produced by GPT, identify differences between GPT-generated feedback and teacher feedback, and examine their respective strengths and potential limitations. In doing so, the future integration of AI into teachers' WCF practice may thus be optimized.

2. Literature Review

2.1 Teachers' WCF

Research on second-language writing feedback has long been a hot topic in the fields of language education and applied linguistics.

Regarding the effectiveness of corrective feedback in second-language writing, there are differing views in the academic community. Some scholars argue that corrective feedback does more harm than good and may even hinder the improvement of learners' writing abilities (Truscott, 1996, 2007). However, another group of scholars hold that corrective feedback can play a positive role in helping learners improve their writing skills (Bitchener, 2008; Bitchener & Knoch, 2009; Fathman, 1990; Robb, Ross, & Shortreed, 1986; Santos, Serrano, & Manchón, 2010; Van Beuningen, De Jong, & Kuiken, 2012). Scholars at home and abroad have conducted numerous empirical studies on the effectiveness of corrective feedback (Bitchener & Knoch, 2009; Chandler, 2003; Ferris & Roberts, 2001; Sheen, 2007). Only a few researchers consider corrective feedback ineffective (Kepner, 1991; Truscott, 1996, 2007; Truscott & Hsu, 2008). However, more disputes center on which type of feedback is more effective (Fathman, 1990; Lalande, 1982).

Ellis (Ellis, 2009) provided a systematic overview of the types of feedback. Based on different sources of feedback, it can be classified into teacher feedback, peer feedback, and software feedback. According to the clarity of feedback, it can be divided into direct feedback, indirect feedback, and metalinguage feedback. In terms of the focus of feedback, it can be categorized into focused/selective feedback and unfocused/comprehensive feedback. The basic concepts of each type of feedback are shown in Table 1 as follows:

Table 1. Types and Concepts of Corrective Feedback in Second Language Writing

Dimension	Type	Concept
Source of Feedback	Teacher Feedback	Evaluation and guidance provided by teachers regarding learners' writing content, structure, language use, and other aspects.
	Peer Feedback	Evaluation and suggestions given by learners to each other's written work in terms of content, logic, language expression, and other aspects.
	Software Feedback	Feedback provided by writing-related software or online essay correction platforms based on preset algorithms and rules, addressing learners' grammatical errors, vocabulary use, discourse structure, and other aspects of their writing.
Explicitness of Feedback	Direct Feedback	Explicitly pointing out errors in learners' writing, such as grammatical errors or spelling mistakes, and directly providing correct expressions or improvement methods.
	Indirect Feedback	Not directly identifying errors but guiding learners to discover problems in their writing and consider directions for improvement through questioning, hinting, and other methods.
	Metalinguistic Feedback	Providing linguistic knowledge terminology related to writing, such as rhetorical devices and stylistic characteristics, to explain and correct errors and deficiencies in learners' language use.
Focus of Feedback	Focused/Selective Feedback	Giving focused attention and feedback on specific points in learners' writing, such as the logical argumentation of a particular paragraph or the accuracy of a certain type of vocabulary use.
	Unfocused /Comprehensive Feedback	Providing comprehensive and integrated evaluation and feedback on learners' entire writing from multiple aspects such as theme, structure, richness of content, and fluency of language.

In recent years, academic circles have conducted research on the eight feedback types across the three aforementioned dimensions, yet the focus has largely been on the latter two. These include direct feedback, indirect feedback, and metalinguistic feedback (categorized by explicitness), as well as focused/selective feedback and unfocused/comprehensive feedback (categorized by focus). Within academic discourse, particular attention has been paid to the effectiveness of direct and indirect feedback. Proponents of direct feedback argue that it reduces confusion arising from learners' failure to understand or retain information, providing them with more adequate guidance and thereby enhancing learning efficiency. Van Beuningen, De Jong, and Kuiken (Beuningen, Jong, & Kuiken, 2008) conducted experiments comparing direct and indirect feedback on nine error types, finding that the direct feedback group achieved higher linguistic accuracy than the indirect feedback group. Advocates of indirect feedback contend that guided comments and feedback can prompt students to engage in critical thinking about their writing, improve their linguistic and critical thinking skills, foster their ability to solve problems independently, and facilitate long-term acquisition of the target language. Razie, Mehdi, and Ahmad (Abedi, Latifi, & Moinzadeh, 2010) compared intermediate-level English learners in Iran, with one group receiving direct error-correction feedback and the other receiving indirect feedback that marked error types. Their findings indicated that indirect feedback was more conducive to improving linguistic accuracy. There is also a group of researchers who argue that no significant difference exists between the two feedback forms. Semke (Semke, 1984) divided 141 German learners into four groups, each receiving a different type of feedback: direct error correction; content-focused comments; a combination of direct error correction and content comments; and indirect indication of error types for students to correct independently. The results showed no significant differences among these feedback types. Achieving an appropriate balance between direct

and indirect WCF may therefore be considered essential in writing teachers' practice to suit specific pedagogical goals (Hyland & Hyland, 2006).

Research has also explored whether teachers' feedback should cover all errors (unfocused, also referred to as comprehensive written corrective feedback) or target only specific ranges of error types (focused written corrective feedback). Empirical studies have tested and compared the effectiveness of these two approaches, revealing that focused feedback is more conducive to enhancing linguistic accuracy in writing than unfocused feedback (Benson & DeKeyser, 2019; Suzuki, Nassaji, & Sato, 2019). However, some researchers' studies have only confirmed the effectiveness of both feedback types without establishing which one is more effective (Ellis, Sheen, Murakami, & Takashima, 2008). That said, many representative studies have been conducted under experimental conditions rather than in authentic classroom settings, raising concerns about the practical value of their findings (Lee, Luo, & Mak, 2023).

2.2 ChatGPT-assisted feedback

ChatGPT, a natural language generation (NLG)-based system capable of producing texts exceeding 25,000 words, leverages robust language understanding and generation capabilities to deliver immediate, personalized written corrective feedback for L2 writing. Its functionality rests on five core technologies: (1) Transformer architecture: A neural network for sequential data processing, it uses self-attention mechanisms to model long-distance dependencies, enabling efficient large-scale data handling and contextual information capture—thus enhancing comprehension and generating coherent, contextually fitting responses. (2) Pre-training: Trained on massive text datasets via unsupervised learning, ChatGPT acquires linguistic patterns and rules, gaining extensive language knowledge and contextual awareness to excel in tasks like dialogue generation. (3) Fine-tuning: By integrating guiding instructions or attention weights during training, this technology directs ChatGPT to prioritize task-specific key information, boosting performance in targeted tasks. (4) Human feedback-based reinforcement learning: Real-time human input allows adjustments to generated text, facilitating faster adaptation to tasks or scenarios, aligning outputs more closely with human expectations, and improving accuracy and rationality. (5) Transfer learning: This involves applying general language representations from large-scale text data to specific tasks, enabling quicker adaptation to task requirements and enhancing overall performance and effectiveness.

