

A Study on the Efficacy of End Commentary Generated by ChatGPT and Perceptions of Junior Middle School Students

Xiaojuan Yin¹ Yanqing Lin²

¹ College of Foreign Languages, Minjiang University, Fuzhou, Fujian 350108, China

² Fuzhou No. 7 Middle School, Fuzhou, Fujian 350012, China

¹Corresponding author. Email: yinxiaojuan@mju.edu.cn

ABSTRACT

In this study, through an English writing teaching experiment, questionnaires and interviews, the efficacy of corrective feedback of ChatGPT on 46 participants' English writing and participants' acceptance of the end commentary generated by ChatGPT were investigated among junior high school students in an ordinary middle school in Fuzhou, China. It was found that ChatGPT performed well in identifying ten categories of linguistic errors including grammar, spelling, word choice etc.. The accuracy of students' error corrections based on ChatGPT's feedback was as high as 98%, indicating participants have a high acceptance and effective use of the end commentary generated by ChatGPT. Based on the Technology Acceptance Model (TAM), results of questionnaires and interviews revealed that students rated ChatGPT highly in terms of ease of use and usefulness, with scores of 3.86 and 4.0; good attitude, with a score of 4.2; a strong willingness to use it, with a score of 4.1; and high evaluation of its actual use, with a score of 4.0. However, ChatGPT's feedback on sentence logic, writing content and paragraph structure still needs to be improved. The results of the study provide inspiration to English practitioners and researchers on how to effectively use ChatGPT to assist writing instruction, and highlight the importance of enhancing the awareness and education of teachers and students about Generative AI technology.

Keywords: ChatGPT, End commentary, Middle school students, English writing, Technology Acceptance Model (TAM).

1. INTRODUCTION

With the rapid progress of artificial intelligence technology, large language models (e.g., ChatGPT, Kimi ai, Bard, etc.) are increasingly used in English teaching. With their powerful language processing capabilities, these models not only provide students with personalized learning paths and suggestions, but also play an important role in assisting teachers in providing feedback on students' English writing. Imran & Almusharraf (2023) argue that ChatGPT can deeply analyze students' writing styles, grammatical errors, and logical structures, so as to provide targeted feedback and suggestions, and effectively improve students' English writing skills.

In English writing instruction, teachers' written feedback is proved to be effective, but it is difficult

to provide in-depth analysis and feedback on every student's composition in time because of the large size of English class at middle schools across China. The introduction of large language models will greatly compensate for this deficiency, making it possible for every student to obtain more detailed and comprehensive writing guidance shortly. It has been shown that corrective feedback and end commentary generated by these models can effectively increase students' interest and self-confidence in writing, and also help improve students' language use and logical thinking skills (Bašić & Banovac, 2023).

However, although the application of large language models in English teaching has been explored recently among adult learners, the middle school students' acceptability of ChatGPT, as well

as their attitudes and needs for generative AI is still a topic that deserves in-depth exploration.

In this context, this study aims to understand the efficacy of end commentary generated by ChatGPT and junior middle school students' perceptions of ChatGPT's feedback on their English writing through an experiment, the questionnaire and the interview.

2. LITERATURE REVIEW

2.1 *ChatGPT and English Writing Instruction*

Since November of 2022, ChatGPT has attracted the close attention of a wide range of researchers in the field of language education. As a state-of-the-art natural language processing tool, ChatGPT opens up new possibilities for foreign language education with its powerful language generation and comprehension capabilities.

Currently, researchers' interest in ChatGPT focuses on the following two aspects: first, about users' perceptions. With regard to teachers' opinions on its use, many teachers focus on how to use ChatGPT to optimize the teaching process and enhance the teaching effect. For example, Chen and Lv (2024) showed that teachers can interact with students in real time through ChatGPT so that they can grasp students' learning needs more accurately and develop targeted teaching plans. In addition, students' perception of the use of ChatGPT is another research interest. Students' attitudes and experiences with ChatGPT are directly related to its practical application in foreign language education. Students are found to generally believe that ChatGPT has positive effects in assisting learning and improving language application skills (Liu, 2023). However, Ying, etc. (2024) worry that over-reliance on ChatGPT will affect their independent thinking and expression.

