IRA-International Journal of Education & Multidisciplinary Studies
ISSN 2455–2526; Vol.21, Issue 02 (Q2, 2025)
Pg. no. 41-55.
IRA Academico Research

The Impacts of Multiple Feedbacks (ChatGPT, Peer and Teacher) on English Composition Revision

Jingjing Chen
School of Foreign Languages, Zhejiang University of Finance and Economics, Hangzhou, China.

Type of Work: Peer Reviewed.
DOI: https://dx.doi.org/10.21013/jems.v21.n2.p1

Review history: Submitted: March 06, 2025; Revised: March 31, 2025; Accepted: April 14, 2025

How to cite this paper:
Chen, J. (2025). The Impacts of Multiple Feedbacks (ChatGPT, Peer and Teacher) on English Composition Revision. IRA-*International Journal of Education & Multidisciplinary Studies* (ISSN 2455-2526), 21(2), 41-55. https://dx.doi.org/10.21013/jems.v21.n2.p1

Jingjing Chen /0009-0009-6017-6028

ABSTRACT

This study aims to explore the impacts of ChatGPT, peer, and teacher feedback on English writing revisions. Sixty-seven junior English majors participated in six rounds of writing and feedback activities throughout an academic year. The results reveal that in terms of the number of revisions, ChatGPT and teacher feedback accounted for 85.1%, while peer feedback accounted for 14.9%. For the types of revisions, ChatGPT dominated surface-level revisions, and teacher and peer feedback were mainly responsible for meaning-level revisions. In terms of the effectiveness of revisions, ChatGPT had the highest success rate (82%), and peer feedback had the lowest (56.1%). The study demonstrates that multiple feedback has a positive effect. Teachers should leverage the advantages of each feedback method, such as guiding students to use ChatGPT, enhancing the effectiveness of peer feedback, and strengthening the guiding role of teacher feedback. However, this study lacks a control group. Future research can conduct comparative studies between experimental and control groups to further explore.

Keyword: ChatGPT feedback; Peer feedback; Teacher feedback; English writing revision

1. Introduction

Offering students formative feedback throughout the writing process represents a crucial educational approach that significantly aids students in enhancing their writing capabilities (Graham et al, 2011). By explicitly conveying to students the characteristics of high-quality writing and the means to achieve it, formative feedback steers students towards constructive actions or the enhancement of specific writing skills (Graham et al, 2016). Nevertheless, the substantial time and energy required to furnish feedback to students, particularly to numerous students across multiple classes, pose a formidable challenge for many educators. This often discourages some teachers from delivering essential writing instruction (Applebee & Langer, 2011). Consequently, alleviating the burden on teachers as the sole providers of feedback may create more opportunities for writing and writing instruction. Teachers can utilize various strategies for providing feedback, including direct or indirect corrective feedback (CF), as well as focused or unfocused approaches (Ellis, 2009). However, CF is a multifaceted issue influenced by numerous factors that affect its effectiveness.

Automated Writing Evaluation (AWE) has been a subject of research for years as a method to provide prompt assessment of student writing and reduce the workload of educators in evaluating writing. These systems typically employ natural language processing and artificial intelligence to evaluate writing. Some studies have indicated that such systems can have positive impacts on student engagement, efficacy, writing length, and quality (Graham et al, 2015; Stevenson & Phakiti, 2014; Wilson & MacArthur, 2024;). Although some AWE systems are as efficient and reliable as human raters when it comes to scoring writing, they are generally less accurate, more general and wordier, and sometimes perplexing to those receiving the feedback (Wilson & MacArthur, 2024). Moreover, preparing these tools for educational settings demands a significant amount of time (Chen et al., 2022; Moore & MacArthur, 2016;). Historically, AWE systems have necessitated training on hundreds of essays written in response to the same prompt and iterative calibration with human-provided feedback. These requirements not only increase their cost but also limit teachers' flexibility in using these systems, as the evaluation is restricted to the types of writing prompts used for training.

However, new-generation generative AI, such as ChatGPT, functions differently from previous versions of AWE software and older AI systems. The advent of artificial intelligence (AI) has the potential to revolutionize traditional educational paradigms. ChatGPT, which was released in November 2022, has not only captured the public's interest but also ignited extensive discussions within the higher-education academic sphere. As a powerful language model, ChatGPT can automatically generate natural sounding texts when users simply key in appropriate prompts and make writing a piece of essay easier than ever. At present, it is cost-effective and easily accessible. It is possible that new generative AI like ChatGPT can offer feedback that is timely, targeted, adaptable, and beneficial, all of which can help students improve their writing.