Few studies have thoroughly compared GPT-generated written corrective feedback (WCF) with teacher-provided WCF. Yoon et al. (Yoon, Miszoglád, & Pierce, 2023) assessed GPT's feedback quality using 50 L2 English essays from a public corpus, where GPT offered comments, scores, and illustrative examples. Quality was measured via accuracy and usability metrics: writing experts coded feedback at the sentence level into direct or indirect types based on explicitness, while usability focused on linguistic specificity and clarity. Unlike prior research on human WCF, GPT favored indirect feedback over direct forms but produced generic, superficial comments—overemphasizing cohesive devices and missing key issues, especially in low-scoring texts. Guo and Wang (Guo & Wang, 2024) compared GPT and teacher feedback in terms of quantity and form, analyzing GPT's responses to 15 EFL learners' argumentative essays against feedback from 5 Chinese EFL instructors. They quantified feedback across content, language, and organization, categorizing it as 'directive', 'informative', 'query', or 'praise'. GPT generated more feedback overall, distributing it evenly across the three domains, whereas teachers focused specifically on content and language issues.

The nature of prompts used to generate WCF is also likely a key factor in its success. A prompt, as user-provided natural language input, supplies relevant information to guide the model in specific tasks (Si et al., 2022) and enables emergent abilities that allow it to produce outputs beyond its training data (Wei et al., 2022). Prompt design significantly affects task performance, with well-crafted prompts yielding representative responses (Hou, Dong, Wang, Li, & Che, 2022; Kung et al., 2023). Li et al. (Li et al., 2023) identified two primary prompt-formulation strategies: template-based (using predefined templates with all required information) and regeneration-based (reevaluating and improving initial responses). Prompt complexity ranges from basic single-shot templates, contextual, role-based, and constraint-based prompts at the lower end to regeneration-based nested and iterative prompts at the higher end.

Compared with automated writing evaluation (AWE) systems, NLG systems, leveraging more advanced natural language processing (NLP) and natural language generation (NLG) technologies, can automatically generate artificial intelligence-generated content (AIGC) that meets users' needs based on the information provided. In writing instruction, issues such as the differences between ChatGPT and teacher-provided written corrective feedback, as well as how to make good use of artificial intelligence technologies to better enhance teaching and learning effectiveness, merit in-depth reflection and exploration by educators. Against this backdrop, this study uses samples of college English essays as corpus to conduct a comparative analysis of the characteristics of teacher feedback and ChatGPT feedback. It aims to fully demonstrate the multiple possibilities of artificial intelligence-assisted L2 English teaching and provide a model for the widespread application of ChatGPT in higher education.

3. Method

3.1 Research Questions

- (1) How does the written corrective feedback provided by ChatGPT perform?
- (2) What are the similarities and differences between the written corrective feedback provided by teachers and ChatGPT?

3.2 Data Collection

The corpus of this study comprises 277 essays, including both regular practice essays and final-term essays written by second-year undergraduates from various majors at the author's affiliated university. Each essay is approximately 200-300 words in length, with a

total word count of 69,254. The essay types cover CET-4 and CET-6 compositions, argumentative essays, resume writing, and others. The participating students completed a total of 96 class hours of basic English courses over two semesters during their first year of university, among whom approximately one-fourth have passed the College English Test Band 4 (CET-4).

Three senior teachers (with over ten years of experience in teaching English writing) were invited to manually grade these 277 essays. Concurrently, a researcher used ChatGPT-4 to generate corrective feedback for the same set of essays via pre-set prompt instructions, operating within a fixed time window and on the same computer device. To ensure the objectivity and accuracy of the feedback results, a new conversation session was initiated for each essay when using ChatGPT for grading. This procedure effectively avoids model memory bias and prevents interference from historical interaction data on current feedback, thereby ensuring the reliability of the research data.

3.3 Implementation Procedures

The specific implementation procedures are as follows:(1) Systematically categorize the collected teachers’ written corrective feedback, and ensure the accuracy and consistency of the classification results through cross-validation;(2) Construct an evaluation dimension system for ChatGPT’s written corrective feedback (see Table 2) by referring to the introduction of ChatGPT’s functions in ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI (Atlas, 2023). Among them, primary-level feedback includes punctuation correction, spelling errors, vocabulary misuse correction, vocabulary replacement, grammar modification, and structure replacement; intermediate-level feedback covers semantic fluency and textual coherence; advanced-level feedback involves stylistic appropriateness and logical reasoning (Note 1¹); (3) Merge and reorganize the categorized teachers’ feedback types with reference to ChatGPT’s feedback dimensions to achieve standardized correspondence between the two types of feedback data;(4) Select representative typical cases from each dimension based on the frequency of feedback occurrence;(5) Input modification instructions into ChatGPT in accordance with the instruction specifications proposed by Atlas(Atlas, 2023) , obtain corresponding written corrective feedback, and conduct a systematic comparative analysis with teachers’ feedback.

Table 2. Main Dimensions of Written Corrective Feedback

Feedback Level	Types of Written Corrective Feedback	Specific Description
Primary	Punctuation Correction	Revise cases of misused punctuation marks
	Spelling Errors	Identify and correct misspelled words caused by spelling mistakes, slips of the pen, etc.
	Vocabulary Misuse	From the perspective of context adaptation, point out issues such as confusion of near-synonyms, misuse of spoken and written language, etc.
	Vocabulary Replacement	Under the premise of maintaining the intended meaning, select more precise, appropriate, and context - style - fitting vocabulary for replacement
	Grammar Modification	Correct grammar problems such as incomplete sentence components, tense errors, improper word order, etc.
	Structure Replacement	The original wording does not constitute a grammar error, but it is restated using another expression
Intermediate	Semantic fluency	Adjust the content of sentences and paragraphs in the text to eliminate ambiguous and obscure expressions and ensure clear and coherent meaning
	Stylistic Appropriateness	Adjust the language style according to the application scenario of the text to make the expression conform to stylistic requirements. For example, convert colloquial expressions to written language, use vivid rhetoric in literary creation, and use professional terms in scientific and technological texts, etc.
Advanced	Logical reasoning	Comprehensively examine the thread of the text and correct logical problems such as one - sided views, insufficient arguments, and illogical reasoning

3.4 Design of Prompts

As a key element in generating written corrective feedback, prompts significantly influence the final feedback effect. Li et al.(Li et al., 2023) summarized three common prompt templates: inquiry-based, context-based, and example-based. Among these, context-based

prompts can effectively enhance instruction clarity by supplementing task details and specific contexts, thus being widely applied in written corrective feedback tasks (Liu et al., 2021; Wang, Valdez, Basu Mallick, & Baraniuk, 2022). Additionally, Reynolds & McDonell (Reynolds & McDonell, 2021) noted that formulating prompts in natural language and clearly expressing expected instructions is an important strategy for optimizing feedback effects. For instance, using a simple task name (e.g., “Provide written corrective feedback”) versus a detailed descriptive request (e.g., “Provide written corrective feedback on this essay in any form you choose, to help the author revise the text and master language knowledge”) yields distinctly different feedback results.