Secondly, given ChatGPT's powerful writing function, researchers have paid more attention to its auxiliary writing function, and the relevant research results are relatively rich. For academic writing, researchers have explored ChatGPT's auxiliary role in essay conception, literature review, language embellishment and so on. For example, Guo et al. (2023) have pointed out in their studies that ChatGPT can help students quickly generate thesis outlines and provide rich academic vocabulary and sentence patterns, thus improving their academic writing skills.

In general English writing, ChatGPT also shows its unique advantages. It not only provides diversified language expressions to help students overcome writing difficulties, but also plays an important role in error correction. Mao et al. (2024) have found that using ChatGPT for error correction can significantly improve the quality of students' writing and reduce grammatical and spelling errors.

Generally speaking, the research on ChatGPT has achieved great results, but there is still gap in the existing research. Firstly, the research on ChatGPT has basically focused on adults, and there is a lack of in-depth research on teenagers, especially among junior middle school students. Second, previous studies have focused on ChatGPT's feedback on linguistic errors in writing, but lacked studies on ChatGPT's non-corrective feedback (i.e, overall end commentary).

Therefore, this study aims to investigate the efficacy of ChatGPT-generating end commentary and junior middle school students' perceptions. By analyzing junior middle school students' acceptability of ChatGPT-generating end commentary, as well as their experiences and feedback on modifying their compositions, aiming to provide useful insights for the application of ChatGPT in secondary school foreign language education.

2.2 *Teachers' Feedback on English Writing*

In the field of second language writing, teacher's written feedback, as an important form of feedback, have been attracting much attention. In this part, a review on the existing research in terms of the content, types and effects of teachers' written feedback will be presented, aiming to reveal the current status and the research gap.

There are various types of teacher's written feedback, which have been categorized by different researchers according to different criteria. Among them, Li's (2013) research pointed out that teacher's written feedback includes three types: scoring, corrective feedback and non-corrective feedback (e.g., end commentary). Teacher's end commentary can be an overall evaluation of a student's composition or a critique of specific paragraphs or sentences. It can be either an affirmation of students' strengths or a pointing out and suggestion of deficiencies.

The effectiveness of teacher's written feedback is an important indicator of their quality. Most

studies show that positive and targeted comments can have a positive effect on students' writing. For example, Li & Wan (2018) found through empirical research that teacher comments can increase students' interest and self-confidence in writing and promote the improvement of students' writing skills. At the same time, effective teacher's written feedback can also stimulate students' writing motivation and make them participate in writing activities more actively. Moreover, although teachers tend to avoid writing comments in the imperative form, imperatives were found to be more influential on revisions than questions or statements (Sugita, 2006).

The content of teacher's written feedback is one of the key factors in evaluating their quality. Most studies point out that the content of teacher's written feedback should be specific and targeted, able to point out the strengths and weaknesses in students' compositions and provide specific suggestions for improvement. For example, Li (2007) mentioned in his study that an effective teacher's written feedback should contain evaluations of students' language expression, passage structure, logical thinking, etc., and at the same time, it should pay attention to the exploration of students' personality traits and development potential.

In summary, teacher's written feedback has important research value and application prospects in second language writing. However, it consumes teachers' a lot of time and energy to write a targeted end commentary, so it is impossible for teachers to provide detailed end commentary for every student' essay because of the large class size in Chinese EFL context. Therefore, it is of great significance to explore ChatGPT's potentials of generating good end commentaries.

2.3 Technology Acceptance Model

In order to theoretically support the measurement of middle school students' acceptance of ChatGPT-generating English writing end commentary, this study draws on the Technology Acceptance Model (TAM) developed by Fred Davis (cited from Al-Emran et al. 2018), a leading scholar in the field of information management. Over the past three decades, the TAM has been recognized as an influential model that explains individuals' acceptance, rejection, and use of information technology in the areas of information management, digital sociology, educational technology, and Computer Assisted Language

Learning (Alfadda&Mahdi, 2021). As a powerful framework for predicting individual adoption of new technologies, the TAM model has at its core the attitude-intention-behaviour relationship. As Davis (1989) points out, attitude is the degree to which people are interested in and actively evaluate the use of a particular technology. Two key predictors of a person's attitude are perceived ease of use, which is "the extent to which a person believes that using a particular system can be effortless", and perceived usefulness, which points to "the extent to which a person believes that using a particular system can improve his or her job performance". Attitudes can influence people's behavioral intentions to use technology, and ultimately can predict actual behaviour in using technology. Notably, despite attempts to incorporate additional external variables (e.g., subjective norms, computer self-efficacy) to explain differences in technology use behaviors (Bailey, et al., 2022), it is generally accepted that the core components of TAM are perceived ease of use, perceived usefulness, attitude, behavioral intention and actual use. Therefore, this TAM-based study aims to measure and explain middle school students' acceptance of English writing end commentary generated by ChatGPT.