2. Literature Review on Related Studies

2.1 Previous Studies on Traditional AWE Feedback

Automated Writing Evaluation (AWE) systems were first developed in the United States during the 1960s as tools for scoring large-scale examinations. Over the following decades, these systems expanded their functionality and application, becoming integral to formative assessments in educational settings. Prominent AWE systems, such as My Access!, Writing Roadmap, and Criterion, now provide features beyond scoring, including written feedback, writing resource repositories, teacher assessment tools, and management of learning processes. The adoption of AWE systems in classrooms is grounded in formative assessment theory, which emphasizes the role of feedback in monitoring and improving student learning. This approach, often described as "assessment for learning" (Black & Wiliam, 1998), aims to enhance both student learning and teaching quality by using assessment results to guide instructional adjustments.

The effectiveness of AWE systems can be assessed from two perspectives: the quality of feedback they provide and their influence on the writing process. Studies comparing AWE feedback with teacher feedback highlight significant differences in quality. Research by Dikli (2010) on My Access! revealed that while the system offers detailed feedback on grammar, technical issues, and conventions, it often repeats the same explanations for similar errors and provides impractical, verbose suggestions. In contrast, teacher feedback is concise, specific, and targeted, making it more actionable for learners. Similarly, research on the Pigai system (Shi, 2012) found that it provides fewer actionable insights on content and structure compared to teachers, who offer more intuitive and explicit feedback. These findings suggest that AWE systems, while efficient in generating feedback, fall short of delivering personalized and nuanced suggestions tailored to individual learner needs. The extent to which students engage with and utilize AWE feedback varies widely. Research shows that a significant portion of feedback provided by AWE systems is ignored or results in superficial revisions. For example, Chodorow et al. (2010) observed that only 32% of students implemented Criterion feedback, with the majority making changes at the word or phrase level, deleting errors, or skipping feedback altogether. Similarly, Chapelle et al (2015) reported that students adopted only 49% of Criterion feedback, with most changes focusing on surface-level corrections rather than substantive revisions.

Adoption rates for AWE feedback vary depending on the type of error and the system in use. For example, Lavolette et al. (2015) found higher modification rates for grammar-related errors like subject-verb agreement and verb forms compared to issues like preposition use or punctuation. In China, adoption rates for Pigai feedback ranged from as high as 80% (Lu, 2016) to as low as 11.5% (Bai & Hu, 2017). The discrepancy is largely due to whether synonymous word suggestions, which account for 70% of feedback but have adoption rates below 3%, are included in the analysis (Huang & Zhang, 2018). Despite these variations, both domestic and international studies indicate that most student revisions focus on linguistic forms (83%), with limited attention given to meaning-based changes involving content and structure (17%) (Wu & Zhang, 2016).

The primary limitations of AWE systems lie in their narrow focus and mechanical approach to feedback. First, they evaluate writing primarily based on linguistic features, offering overly general suggestions on content, organization, and tone. This lack of depth reduces the usefulness of feedback for improving overall writing quality (Li et al., 2015; Li, 2015). Second, AWE systems tend to prioritize formal aspects, such as vocabulary complexity, text length, and conjunction use, leading to formulaic feedback that may include inaccuracies (Chen & Cheng, 2008; Yang, 2013). While AWE systems offer the potential for scalable and efficient feedback, their current limitations highlight the need for further refinement to fully support student learning and writing development. As technology evolves, combining automated and human feedback could provide a more holistic approach to writing instruction, balancing efficiency with pedagogical depth.

The traditional AWE system can greatly reduce the feedback pressure of teachers, but there are also problems such as weak pertinence and insufficient personalized feedback. Since 2022, the application of generative AI in the field of text generation has provided new possibilities for addressing the challenges faced by traditional AWE systems.

2.2 Previous Studies on Large Language Model

Deep learning has become one of the techniques for enhancing the precision and efficacy of AWE. Methods of AES grounded in deep learning make use of artificial neural networks. These neural networks imitate the operation of the human brain via layered algorithms and computational units. Differing from traditional machine learning, deep learning is capable of learning independently from the surroundings and past mistakes without human involvement. This characteristic allows deep-learning models to form nonlinear relationships, thereby achieving higher accuracy. The recent progress in deep learning has given rise to the creation of transformers. These are especially efficient in learning text representations. Prominent instances are Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) and the Generative Pretrained Transformer (GPT) (OpenAI).

Language model-based chatbots, such as ChatGPT, represent a form of generative AI capable of creating new content based on the data they were trained on. These large language models (LLMs) process text using “tokens,” which can range from individual letters and combinations of letters to complete words. The outputs generated by these chatbots are essentially predictions derived from their training data, making their responses heavily dependent on that information (Kim et al., 2023). Additionally, LLMs have a set limit on the number of tokens they can process simultaneously, which impacts the scope of questions they can handle. ChatGPT, developed by

OpenAI, owes its ability to deliver human-like responses across diverse prompts to the extensive dataset on which it was trained (Barrot, 2023). However, OpenAI has not disclosed specific details about the composition of this dataset.