However, existing studies have shown that language teachers exhibit significant differences in their ability to compose complex and effective prompts (Ma, Crosthwaite, Sun, & Zou, 2024). Based on this, the present study focuses on L2 English teachers who have only received basic prompt training, taking their typical prompting strategies in actual teaching as a starting point to explore methods for improving the quality of written corrective feedback by designing contextualized prompts. Meanwhile, drawing on the approach of Reynolds & McDonell (Reynolds & McDonell, 2021), descriptive requests rather than concise task names are adopted in designing prompts for both ChatGPT and teachers, aiming to enhance the effectiveness of written corrective feedback (see Table 3).

Table 3. Prompts Designed for Teachers/ChatGPT

	Content of Prompts
Teacher	Please manually grade the problems in the following text in terms of vocabulary use, grammatical structure, punctuation marks, word spelling, semantic expression, text structure, logical argumentation, and presentation of views. Provide corrective feedback in diverse forms to help the author revise and perfect the text. In addition, combine the grading situation to refine targeted modification suggestions to provide reference for learners to optimize their compositions.
ChatGPT	Suppose you are an English teacher. You need to address the problems in the following text regarding vocabulary use, grammatical structure, punctuation marks, word spelling, semantic expression, text structure, logical argumentation, and presentation of views. Provide written corrective feedback by clearly marking the modification positions to help the author revise and improve the text.

4. Results and Discussion

4.1 Data Statistics

Through systematic collection and collation, this study selected 277 student essays as analytical samples, obtaining a total of 5,085 items of written corrective feedback from teachers and 9,753 items from ChatGPT. In-depth analysis revealed that the two types of feedback exhibit multiple differences and commonalities in terms of content dimensions, feedback depth, and guidance methods. These similarities and differences not only reflect the distinct advantages of artificial intelligence and human teachers in language teaching but also provide empirical evidence for optimizing English as a Second Language (ESL) writing instructional strategies and constructing learning support systems. For specific results, refer to Table 4.

Based on the data in Table 4, the characteristics of the written corrective feedback from teachers and ChatGPT can be summarized as follows:

1) Primary-level Feedback: Different Emphases in Basic Revision

Teachers’ feedback focuses on the accuracy of grammar and vocabulary. Among them, grammar modification accounts for 35.7%, and vocabulary misuse correction accounts for 21.3%, reflecting their emphasis on constructing the language norm system. Notably, when students’ expressions have no obvious grammar errors, teachers rarely perform structure replacement (accounting for 0.7%) and only carry out a small amount of vocabulary replacement (3.6%). In contrast, relying on a large - scale corpus, ChatGPT tends to optimize expressions through synonym replacement of vocabulary (26.3%) and structures (30.1%). However, its sensitivity in identifying details such as misspelled words and vocabulary misuse is significantly lower than that of teachers. In addition, in terms of feedback on English character norms and punctuation formats, teachers, with the intuitive form of circling and commenting, have a higher proportion than ChatGPT; while ChatGPT, restricted by the text feedback form, can only give prompts by restating sentences and is difficult to achieve precise marking.

2) Intermediate-level Feedback: Functional Complementarity in Expressing Meaning and Coherence

In the dimension of “semantic fluency”, the proportion of ChatGPT’s feedback is slightly higher than that of teachers, indicating that it has certain advantages in constructing textual coherence. In the feedback of “stylistic appropriateness”, the proportions of teachers and ChatGPT are 4.3% and 5.2% respectively, indicating that they have comparable abilities in controlling stylistic adaptability. From the overall data, neither the quantity nor the proportion of the two in the intermediate - level feedback has formed a significant difference,

which shows that their functions at this stage are highly complementary.

3)Advanced-level Feedback: Significant Differences in Thinking Cultivation

In terms of “logical reasoning”, the proportion of teachers’ feedback (3.3%) is significantly higher than that of ChatGPT (1.7%), highlighting the irreplaceability of teachers in guiding learners to improve logical thinking and deepen argumentation ability, while ChatGPT is relatively weak in high-order thinking intervention.

Table 4. Statistics Table of Written Corrective Feedback Types: Teachers vs ChatGpt

Feedback Level	Types of Written Corrective Feedback	Teacher Feedback		ChatGPT's Feedback	
		Quantity/ piece	Proportion/ %	Quantity/ piece	Proportion/ %
Primary	Punctuation Correction	207	4.0%	136	1.4%
	Spelling Errors	635	12.4%	166	1.7%
	Vocabulary Misuse	1085	21.3%	634	6.5%
	Vocabulary Replacement	185	3.6%	2565	26.3%
	Grammar Modification	1820	35.7%	115	11.8%
	Structure Replacement	40	0.7%	2935	30.1%
Intermediate	Semantic Fluency	753	14.7%	1492	15.3%
	Stylistic Appropriateness	223	4.3%	507	5.2%
Advanced	Logical Reasoning	137	3.3%	1203	1.7%

4.2 Case Analysis

4.2.1 Primary -level Feedback

Primary-level feedback encompasses punctuation correction, spelling errors, vocabulary misuse correction, vocabulary replacement, grammar modification, structure replacement etc. The corpus used here is derived from an essay training assignment for the College English course. The topic is “Should one pursue graduate studies or enter the workforce after completing undergraduate studies?” After posting this topic on the teaching platform, the teacher received assignments submitted by students and provided feedback (see Figure 1).