3. EXPERIMENT DESIGN

3.1 Research Questions

What is the efficacy of ChatGPT's feedback in terms of its capability of error recognition in the end commentary?

What are subjects' perceptions of ChatGPT's feedback on their English essays?

What factors influence the subjects' attitudes toward ChatGPT's feedback?

3.2 Context and Participants

China has the largest number of EFL learners in the world, including a great number of middle school students. So it is meaningful to understand their perceptions and attitudes toward this new type of end commentary, so as to provide a reference for the improvement of English writing teaching in junior middle schools.

In this study, the third-grade students in the general class of a junior middle school in Fuzhou, China were chosen as the participants out of convenience. Since they are about to face the high

school entrance exam, they have more writing practices and have a better understanding of the requirements and scoring criteria of English compositions. Therefore, it is important to assess their acceptability of the end commentary generated by ChatGPT.

In the process of participants' selection, the researcher communicated fully with their English teachers to ensure the smooth progress of the experiment. At the same time, the researcher also made a preliminary understanding of the students' English grades and learning attitudes in order to better analyze the experimental results.

The study involved 46 participants, who were between the ages of 14 and 16, with roughly equal gender ratios. Most of these students come from ordinary families and have a solid foundation in English, but there is still some room for improvement in writing ability and depth of thinking. In their daily studies, they are generally able to complete their homework and exercises according to the teacher's requirements, but they often lack sufficient guidance and feedback on writing and revising their English compositions. The total number of interviewees was eight, with a good gender balance, i.e. four males and four females. The eight individuals were distributed as follows, and were divided into three groups (high, medium and low), based on their latest mid-term examination scores: the high English proficiency level group (scores 130 or above out of 150): 3 interviewees in total, including 2 males and 1 female. The medium level group (scores between 110 and 130): 2 interviewees, including 1 male and 1 female. The low level group (scores below 110): 3 interviews, including 1 male and 2 females. This distribution was intended to ensure that the interviewees were balanced in terms of both gender and level of English proficiency, thus enabling a more comprehensive and objective understanding of the views and feelings of different groups.

In order to guarantee the precision and validity of this study, the participants' English compositions recently finished in a unified examination were selected as the study sample. The uniformity of the essay topics in this exam ensured that all participating students were writing under the same standard, which could more accurately reflect their English writing level. In collecting the essay samples, a strict screening criteria was adopted to ensure that the selected essays were representative and comparable. First, compositions from 84 students in two classes of the school's junior high

school were chosen, which were representative of the students' English learning and could reflect the general level of the school's junior high school students. Second, some compositions were excluded, such as blank papers or compositions with only a few words (the requirement of the essay is 80 words, so if the essay is less than 50% of the required words, it is excluded) to ensure the quality of the selected compositions. In the end, selected a total of 46 qualified compositions were selected as the study sample. In the process of collecting the essay samples, the researcher strictly adhered to the relevant regulations and procedures of the school to ensure that the privacy and rights of the students were safeguarded.

3.3 Methods

3.3.1 Questionnaires

The questionnaire, which is divided into three crucial sections, is carefully designed by the author and her mentor. Its framework is inspired by the classic questionnaires of scholars such as Liu and Ma (2023), aiming to ensure its professionalism, scientificity, and effectiveness. Part I and Part II of the questionnaire used nominal scales to collect information about participants' gender, age, knowledge of the ChatGPT, and writing and revising habits. Part III consisted of a five-point Likert scale (ranging from "completely disagree" 1 to "completely agree" 5) based on our findings from Davis and later TAM instruments (cited from Su & Mei, 2022). The five constructs are perceived ease of use, perceived usefulness, attitude, behavioral intention and actual use. Part III comprises 10 items in total, with items 1-2 corresponding to perceived ease of use, items 3-4 to perceived usefulness, items 5-6 to attitude, items 7-8 to behavioral intention, and items 9-10 to actual use. In a pilot study, the questionnaire was filled out by three non-subject students at the same middle school to ensure content validity and accuracy and non-ambiguity. After three rounds of revision, the questionnaire containing 10 formal items was finally generated. To ensure the accuracy and reliability of the questionnaire, a reliability and validity analysis were conducted. In terms of reliability, the overall Cronbach's Alpha coefficient of the questionnaire reached 0.85, indicating good internal consistency. The coefficients for each sub-dimension also exceeded 0.7, further demonstrating the stability of the questionnaire. In terms of validity, the questionnaire design took into full consideration both content validity and structural