Barrot (2023) suggests that tools like ChatGPT can alleviate the workload of L2 writing teachers by providing feedback that supports student learning. The chatbot is noted for its ability to deliver personalized and timely feedback, a feature also emphasized by Falk (2023). Despite these advantages, Barrot (2023) acknowledges limitations in ChatGPT's performance, particularly when addressing topics requiring deep understanding or higher-order thinking. Pfau et al. (2023) highlight that ChatGPT's primary function is not to correct linguistic errors, contrasting it with tools like Grammarly, which relies on prescriptive grammar rules rather than machine learning. Unlike Grammarly, ChatGPT generates responses based on probability calculations, allowing for variability in its answers to the same prompt. The quality of its responses depends significantly on the clarity and specificity of the prompts provided.

How to integrate automated assessment systems with manual feedback to construct a multifaceted feedback mechanism that leverages the strengths of peer and teacher feedback while addressing the limitations of automated writing evaluation systems has become a key research focus. Tang (2014) explored a pioneering model of writing instruction based on multidimensional feedback using the WRM automated evaluation system. The results showed that this model enhanced students' writing skills and positively influenced the teaching process. However, the study primarily examined the effects of the multidimensional feedback model on students' writing abilities without specifically investigating its impact on students' revisions. Additionally, the study focused on high school students, leaving the applicability and effectiveness of this approach for other populations uncertain. Huang and Zhang (2014) investigated the effects of feedback from multiple sources, including Pigai.net, peer feedback, and teacher feedback, on students' revision processes and writing behavior. They found that such multifaceted feedback could improve students' writing practices, enhance their agency, and boost text quality. While these studies explored the impact of multifaceted feedback on revision types, functions, and text quality, they did not specifically examine its effects on the outcomes of students' revisions. Other research has indicated that feedback delivered via computer-mediated platforms can be more effective in facilitating language acquisition than feedback provided by teachers (Wang et al., 2018). As a teaching and assessment tool, multifaceted feedback requires further theoretical and practical validation and broader empirical testing to evaluate its effectiveness.

Based on this, the present study employs data collected from natural classroom settings and leverages ChatGPT's powerful interactive capabilities to explore how ChatGPT-based multifaceted feedback affects students' revision quantity, revision types, and revision outcomes in English writing. This study aims to provide empirical evidence to enrich research on writing feedback and expand its scope while offering a new operational model for English writing instruction.

3. Research Methods

3.1 Research Participants

This study was carried out with 67 third-year undergraduate students majoring in English. These students were from two natural classes within an academic English writing course at a comprehensive university located in southern China. At the time of the study, the participants were in their sixth semester of undergraduate study, standing at the threshold of commencing their thesis writing journey. Among these students, the gender distribution was as follows: 14 were male, while 53 were female. In terms of their English proficiency, they had demonstrated remarkable performance. A high 91.2% of them passed the national Test for English Majors-Band 4 (TEM-4), which is a significant benchmark for English majors in China. Moreover, an impressive 63.7% of the students achieved excellent scores in this test. According to the China Standards of English Proficiency (CSE), their English proficiency levels were scattered between levels 5 and 6, indicating a relatively solid and intermediate-to-advanced level of English language command.

3.2 Feedback Procedures

The study was meticulously designed to span an entire academic year, commencing in October 2023 and concluding in June 2024. Throughout this period, students actively engaged in a series of six rounds of writing and feedback activities, which were seamlessly integrated into their regular English writing instruction. These activities were carefully structured to enhance students' writing skills through a multi-faceted feedback approach. The feedback procedures were composed of three distinct and sequential stages:

1) First, once students had completed their initial drafts, they were required to upload them to a shared document. Two graduate assistants then took on the task of compiling all the drafts. They interacted intensively with ChatGPT to obtain comprehensive feedback for each draft. Students were given a set timeframe within which they could revise their drafts based on ChatGPT's feedback, and they had the flexibility to submit multiple revisions.

2) Peer Feedback: The instructor carefully divided the students into groups. In these groups, they exchanged and thoroughly reviewed each other's revised drafts, which already incorporated ChatGPT feedback. After peer evaluations, students were expected to make further revisions and submit their work again. During this phase, open communication and collaboration were highly encouraged among students to foster a more in-depth understanding of writing improvements.

3) Teacher Feedback: The instructor provided detailed and in-depth feedback on the drafts that had been revised based on peer evaluations. Students were then tasked with making further revisions in response to the teacher's feedback and submitting their final versions within the allotted time, thus completing the cycle of writing improvement.