As shown in Figure 1, there are several obvious grammatical issues in the student's essay: (1) The simultaneous use of "although" and "but," with one of them needing to be deleted; (2) The phrase "pure" combined with "do not have so much trouble" is semantically unclear and exhibits obvious traces of Chinglish; (3) The word "just" is too colloquial and unsuitable for written language; (4) Misuse of the preposition "in"; (5) A mixture of main and subordinate clauses, which should be separated by commas. The teacher adopts indirect and focused feedback, pointing out the grammatical and punctuation issues in the student's essay.

题量: 1 满分: 100
作答时间: 09-22 17:19 至 09-29 17:19

一. 简答题 (共1题, 100分)

1. (简答题) Suppose you have two options upon graduation: one is to take a job in a company and the other to go to a graduate school. **You are to make a choice between the two. Write an essay to explain the reasons for your choice. You should write at least 120 words but no more than 180 words.**

72 分

我的答案:

For undergraduates college students, deciding what to do upon graduation can be a very tough decision to say the least. As far as I am concerned, I prefer to pursuing postgraduate students.

For one thing, personally speaking, though college just like a some society but people in there are all kind and pure do not have so much trouble. For another thing, entering into a postgraduate school, I can improve my job prospects. Specializing at postgraduate level can make you a stronger job applicant. All in all, when in comes time for university graduates to decide their next step in life, there is no right or wrong choice for everyone. Rather, each students must reach his or her own conclusion.

正确答案:

教师批语:

For one thing, personally speaking, though college just like a some society but people in there are all kind and pure do not have so much trouble. 该句标点与语法均有问题, 观点表达模糊, 逻辑关系不清晰。

Figure 1. Corrective Feedback Provided by the Teacher

Compared with the corrective feedback provided by the teacher, ChatGPT offers more detailed direct and focused feedback on issues such as vocabulary misuse (problems (1) and (4)), colloquial expressions (problem (3)), structural confusion, and punctuation modifications (problem (5)), while preserving the original sentence structure (see Figure 2). Additionally, for problem 2, ChatGPT combines focused and unfocused feedback, with its advantages mainly manifesting in two aspects: On the one hand, the student's vocabulary misuse stems from their literal translation of "单纯" as "pure," whereas "pure" when used to express "morally good" typically modifies objects, such as "a pure life/motive (Note 2ⁱⁱ)". ChatGPT accurately captures the student's intention and revises "pure" to "moral." This analysis helps students re-examine the multiple meanings of "pure" and make further revisions, enabling them to understand the "why" behind the change. On the other hand, ChatGPT interprets "do not have so much trouble" in problem 2 as a concrete manifestation of "moral," with the logic being that "people in university are kind and simple, so (those people) do not cause much trouble." However, the student's subsequent feedback indicates that their original intention was to use causality for argumentation, i.e., "because people in university are kind and simple, so (you) choosing to continue graduate studies will not cause trouble." Since there are no causal markers in the student's essay, ChatGPT can only modify based on the literal meaning, failing to accurately "understand" the student's "implied meaning."


 Sure, here's a corrected version with proper grammar and punctuation "Personally speaking, although college can feel like a society, the people there are generally kind and moral, and don't cause too many problems." I made a few changes to the original sentence to improve its clarity and correctness. These changes include: 1. Removing the word "just" and rephrasing the first part of the sentence for clarity. 2. Using "moral" instead of "pure" to convey a sense of overall ethical behavior. 3. Adding a comma after "moral" to properly separate the clauses in the sentence. 4. Removing "do not have so much trouble" and replacing it with "don't cause too many problems," which is a clearer and more precise way of conveying the idea.

Figure 2. Corrective Feedback Provided by ChatGPT

4.2.2 Intermediate-level Feedback

Intermediate-level feedback covers semantic fluency and textual coherence, aiming to enhance students' vocabulary richness, improve the accuracy and conciseness of their writing, and make it more suitable for specific stylistic conventions. In writing training, when asking students to express their opinions on doctor-patient relationships, the original draft of a student's essay is shown in Figure 3.

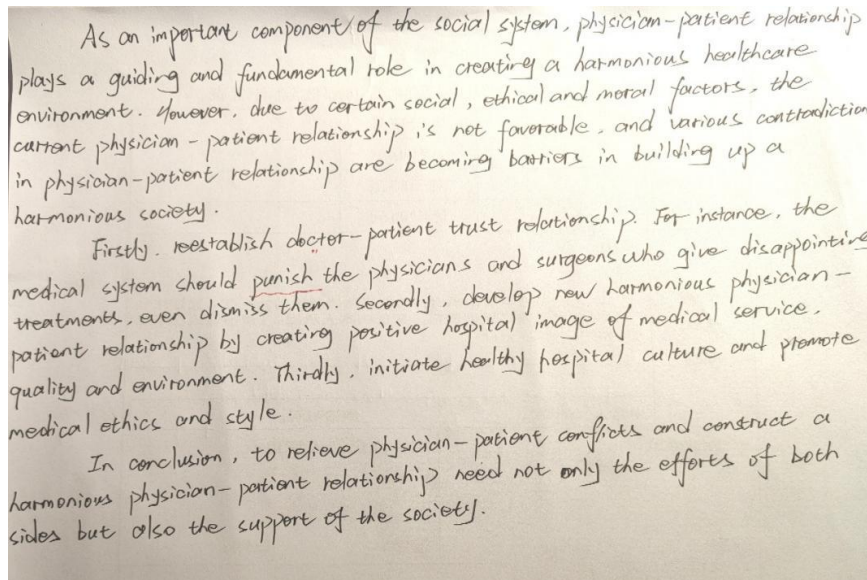


Figure 3. Student's manuscript

Teacher's feedback: "Punish" is a strong word. Please use an alternative. The teacher provides focused indirect feedback, suggesting that the word "punish" should be replaced with another term. After receiving the indirect feedback from the teacher, the student further communicated with the teacher and stated in their subsequent feedback that they were "unsure how to choose a more idiomatic word." This demonstrates that many students have a relatively clear understanding of the problem of inadequate expression, but due to their current limited English proficiency, they are unable to produce more idiomatic expressions, thus encountering a bottleneck in writing or revision. This also reinforces the view that indirect feedback has limited effectiveness for first-year university students with lower English proficiency, as they may not yet possess the linguistic ability to correct the writing errors they are aware of. ChatGPT, however, effectively addresses this issue. Its direct feedback not only provides multiple vocabulary expressions to replace "punish" but also enriches the discourse expression of the student's essay by varying sentence structures. While focusing on content feedback, it also provides formal feedback, balancing both local and overall feedback (as shown in Figure 4).

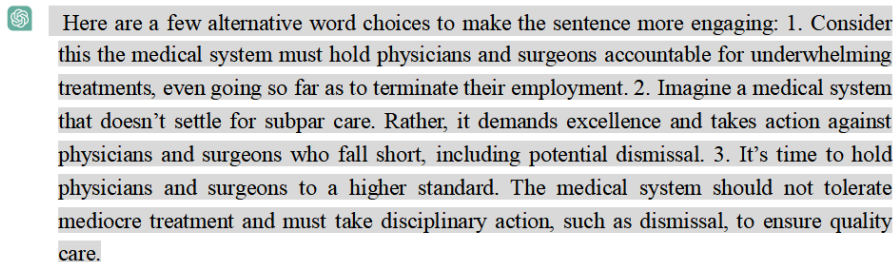


Figure 4. Corrective Feedback Provided by ChatGPT

Example 1 is a piece of writing material selected from essay training exercises. Students were asked to write an essay of about 120 words on the topic of "How to solve the problem of not affording to see a doctor." However, the sentence "Finally, everyone has a lot of responsibility for their own health, and it is for everyone's needs" is semantically unclear, leading to ambiguity and confusion.

Example 1:

Student's composition: *Many people cannot afford to see the doctor, which is a big problem. One way to help is to create more medical places for all people to use, it would make things better. Public education about health could help everyone also, so people can prevent illness. Charities can help poor people get to hospitals, which is very important. Finally, everyone has a lot of responsibility for their own health, and it is for everyone's needs.*

Teacher's feedback: The sentence "Finally, everyone has a lot of responsibility for their own health, and it is for everyone's needs" is semantically unclear.

ChatGPT: Many people cannot afford to see a doctor, which is a significant problem. One way to help is to create more healthcare facilities for all people to use, which would improve access. Public education about health could also help everyone, so people can prevent illness. Charities can help poor people get to hospitals, which is very important. Ultimately, everyone has a lot of responsibility for their own health, and this should be a priority for everyone.

The sentence “Finally, everyone has a lot of responsibility for their own health, and it is for everyone’s needs,” is meant to create ambiguity. “and it is for everyone’s needs” is unclear. It’s not clear what “it” is referring to, and the connection to “responsibility” is unclear. Is it saying that health responsibility is everyone’s needs, or that it fulfills everyone’s needs? Is it something needed by everyone, or is it for the benefit of everyone? The phrasing is vague and doesn’t quite make sense in context. In the corrected version, the sentence is changed to “Ultimately, everyone has a lot of responsibility for their own health, and this should be a priority for everyone.” This resolves the ambiguity and makes the sentence clear and grammatically sound. This emphasizes that health responsibility should be everyone’s priority.

Beyond semantic fluency, ChatGPT can also be utilized for style adjustment. Example 2 is taken from a student's final essay, describing a significant setback they have experienced. Compared to indirect feedback from teachers, the direct feedback provided by ChatGPT can assist students with inadequate English proficiency in "tracing and finding clues," enabling them to clearly understand their errors and quickly make self-corrections. As demonstrated in Example 4, following the instructions, ChatGPT corrected the issue of mixing written and spoken language in the original text and, in accordance with the writing style of a narrative essay, employed more descriptive language or vocabulary to make the story more fluid and vivid, enhancing its dramatic quality and making the "tragic" aspect even more prominent.

Example 2:

Student's composition: *Last month, the weather was so bad, it's been raining for days, which made the campus environment very unpleasant. I was studying hard on campus when I suddenly started to feel really sick. My throat was super painful, I almost cannot talk, and I also had a high fever. Because of that, I must go to the hospital, where the doctor told me I caught a serious flu and needed to take medicine and lots of rest. This illness totally messed up my study plans. I couldn't attend the lectures and was also unable to go to the library to prepare for the exam. Even worse, the group project I was leading was facing problems because I was too sick. This made me very stressed and down, and I am worried about the upcoming exams. The sickness has been a real obstacle to my college life, making it so difficult. I just want to recover soon and return to my normal life and get healthy. I just want to back to normal and healthy.*

Teacher's feedback: The student's essay contains a few language errors and a mixture of written and colloquial expressions. There is room for further polishing in terms of stylistic refinement.

ChatGPT: Here's a revised version of the paragraph to improve its narrative flow: “Last month, the weather was just awful. The rain never stopped, turning the campus into a gray, sad place. I remember being so focused on my books when a sudden sickness hit me. My throat started hurting really bad, my voice became just a whisper, and a hot fever came over me. I had no choice but to go to the hospital. The doctor, looking serious, told me it was a bad flu, and I needed medicine and lots of rest. This illness completely messed up my study plans. Lectures seemed miles away, and the library, a place I usually loved, became unreachable. Even worse, the group project started to fail because I was gone, a key part missing. Stress and worry became a constant, quiet hum in the background, along with thoughts of the exams. The sickness felt like a heavy weight, a real problem in my college life. I just want to get better soon, to feel the sun again, and to go back to a healthy life.” In this revision, I've tried to use more descriptive language to create a clearer and more engaging narrative. I've also reorganized the ideas and sentences to create a more natural flow of events.

4.2.3 Advanced-level Feedback

Advanced-level feedback involves logical reasoning where ChatGPT can also generate outputs according to specific texts or requirements. The following material is taken from CET-6 essay training, where students read the text and then write a synopsis(as shown in Figure 5). From ChatGPT's synopsis writing, it is clear that the generated synopsis meets the requirements of teachers' indirect feedback, which involves providing a comprehensive overview of each paragraph and effectively connecting the topic sentences of each section. Therefore, for certain essay topics that lack reference examples, teachers can fully utilize ChatGPT's advanced output capabilities to generate examples, serving as a supplement to their indirect feedback.

Task 4

1. (写作题, 100分)

Read carefully the following excerpt and then write your response in **NO LESS THAN 120 WORDS**, in which you should:

- 1) summarize the study results, and then
- 2) comment on the relationship between loneliness and wisdom.

You can support yourself with information from the excerpt. Marks will be awarded for content relevance, content sufficiency, organization and language quality. Failure to follow the above instructions may result in a loss of marks.

Loneliness and Wisdom: Are They Related?

Rising rates of loneliness may not be news, but the three periods when it peaks may come as a surprise: More people reported feeling moderate to severe loneliness during their late 20s, their mid-50s and their late 80s than in other life periods, according to the research published in the journal *International Psychogeriatrics*.

The general sense of isolation was also more prevalent than the researchers had expected. A full three-quarters of the study participants reported moderate to high levels of loneliness, said Dr. Philip Jeste, senior author of the study and a professor of psychiatry and neuroscience at the University of California, San Diego.