3.3 Feedback Ways

In this study, the feedback methods implemented were comprehensively divided into direct and indirect feedback, each playing a distinct and crucial role in the learning process. Direct feedback was a straightforward and immediate approach. When applying this method, the feedback providers would clearly identify the errors present in the students' work and then directly offer the correct forms or specific solutions. This allowed students to quickly and efficiently address the issues at hand. For example, if a student made a grammar error in a sentence, direct feedback would point out the exact error and present the corrected sentence structure. This type of feedback was particularly useful for students who needed clear-cut guidance to rectify their mistakes. On the other hand, indirect feedback took a more subtle and exploratory approach. Instead of directly providing solutions, it offered cues or information in the form of queries or hints. For instance, it might ask a question like “Is the verb tense consistent in this paragraph?” or give a hint such as “There seems to be an issue with the logical connection here.” This approach was designed to stimulate students’ critical thinking, encouraging them to independently identify and resolve problems. It promoted a deeper understanding of the writing process as students had to actively engage with their work to figure out the errors and find appropriate solutions

3.4 Data Collection and Analysis

The study employed two primary and complementary methods for data collection to ensure a comprehensive and accurate understanding of the research topic.: 1) To effectively manage the sample size and the volume of data, a strategic selection was made. Three writing tasks were carefully chosen for in-depth textual analysis. All the relevant data were extracted from the shared document, which served as a detailed record-keeper. It logged all the drafts and revisions made by the students during the three tasks. This in-depth analysis was aimed at closely examining the effects of multifaceted feedback on various aspects of the writing process, including the quantity of revisions, the different types of revisions (such as surface-level and meaning-level), and the overall quality of the final written work. 2) Interviews: Prior to the study, students were categorized into three distinct performance groups: high, medium, and low, based on their writing scores. From each of these proficiency levels, three students were purposefully selected for in-depth interviews. These interviews were not only a means to further validate the findings from the textual analysis but also a way to gain unique insights into students’ revision processes and their experiences with the feedback they received. Through these interviews, researchers could understand the thought-processes behind students’ decisions during revisions, as well as their perceptions of the different feedback methods, which provided a more holistic view of the study’s subject matter.

4. Research Results and Discussion

4.1 The Impact of Multiple Feedback on the Number of Composition Revisions

Through text comparison, it was found that ChatGPT, peers, and teachers provided a total of 3,000 feedback points for three writing tasks. Students revised their compositions based on 2,319 of these feedback points (see Table 1). Among them, 45.6% of the revisions were based on ChatGPT feedback, 14.9% were based on peer feedback, and 39.5% were based on teacher feedback. The proportions of the three feedback methods in the number of revised and unrevised items are slightly different (see Table 2). Students made the most revisions based on ChatGPT feedback, with a total of 1,057 revisions, accounting for 87%, and the least number of unrevised items, accounting for 13%. The number of revisions made by students based on teacher feedback was 915, accounting for 74%, and the number of unrevised items was 321, accounting for 26%. In contrast, the number of revisions made by students based on peer feedback was the least, with a total of 347 revisions, accounting for 63.2%, and the number of unrevised items was 202, with the highest proportion.

Table 1. Classification Statistics of the Impact of Multiple Feedback on the Number of Composition Revisions (1)

	ChatGPT Feedback	Peer Feedback	Teacher Feedback	Total
Quantity of Revisions	1057	347	915	2319
Proportion（%）	45.6%	14.9%	39.5%	100%

Table 2. Classification Statistics of the Impact of Multiple Feedback on the Number of Composition Revisions (2)

Feedback Source	Feedback Points (%)	Revised Items (%)	Unrevised Items (%)
ChatGPT	1215(100)	1057(87%)	158(13%)
Peer	549(100)	347(63.2%)	202(36.8%)
Teacher	1236(100)	915(74%)	321(26%)

From the above data, first of all, in terms of the total amount of revisions, the proportion of revisions based on ChatGPT feedback and teacher feedback is 85.1%, while the proportion of revisions based on peer feedback is relatively low, accounting for only 14.9%. Secondly, regarding the number of revised and unrevised items, the number of revised items for all three feedback methods is over 60%. However, comparatively, the number of unrevised items in peer feedback is the highest, accounting for 36.8%. This result indicates that all three feedback methods can be accepted and adopted by students, but students' acceptance of peer feedback is relatively low. The interview results show that the main reason for students' low acceptance of peer feedback is their lack of confidence in their peers' language proficiency and evaluation ability, which is consistent with some existing research findings on peer feedback (Yang et al., 2006). Nevertheless, students still believe that peer feedback can play a positive role in helping with composition revisions. In particular, peer feedback can increase interaction among classmates, expand writing ideas, and they sometimes feel that the gains as reviewers are even greater than those as feedback recipients. We think that students have such a feeling because peer feedback has stronger sociality and interactivity. This process of transforming from the evaluated object to the evaluation subject is conducive to stimulating learning interest, enhancing learning motivation, and increasing writing confidence. It can be seen that although the acceptance of peer feedback is relatively low, it has its necessity and rationality in the multi-feedback mechanism. However, how to better exert the role of peer feedback still requires further in-depth research and practice.