"One thing to remember is that loneliness is subjective. Loneliness does not mean being alone; loneliness does not mean not having friends," said Dr. Jeste, who is also director at the UC San Diego's Center for Healthy Aging. "Loneliness is defined as 'subjective distress'." It is the discrepancy between the social relationships you want and the social relationships you have, he said.

Within the dark clouds, Dr. Jeste also found a silver lining: An inverse relationship exists between loneliness and wisdom. "In other words, people who have high levels of wisdom didn't feel lonely, and vice versa," he said.

学生答案:

According to the research, more people sink into the most severe loneliness during their late 20s, mid-50s and late 80s. As Dr. Jeste said, there is an inverse relationship exists between loneliness and wisdom. Some people who have higher levels of wisdom didn't feel lonely. This conclusion means that wisdom can protect against loneliness. In my opinion, I really agree with this point of view.

Figure 5. Student's composition

Teacher's comments: The summary should condense each paragraph into one sentence and make logical connections between them, but sentences 1 and 2 lack logical connections. Sentences 2 and 3 contain repeated content. However, ChatGPT directly generates an example (as shown in Figure 6).

The research paper "Loneliness and Wisdom: Are They Related?" explores the relationship between loneliness and wisdom. The study finds that people tend to experience moderate to severe loneliness during their late 20s, mid-50s, and late 80s, with a significant number of participants reporting high levels of loneliness. The authors note that loneliness is subjective, defined as the discrepancy between the social relationships one wants and those they have. However, the study also highlights a potential silver lining, with an inverse relationship existing between loneliness and wisdom. The research suggests that people with high levels of wisdom tend to experience less loneliness, and vice versa.

Figure 6. Corrective Feedback Provided by ChatGPT

5. Conclusions and Implications

5.1 Conclusions

This study takes 277 English as a Second Language (ESL) compositions as the research object, adopting a mixed - methods approach (qualitative and quantitative) to conduct a comparative analysis of the written corrective feedback provided by teachers and ChatGPT. The conclusions are as follows:

Feedback Quantity: Teachers' feedback features "precise focus", mainly correcting core grammar and vocabulary errors in students' compositions, and paying less attention to minor flaws. In contrast, ChatGPT, based on instruction requirements, comprehensively corrects all erroneous expressions and frequently implements vocabulary and structure replacements. Data shows that the proportion of vocabulary and structure replacements by ChatGPT is as high as 56.4%, reflecting its "comprehensive coverage" feedback logic.

Feedback Mode: Leveraging its vast corpus and powerful computing and data-processing capabilities, ChatGPT can quickly retrieve linguistic knowledge, generate coherent texts through open-ended understanding, and achieve comprehensive and non-focused feedback on compositions. It performs particularly well in optimizing written-language style and reformulative transcription. Meanwhile, it can also provide local feedback targeting language aspects. Conversely, teachers tend to adopt indirect feedback and focused local feedback strategies. They provide targeted guidance in a guiding manner based on the explicit language errors in learners' compositions. This difference stems from their distinct operation mechanisms: ChatGPT is driven by algorithms and data and is not restricted by individual differences. However, teachers need to consider the differences in proficiency levels of ESL learners, the diversity of cultural backgrounds,

and the complexity of language, making it difficult to achieve absolute comprehensiveness in feedback content.

Language Difficulty of Feedback: Teachers' feedback strictly follows the "i + 1" comprehensible input principle. It is based on learners' current language proficiency, avoiding complex English expressions. When necessary, annotations in learners' native language are used to reduce comprehension difficulty. In contrast, the language of ChatGPT's feedback often exceeds learners' actual proficiency levels. The frequent reformulative transcription is likely to increase cognitive load and weaken the feedback effect. Although teachers can quickly evaluate the effectiveness of AI-generated feedback, for non-native speakers, the lengthy feedback with mismatched language difficulty has limited practical teaching value.

Emotional Value Empowerment: Teachers' feedback is student-centered. Besides written error correction, it also incorporates positive and encouraging remarks. For example, comments like "Through continuous practice, I believe your English will get better and better" are attached after composition grading. These can effectively boost students' confidence, promote critical thinking, and provide emotional value. In contrast, ChatGPT's feedback is highly standardized and lacks personalized emotional care, making it difficult to replace the humanistic educational value embedded in teachers' feedback.

5.2 Optimization Suggestions

ChatGPT, with its powerful natural language generation capabilities, provides innovative technological support for ESL writing teaching. However, how English teachers can effectively utilize ChatGPT to assist in providing corrective feedback remains a question worth exploring. Based on the conclusions of this study, the following optimization suggestions are proposed:

Teachers can use ChatGPT to quickly generate multi-dimensional modification plans, applying it to aspects such as basic grammar error detection, written-language expression enhancement, word-order and structure adjustment, and discourse cohesion optimization. By automating repetitive grading tasks, teachers can devote more energy to personalized guidance and complex problem analysis, thereby achieving a dual improvement in grading efficiency and quality.

Utilizing the interactive features of ChatGPT, teachers can develop personalized modification suggestions tailored to learners' language proficiency levels. During this process, it is recommended that teachers shift their teaching focus to in-depth explanations of cultural connotations and logical understanding. Meanwhile, combined with learners' actual proficiency levels, they should appropriately simplify vocabulary and syntactic difficulties, helping learners construct expression methods that match their proficiency levels and achieving differentiated teaching goals.

With the text-processing capabilities of ChatGPT, teachers can create a dynamic "electronic error-collection notebook" to systematically integrate scattered feedback data. Through long-term tracking and analysis of this data, teachers can accurately grasp learners' writing characteristics and common error patterns, and then optimize teaching plans, enhancing the pertinence and effectiveness of teaching strategies.

ESL teachers need to continuously improve their ability to use AI tools and deeply understand the crucial impact of instruction design on feedback effects. It is suggested to use contextualized prompts (e.g., "Assume you are a learner with an IELTS score of 5" "Simulate the role of a senior English teacher") to construct clear task scenarios, and replace general task names with detailed descriptive requests, thereby improving the accuracy and effectiveness of ChatGPT's feedback. In addition, considering the prevalence of students' independent use of ChatGPT outside of class, teachers need to systematically explain the advantages and limitations of this tool in written corrective feedback. For example, clearly point out that it lacks emotional perception and life experience, making it difficult to capture personalized expressions and identity characteristics in students' writing, and guide students to use technological tools rationally.

5.3 Limitations and Future Research Directions

While technology-enabled education brings many conveniences, we should be more aware of the limitations of ChatGPT in ESL teaching applications. Although ChatGPT can output more comprehensive and numerous modification suggestions than teachers' feedback, its high-frequency and high-density feedback may exceed learners' cognitive load, leading to information overload. Beyond this, technological defects cannot be ignored: ChatGPT has problems such as error-correction mistakes and over-modification. It often makes unnecessary rewrites of originally correct expressions and even deletes key information by mistake. In terms of identifying logical errors, its performance is particularly weak, making it difficult to accurately judge argumentative loopholes and thinking deviations in texts. This may further cause learners to have confused understanding of English vocabulary and grammatical structures. Moreover, although ChatGPT can provide multiple rounds of feedback, it still requires students to articulate clear instructions or needs, which poses a significant challenge for those with inadequate English proficiency. Therefore, teacher assistance is indispensable.

Additionally, English writing is not only an output of English but also a "transition" of thought and an act of cross-cultural communication. This means that both teachers and students need to jointly participate in the process of writing creation and revision, repeatedly discussing issues such as language use, thought transformation, and cross-cultural negotiation in writing. Another critical concern lies in how ChatGPT's advanced output capabilities pose a significant challenge to writing instruction. Its vast corpus, efficient text generation speed, and human-like language expression have raised concerns among educators and posed numerous challenges to the academic community (Khalil & Er, 2023; Ling, Wang, & Yuan, 2023; Rudolph, Tan, & Tan, 2023) -whether students will excessively rely on AI, or even use ChatGPT to write papers for them, potentially undermining their writing and critical thinking abilities, remains a concern worth addressing. Compounding these issues, ChatGPT is not flawless. For example, it still has issues with logical reasoning, metaphors, and

understanding context (Bang et al., 2023; Le, 2023). When generating academic papers, ChatGPT may output inappropriate remarks or viewpoints and even fabricate references (de Winter, 2023; Kim, Htut, Bowman, & Petty, 2022). In summary, whether ChatGPT can truly improve students' writing abilities (Basic, Banovac, Kruzic, & Jerkovic, 2023; Fyfe, 2023) and how effective its application in writing instruction is require more empirical research for verification.

Notably, This study still has the following deficiencies: First, ChatGPT lacks the ability to accurately assess English proficiency levels and cannot effectively distinguish the composition differences among beginners, intermediate, and advanced learners, making it difficult to meet the needs of differentiated teaching. Due to the limitation of research length, this study has not deeply examined the actual effects and adaptability of different - proficiency ESL learners using ChatGPT. Future research can focus on this, exploring the matching relationship between technological tools and learners' language proficiency through empirical analysis, so as to provide more targeted strategic references for optimizing ESL writing teaching. Second, it has not conducted an in - depth evaluation of the effectiveness of ChatGPT's feedback on different types of errors. And existing research has shown that it has problems such as over - correction and generating illogical or inaccurate content (Wei et al., 2022). Third, the prompt strategy adopted in this study is only one of many possibilities. In the future, more diversified prompt designs and application scenarios can be explored to further optimize its performance. Fourth, there is a lack of longitudinal data. This study has not explored the long - term impact of ChatGPT's feedback on students' writing skills. Longitudinal research can deeply reveal the effectiveness of AI - assisted feedback.

Acknowledgments

We greatly appreciate the valuable contributions of our research team members, as well as the participants in this study [Project Number: 2025GYB132]& [Project Number: ZN-D-042]&[Project No.: JXJG-23-40-2]. We sincerely thank the Jiangxi Provincial Education Science Planning Office, Jiangxi Provincial Association of Higher Education, and the authority in charge of Jiangxi Provincial Higher Education Teaching Reform Projects. Their funding for this study provided essential support for the study's design, AI-assisted feedback analysis, and data collection.

AI Acknowledgments

The authors acknowledge the use of ChatGPT 4.0 as a tool for analyzing research subjects. The prompts employed include: "Correct basic language issues such as grammar, sentence patterns, punctuation, and spelling"; "Polish this article to make it more fluent and coherent"; and "Summarize this article." Outputs generated from these prompts were used to assist in analyzing the research subjects. While acknowledging the use of artificial intelligence, the authors confirm that they are the sole authors of this article and assume full responsibility for its content, in accordance with the recommendations of the Committee on Publication Ethics (COPE).

Authors' contributions

Dr. Jianmei Yang designed the study, collected data, drafted the initial manuscript, and revised the final version. Caisang Hou and Xuanxiao Lu, as co-first authors, participated in data collation, analysis of feedback differences between teachers and ChatGPT, and contributed to manuscript refinement. Yong Wang, the corresponding author, supervised the study design, verified data reliability, and provided critical revisions to the manuscript. All authors reviewed, commented on, and approved the final manuscript.

Funding

This work was funded by three projects: (1) the General Project of Jiangxi Provincial Education Science Planning, entitled Exploration on the High-Quality Development Path of Foreign Language Teaching Under Digital Transformation [Project No.: 2025GYB132]; (2) the Special Project of Jiangxi Provincial Association of Higher Education, entitled A Study on Artificial Intelligence-Assisted Feedback in College English Writing - A Case Study of ChatGPT [Project No.: ZN-D-042]; and (3) the Teaching Reform Project in Higher Education Institutions of Jiangxi Province, entitled Construction and Application of the "Three-Stage and Nine-Step" Teaching Model for ESP Empowered by Digital Transformation [Project No.: JXJG-23-40-2]. We would like to express our sincere gratitude to the above-mentioned institutions for their support.

Competing interests

The authors declare no competing interests.

Informed consent

The authors have obtained informed consent from all participants.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available