4.2 The Impact of Multiple Feedback on the Types of Composition Revisions

In this study, the types of revisions are classified into two major categories: surface-level revisions and meaning-level revisions. Surface-level revisions involve formal modifications while retaining the meaning, and meaning-level revisions involve micro-structure and macro-structure aspects.

ChatGPT is the main source of surface-level revisions. Specifically, in terms of formal modifications, the proportion of revision points based on ChatGPT feedback is large, that based on teacher feedback is small, and that based on peer feedback is the smallest. The proportion of meaning-level revision mainly concentrated in the micro-structure, with teacher feedback being the dominant source followed by peer feedback and ChatGPT feedback. The amount of macro-structure revisions is relatively small, only accounting for 1% of the total meaning-level revisions, indicating that students can generally correctly understand and grasp the composition topics and content, and can basically complete the writing tasks as required. Overall, surface-level revisions are mainly based on ChatGPT feedback, while meaning-level revisions are mainly based on teacher and peer feedback. Based on this, the following three conclusions can be drawn:

First, ChatGPT mainly provides "sentence-by-sentence comments" on students' compositions. Its feedback mainly focuses on the language and text level, including spelling, vocabulary, collocations, sentences, and Chinglish. However, for the content, structure, and rhetoric of the article, it mainly gives summary comments, which are general and formatted, such as "There are too few cohesive elements, and the article structure is not good." It lacks personalized and targeted guidance. Therefore, students mainly make revisions at the levels of spelling, punctuation, vocabulary, and grammar based on ChatGPT's feedback. However, since ChatGPT sometimes fails to identify some serious grammar errors, these types of errors are mainly corrected through peer feedback and teacher feedback.

Second, the total amount of peer feedback is the least, but it covers both surface-level and meaning-level aspects. The research subjects in this study tend to comment more on the content and structure of the article during the evaluation. However, they can also provide relatively pertinent feedback on the language. For example, Student A commented on the sentence "Changement of students' attitude from schools to colleges has been confirmed." in a peer's composition, saying, "It should be 'change' instead of 'changement'." Through a comparative analysis of 6 students' compositions, it was found that students' feedback ability is gradually increasing. When providing feedback, they will actively look up dictionaries to give revision suggestions. For instance, Student B commented on the sentence "proficient in English when they come for college education" in a peer's composition, saying, "I'm not sure if 'come for' is correct for entering college, but I think it could be 'proficient in English when they enter college'." This indicates that students have a certain evaluation ability, and teachers should encourage students to actively participate in the feedback activities.

Third, since ChatGPT and peer feedback have already provided corresponding feedback on students' composition spelling, grammar, and vocabulary, teachers only need to provide feedback on the more complex grammatical structures and semantics that ChatGPT and peer feedback cannot address at the language level. There is no need to spend energy on low-level language errors such as spelling, tense, and voice. Instead, teachers can devote more energy to meaning-level feedback, that is, to provide feedback on issues such as whether the paragraphs and sentences in students' compositions are coherent, whether each paragraph has a topic sentence, and whether the details support the argument. Therefore, students mainly make revisions in terms of structure and content based on teacher feedback.

The above-mentioned analysis results show that the three feedback methods can complement each other in helping students revise their compositions, meeting the different needs of different students for composition revisions. However, overall, the total amount of language-level feedback is large, indicating that grammar and vocabulary remain a difficult point in students' writing, and are an important bottleneck restricting the improvement of students' writing level. In future teaching, more training in language knowledge should be strengthened to further consolidate students' language foundation.

4.3 The Impact of Multiple Feedback on the Types of Composition Revisions

In this study, the effectiveness of composition revision is classified into three categories: successful revision, unsuccessful revision, and no revision. A successful revision refers to resolving or improving the issues pointed out in the feedback. An unsuccessful revision means failing to improve the original text or making it even worse. The specific statistical results are shown in Table 3:

Table 3. Classification Statistics of the Impact of Multiple Feedback on the Effectiveness of Students' Composition Revision

Feedback Source	Feedback Points (%)	Successful Revision (%)	Unsuccessful Revision (%)	No Revision (%)
ChatGPT Feedback	1215(100)	997(82)	52(4.3)	166(13.7)
Peer Feedback	549(100)	308(56.1)	31(5.6)	210(38.3)
Teacher Feedback	1236(100)	822(66.5)	85(6.9)	329(26.6)
Total	3000(100)	2127(70.9)	168(5.6)	705(23.5)

As depicted in Table 3, the successful-revision rate of ChatGPT feedback was the highest (82%), followed by teacher feedback (66.5%), and then peer feedback (56.1%). In terms of the non-revision rate, peer feedback exhibited the highest proportion (38.3%), succeeded by teacher feedback (26.6%), while ChatGPT feedback had the lowest non-revision rate (13.7%). Through textual analysis and interviews, the following findings were obtained:

Firstly, the high successful-revision rate of ChatGPT feedback can be attributed to its focus on surface-level errors in compositions. It either rectifies errors directly or identifies the nature of the errors and provides clear and straightforward revision suggestions. Under normal circumstances, students can effectively revise errors based on these suggestions. The relatively small proportion of unsuccessful revisions was primarily due to students' insufficient language knowledge. The main reason for the non-adoption of some revision suggestions was that students deemed the suggestions provided by ChatGPT to be incorrect. It was observed that, due to its mechanical nature, ChatGPT occasionally made errors in correction, such as "quantitative and qualitative methods" should be "qualitative and quantitative methods". This further validates the necessity of incorporating human feedback when utilizing an automated writing evaluation system.