due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Abedi, R., Latifi, M., & Moinzadeh, A. (2010). The Effect of Error Correction vs. Error Detection on Iranian Pre-Intermediate EFL Learners' Writing Achievement. *English Language Teaching*, 3(4), 168-174. <https://doi.org/10.5539/elt.v3n4p168>
- Atlas, S. (2023). ChatGPT for higher education and professional development: A guide to conversational AI.
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... Chung, W. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*. <https://doi.org/10.18653/v1/2023.ijcnlp-main.45>
- Basic, Z., Banovac, A., Kruzic, I., & Jerkovic, I. (2023). Better by you, better than me, chatgpt3 as writing assistance in students essays. *arXiv preprint arXiv:2302.04536*. <https://doi.org/10.35542/osf.io/n5m7s>
- Benson, S., & DeKeyser, R. (2019). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, 23(6), 702-726. <https://doi.org/10.1177/1362168818770921>
- Beuningen, C. V., Jong, N. d., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL-International Journal of Applied Linguistics*, 156(1), 279-296. <https://doi.org/10.2143/ITL.156.0.2034439>
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language Writing*, 17(2), 102-118. <https://doi.org/10.1016/j.jslw.2007.11.004>
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322-329. <https://doi.org/10.1016/j.system.2008.12.006>
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of second language Writing*, 12(3), 267-296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- de Winter, J. C. (2023). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 1-16. <https://doi.org/10.1007/s40593-023-00372-z>
- Ellis, R. (2009). A typology of written corrective feedback types. *ELT journal*, 63(2), 97-107. <https://doi.org/10.1093/elt/ccn023>
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353-371. <https://doi.org/10.1016/j.system.2008.02.001>
- Fathman, A. (1990). Teacher response to student writing: Focus on form versus content. *Second language The Long-Term Effectiveness of Written Corrective Feedback in an EFL Writing Class*, 29. <https://doi.org/10.1017/CBO9781139524551.016>
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of second language Writing*, 10(3), 161-184. [https://doi.org/10.1016/S1060-3743\(01\)00039-X](https://doi.org/10.1016/S1060-3743(01)00039-X)
- Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI & SOCIETY*, 38(4), 1395-1405. <https://doi.org/10.1007/s00146-022-01397-z>
- 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(7), 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Hou, Y., Dong, H., Wang, X., Li, B., & Che, W. (2022). MetaPrompting: Learning to learn better prompts. *arXiv preprint arXiv:2209.11486*.
- Hyland, K. (2013). Student perceptions of hidden messages in teacher written feedback. *Studies in Educational Evaluation*, 39(3), 180-187. <https://doi.org/10.1016/j.stueduc.2013.06.003>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, 39(2), 83-101. <https://doi.org/10.1017/S0261444806003399>
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The modern language journal*, 75(3), 305-313. <https://doi.org/10.1111/j.1540-4781.1991.tb05359.x>
- Khalil, M., & Er, E. (2023). Will ChatGPT G et You Caught? Rethinking of Plagiarism Detection. Paper presented at the International Conference on Human-Computer Interaction. <https://doi.org/10.35542/osf.io/fnh48>

- Kim, N., Htut, P. M., Bowman, S. R., & Petty, J. (2022). $\$^{\wedge} 2\$: Question Answering with Questionable Assumptions. *arXiv preprint arXiv:2212.10003*.$
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... Maningo, J. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS digital health, 2*(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Lalande, J. F. (1982). Reducing composition errors: An experiment. *The modern language journal, 66*(2), 140-149. <https://doi.org/10.1111/j.1540-4781.1982.tb06973.x>
- Le, T. T. (2023). A preliminary example of utilizing AI text generation to support academic writing “Humans befriending their creations some notes on the human-AI relationship”. OSF Preprints. *OSF Preprints*. <https://doi.org/10.31219/osf.io/bsxy>
- Lee, I., Luo, N., & Mak, P. (2023). Issues of error selection for focused written corrective feedback in authentic classroom contexts. *RELC Journal, 54*(3), 616-629. <https://doi.org/10.1177/00336882211028425>
- Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., ... Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence, 4*, 100140. <https://doi.org/10.1016/j.caeai.2023.100140>
- Ling, X., Wang, D., & Yuan, J. (2023). Reflection on Technology Ethics and Academic Ethics in the Context of ChatGPT Craze [J]. *Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences), 44*(04), 123-136.
- Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., & Chen, W. (2021). What Makes Good In-Context Examples for GPT-3? *arXiv preprint arXiv:2101.06804*. <https://doi.org/10.18653/v1/2022.deeLIO-1.10>
- Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*. <https://doi.org/10.1016/B978-012589042-7/50015-3>
- Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. *Computers and Education: Artificial Intelligence, 7*, 100278. <https://doi.org/10.1016/j.caeai.2024.100278>
- Mahfoodh, O. H. A. (2017). “I feel disappointed”: EFL university students’ emotional responses towards teacher written feedback. *Assessing Writing, 31*, 53-72. <https://doi.org/10.1016/j.asw.2016.07.001>
- Reynolds, L., & McDonell, K. (2021). *Prompt programming for large language models: Beyond the few-shot paradigm*. Paper presented at the Extended abstracts of the 2021 CHI conference on human factors in computing systems. <https://doi.org/10.1145/3411763.3451760>
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL quarterly, 20*(1), 83-96. <https://doi.org/10.2307/3586390>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching, 6*(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Santos, M., Serrano, S. L., & Manchán, R. M. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies, 10*(1), 131-154. <https://doi.org/10.6018/ijes/2010/1/114011>
- Semke, H. D. (1984). Effects of the red pen. *Foreign language annals, 17*(3), 195-202. <https://doi.org/10.1111/j.1944-9720.1984.tb01727.x>
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL quarterly, 41*(2), 255-283. <https://doi.org/10.1002/j.1545-7249.2007.tb00059.x>
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2022). Prompting gpt-3 to be reliable. *arXiv preprint arXiv:2210.09150*.
- Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System, 81*, 135-145. <https://doi.org/10.1016/j.system.2018.12.017>
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning, 46*(2), 327-369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
- Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately. *Journal of second language Writing, 16*(4), 255-272. <https://doi.org/10.1016/j.jslw.2007.06.003>
- Truscott, J., & Hsu, A. Y. P. (2008). Error correction, revision, and learning. *Journal of second language Writing, 17*(4), 292-305. <https://doi.org/10.1016/j.jslw.2008.05.003>
- Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language learning, 62*(1), 1-41. <https://doi.org/10.1111/j.1467-9922.2011.00674.x>
- Wang, Z., Valdez, J., Basu Mallick, D., & Baraniuk, R. G. (2022). *Towards human-like educational question generation with large language models*. Paper presented at the International conference on artificial intelligence in education. https://doi.org/10.1007/978-3-031-11644-5_13

- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... Metzler, D. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.
- Yoon, S. Y., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion. *arXiv preprint arXiv:2310.06505*.
- Yu, S., Zheng, Y., Jiang, L., Liu, C., & Xu, Y. (2021). "I even feel annoyed and angry": Teacher emotional experiences in giving feedback on student writing. *Assessing Writing*, 48, 100528.

ⁱ Note1: ChatGPT's advanced output capabilities also include opinion output and essay examples. For details, please refer to Atlas, S. (2023). ChatGPT for higher education and professional development: A guide to conversational AI. Due to space limitations, this study will no longer present examples of ChatGPT outputting opinions and writing essays based on specified topics.

ⁱⁱ Note 2: For details, please see the Oxford Advanced Learner's English-Chinese Dictionary (9th Edition), page 1723, entry 5 of "pure".