Secondly, peer feedback demonstrated the lowest successful-revision rate and the highest non-revision rate. The reason lies in that for content in which they were confident, peers typically provided the correct form directly or identified the errors and offered specific revision suggestions, resulting in a relatively high successful-revision rate for this part. However, for content they were uncertain about, in most cases, they merely pointed out the existence of errors without clearly defining the nature of the errors or providing corresponding solutions. Their expressions often included ambiguous terms such as "it might be better...", "it could be...", "I feel that...", etc. For example, as previously mentioned, "I'm not certain whether 'come for' is appropriate for 'entering college', but I think it could be 'proficient in English when they enter college'." Moreover, the comments on content and structure were rather general. As stated in an interview, "Peers are unable to provide clear suggestions as teachers do. For instance, if they claim the beginning is not good, they should explain precisely what is wrong and what should be emphasized." Consequently, students either failed to revise successfully or disregarded the revision suggestions.

Teacher feedback predominantly employed the approach of indirect feedback. Similar to peer feedback, the more explicitly and comprehensively teachers identified errors and provided suggestions, the higher the likelihood of students' successful revisions. Nevertheless, as indicated in Table 3, 6.9% of teacher-feedback revisions were unsuccessful, and nearly 26.6% of the feedback was not adopted. Through interviews, several factors were identified: Firstly, language-level revisions necessitate a solid language foundation. Although students acknowledged feedback such as "inappropriate word usage", "improper collocations", or "syntactic errors in this sentence", due to their limited language proficiency and a lack of perseverance in overcoming difficulties, they often either failed to revise successfully or abandoned the revision. As one student remarked, "Teacher feedback is relatively comprehensive, yet sometimes, even after the teacher's revision, my article still contains numerous issues. Due to my own limitations and the absence of mandatory requirements for composition revisions from the teacher, I made few revisions to the problems pointed out by the teacher." Secondly, meaning-level revisions demand that learners possess a high level of logical thinking ability to engage in higher-order thinking regarding the issues. Sometimes, even though students recognized the problems, constraints in text-processing skills, logical reasoning, and language-expression capabilities led to suboptimal final revision effects or caused them to abandon the revision due to not knowing where to commence (Wu, 2015; Yan, 2011). Thirdly, there were discrepancies in the perception of "revision suggestions" between students and teachers. For example, students disagreed with feedback such as "lacking two research implications" or "the second paragraph lacks elaboration of the argument", and thus were reluctant to revise. This indicates that when confronted with fundamental issues such as content and structure in the article, students sometimes do not readily concede and revise, demonstrating a strong "ownership of text" awareness. However, according to Tsui & Ng (2000), the enhancement of this "ownership of text" awareness can assist students in reducing their reliance on teachers and becoming self-assured writers. Teachers should refrain from arbitrarily demanding that students revise the article in strict accordance with their wishes. Instead, they should focus on strengthening communication and interaction with students to bridge the gap in understanding and improve the effectiveness of revisions.

Evidently, the feedback method is a crucial factor influencing the effectiveness of students' composition revisions. Furthermore, the higher the degree of explicitness (the corrective information inherent in the feedback) (Zhu & Wang, 2005), the greater the probability of students' successful revisions. Nevertheless, whether students can ultimately revise successfully based on the feedback is also contingent upon factors such as their language proficiency, cognitive abilities, and the effort they invest.

5. Conclusions

Based on the above-mentioned analysis results, we have identified that multiple feedback exerts a positive influence on the revision of English compositions. Within the multiple-feedback mechanism, ChatGPT feedback, peer feedback, and teacher feedback each possess distinct functions and roles in assisting students with composition revisions. Teachers should, in their teaching, pay attention to leveraging the characteristics and advantages of different feedback methods to enhance the effectiveness and quality of students' composition revisions using these various approaches. To this end, this study proposes the following suggestions from the perspectives of facilitating learning, promoting learning, and guiding learning:

Leveraging ChatGPT's evaluative and learning-facilitating role to enhance students' autonomous writing and revision abilities. ChatGPT not only has the capabilities to score and provide feedback but also serves as a learning-facilitating tool. Endowed with a vast corpus of language data, ChatGPT is a comprehensive learning repository integrating bilingual dictionaries, collocation dictionaries, and thesauruses. Students can interact with ChatGPT to learn about word classes, collocations, and idiomatic language expressions. It can generate customized extended exercises, collocation recommendations, and reference sentences for specific knowledge points, offering guidance tailored to students' language proficiency levels. However, due to students' unfamiliarity with ChatGPT's generalized language expressions and the potential impropriety of the prompts they input when using ChatGPT independently, they have not been able to fully and effectively utilize the learning resources and feedback information provided by the system during the writing and revision processes. Therefore, in teaching, teachers should instruct students on how to use large-language models represented by ChatGPT and familiarize them with the prompt design of such models. By doing so, the learning-facilitating role of ChatGPT feedback can be fully realized, and students can make use of the revision and learning cues provided by ChatGPT to continuously improve their autonomous writing and revision abilities.

Harnessing peer feedback's peer-evaluation and learning-promoting role to transform students from passive recipients to active thinkers. Peer feedback highlights the dominant position of students, enabling their genuine participation in writing evaluation. This is more conducive to students' transformation from passively and mechanically accepting revision suggestions to actively and proactively thinking about revisions. During the feedback process, students, assuming the role of evaluators, can objectively and directly identify various writing techniques and goals, and personally experience different ways in which other students handle the same writing content. This allows them to learn from each other's strengths and promote their own learning. To enhance the effectiveness of peer feedback, teachers should, through various means such as training, demonstration, and guidance, help students clarify feedback goals, become familiar with evaluation criteria, and master feedback methods. Additionally, corresponding measures should be taken to encourage and guide students to actively participate in peer-evaluation feedback activities. By doing this, students can experience the "scaffolding" role they play in helping their peers improve the quality of their written texts, thereby strengthening their feedback motivation and stimulating their subjective initiative.

Utilizing teacher feedback's evaluative and learning-guiding role to comprehensively improve students' writing and revision abilities. After two rounds of feedback and revision by ChatGPT and peers, the remaining problems in students' compositions often prove insurmountable for the students themselves. Hence, the evaluative and learning-guiding role of teacher feedback becomes particularly crucial. Regarding these issues in the compositions, the feedback information provided by teachers should be instructive and forward-looking. It should not only assist students in correcting errors and improving the text quality but also enable students to clearly identify the issues to be noted and the directions for future efforts. This forms a complete cycle of teaching, learning, and evaluation, effectively enhancing students' writing and revision abilities. On one hand, when choosing feedback methods and types of comments, teachers should consider the potential revision effects they may trigger and guide students on how to effectively revise their compositions by reading teacher comments, accumulate knowledge, and improve their skills. On the other hand, teachers need to shift their role perception. They should not only act as language teachers but also as readers of students' compositions and writing consultants. Teachers should communicate with students on an equal footing from a communicative perspective regarding composition structure, information content, etc. (Wang, 2006), guiding students to focus not only on language forms but also on the content of the article. By helping students master writing knowledge and techniques and broaden their writing ideas, teachers can continuously improve students' writing and revision abilities.

The multiple-feedback approach based on the automated evaluation system emphasizes the monitoring and regulation of students' writing processes. By adopting a feedback model that combines system feedback, peer feedback, and teacher feedback, learning spaces can be expanded, and an organic combination and mutual supplementation among different types of feedback can be achieved. This stimulates students' initiative and enthusiasm in writing and revision. Moreover, the combination of human-machine cooperation in assessment introduces interpersonal interaction into writing feedback, overcoming the limitation of automated writing evaluation systems that only focus on cognitive processing. Thus, the purpose of writing communication can be effectively achieved both in form and meaning.

Acknowledgement

This paper is supported by the research project [24XJKT065] at Zhejiang University of Finance and Economics. A Study on Writing Behavior and Effectiveness of Second Language Learners in a Mixed Human-machine Feedback Environment.

References

Applebee, A. N., & Langer, J. A. (2011). "EJ" extra: A snapshot of writing instruction in middle schools and high schools. *English Journal*, 100(6), 14–27.

Bai, L.F. & G.W. Hu. (2017). In the face of fallible AWE feedback: how do students respond? *Educational Psychology*, 37(1): 67-81.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31.

Chapelle, C.A., E. Cotos & J. Lee. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3): 385-405.

Chen, C.F.E. & W.Y.E. Cheng. (2008). Beyond the design of automated writing evaluation: pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2): 94-112.

Chen, D., Hebert, M., & Wilson, J. (2022). Examining human and automated ratings of elementary students' writing quality: A multivariate generalizability theory application. *American Educational Research Journal*, 59(6), 1122–1156.

Chodorow, M., M. Gamon & J. Tetreault. (2010). The utility of article and preposition error correction systems for English language learners: feedback and assessment. *Language Testing*, 27(3): 419-436.

Devlin J, Chang MW, Lee K, Toutanova K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Association for Computational Linguistics*, 4171–4186.

Dikli, S. (2010). Nature of automated essay scoring feedback. *CALICO Journal*, 28(1): 99-134.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107.

Falk, J. (2023). AI och skolan: Praktiska tips och framtidsspaningar om chattbottar i skolan: Vol. Version 2.7. github.com/Itangalo/AI-och-skolan.

Graham, S., Harris, K., & Chambers, A. B. (2016). Evidence-based practice and writing instruction. *Handbook of Writing Research*, 2, 211–226.

Graham, S., Harris, K., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment.* A report from Carnegie Corporation of New York. New York: Carnegie Corporation of.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523–547.

Kim, S., Shim, J., & Shim, J. (2023). A Study on the Utilization of OpenAI ChatGPT as a Second Language Learning Tool. *Journal of Multimedia Information System*, 10(1), 79–88.

Lavolette, E., C. Polio & J. Kahng. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2): 50-68.

Li, Z., S. Link, H. Ma, H. Yang & V. Hegelheimer. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System*, (44): 66-78.

Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research*, 8(1), 149–175.

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, 2(3), 100083.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65.

Wilson, J., & MacArthur, C. (2024). Exploring the role of automated writing evaluation as a formative assessment tool supporting self-regulated learning in writing. *In The Routledge international handbook of automated essay evaluation* (pp. 197-220). Routledge.

Yang, M., Badger, R. & Z. Yu. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, (3):179-200.

黄爱琼，张文霞. (2018). 英语作文自动评价反馈对学生词汇修改的影响——以批改网为例. 现代

教育技术, (7):71-78. [Translation]: Huang, A. & Zhang, W. (2018). The Effect of Automated Writing Evaluation Feedback on Students' Vocabulary Revision——Taking Pigai.org for Example. *Modern Educational Technology,* (7):71-78.

黄静，张文霞. (2014). 多元反馈对大学生英语作文修改的影响研究. 中国外语, (1):51-56. [Translation]: Huang, A. & Zhang, W. (2014). The Impact of the Integrated Feedback on Students' Writing Revision. *Foreign Languages in China,* (7):71-78.

李奕华. (2015). 基于动态评估理论的英语写作反馈方式比较研究. 外语界, (3):59-67. [Translation]: Li, Y. (2015). A comparative study of three types of feedback in EFL writing based on the dynamic assessment theory. *Foreign Language World,* (3):59-67.

卢鹿. (2016). 基于自动评价系统的第二写作过程研究. 外语界, (2):88-96. [Translation]: Lu, L. (2016). A study of the second writing process based on an automated essay evaluation tool. *Foreign Language World,* (3):59-67.

石晓玲. (2012). 在线写作自动评改系统在大学英语写作教学中的应用研究——以句酷批改网为例. 现代教育技术, (2):88-96. [Translation]: Shi, X. (2012). A Tentative Sudy on the Validity of Online Automated Essay Scoring Used in the Teaching of EFL Writing ——Exemplified by http//www.pigai.org. *Modern Educational Technology,* (2):88-96.

唐锦兰. (2014). 探究写作自动评价系统在英语教学中的应用模式. 外语教学理论与实践, (1):49-57. [Translation]: Tang, J. (2014). How to integrate an automated writing assessment tool in the EFL classroom. *Foreign Language Learning Theory and Practice,* (1):49-57.

王俊菊． (2006). 总体态度、反馈类型和纠错种类——对大学英语教师作文 书面反馈的探讨. 国外外语教学, (3):24-30. [Translation]: Wang, J. (2006). General attitude, types of feedback and types of error correction: A probe into the written feedback of college English teachers. *Foreign Language Learning Theory and Practice,* (3):24-30.

王利娜，牟 蕾，吴勇毅，Beth Cappelletti. (2018). 书面修正性反馈有效性的元分析研究. 西安外国语大学学报, (4):40-46. [Translation]: Wang, L., Mou, L., Wu, Y.& Cappelletti, B. (2014). A meta-analysis on the efficacy of written corrective feedback. *Journal of Xi'an International Studies University,* (4):40-46.

武永，张文霞. (2016). 作文自动评价系统和教师反馈对大学生英语作文修改的影响研究. 中国外语教育, (1):12-19. [Translation]: Wu, Y., & Zhang, W. (2016). Impact of automated writing evaluation system and teacher feedback on students' writing revision. *Foreign Language Education in China*, (1):12-19.

杨玲. (2013). 作文自动评价系统在高水平学生英语写作学习中的应用. 现代教育技术, (5):73-77. [Translation]: Yang, L. (2013). On the Application of AWE System in High-level Students' EFL Writing Learning. *Modern Educational Technology*, (5):73-77.

朱 晔，王 敏． (2005). 二语写作中的反馈研究: 形式、明晰度及具体效果. 现代外语，(2):170-180． [Translation]: Yang, L. (2013). A study of written corrective feedback: Forms, explicitness and effects. *Modern Foreign Languages*, (2):170-180.