validity. Based on relevant literature and advice from mentors, revisions have been made through pre-testing to ensure that the questionnaire content accurately reflected the research objectives. The results of factor analysis showed that the questionnaire items could be reasonably grouped into preset dimensions, and there were clear distinctions between the dimensions, which proved the rationality of the questionnaire structure. In summary, this questionnaire possesses high reliability and validity, capable of reliably reflecting the attitudes and opinions of junior high school students, providing strong support for subsequent data analysis and research.

3.3.2 *The Semi-structured Interview*

This study adopts a semi-structured interview methodology, aiming to explore the participants' perceptions of the efficacy of ChatGPT-generating end commentary and their writing needs. The method maintains the coherence of the question framework and can be flexibly adjusted according to respondents' feedback to ensure a more precise understanding of students' actual perceptions and feelings about the end commentary, and also ensures the accuracy and reliability of the survey results by building trust and collecting authentic feedback through direct interaction with students.

In the design of the interview outline, the following principles were followed:

- Question focus: ensure that each question is closely related to the survey topic and can directly reflect students' acceptability and perceptions of ChatGPT-generating end commentary.
- Hierarchical progression: the questions were designed in a principle of increasing difficulty, starting from understanding students' daily writing habits, gradually transitioning to perceptions and acceptability of ChatGPT-generating end commentary, and finally exploring future willingness to use them.

A combination of open and closed questions: there were both closed questions in order to collect quantitative data and open questions in order to obtain more in-depth qualitative information.

The purpose of this interview was to gain insight into middle school students' acceptability of the English writing end commentary generated by ChatGPT and their actual needs in English writing. By asking students about their daily habits of

revising and the reasons behind, their attitudes and the importance they attach to writing will be explored. Then, students' perceptions of ChatGPT-generating end commentary will be probed to find out whether they like this type of feedback and whether these end commentaries help them improve their writing. In addition, the ways in which the ChatGPT-generating end commentary has been most helpful to students' writing will be analyzed and specific examples will be collected to deepen our understanding. Finally, students' willingness to use it in the future, as well as their changes in beliefs on what could be improved and what they expect from the ChatGPT-generating end commentary will be investigated, thus providing a valuable reference for subsequent research and improvement.

3.4 *Procedures*

This study explores the effects of ChatGPT-assisting English writing instruction among junior high school students in an ordinary secondary school in Fuzhou. First, students' essays finished in the latest mid-term exam were collected and transcribed electronically, and errors were identified and end commentary in Chinese. Considering that the subjects were junior high school students with insufficient vocabulary and limited English proficiency, the feedback results from ChatGPT to students' compositions in this study were mainly in Chinese through ChatGPT 3.5. Then, the end commentary was printed and returned to the students, who were asked to use 10-15 minutes to revise their compositions at recess. Afterwards, the revised drafts were collected, and the first and revised drafts were compared to assess the students' revisions. Finally, the survey and semi-structured interview were conducted to understand students' attitudes and acceptance of the end commentary provided by ChatGPT. "Table 1" shows the procedure.

Table 1. The experimental procedure

Step	Contents	Date
1	Prompts design: Based on previous literature, the prompts were designed with the help of the supervisor and then tested by interacting with ChatGPT.	December 1st, 2023 - December 7th, 2023
2	Samples choice: Clarify the target audience of the collected essays, screen qualified samples and organize them into a Word document.	November 12th, 2023 - November 18th, 2023
3	Questionnaire design: The questionnaire was designed based on the TAM theoretical model.	December 20th, 2023- January 3rd, 2024
4	ChatGPT's feedback: Based on the prompts, student essays were attached and ChatGPT was asked to provide error detection feedback and end commentary.	January 2nd, 2024 - January 4th, 2024
5	English writing revision: Students received ChatGPT's feedback and were required to revise their essays accordingly in 10-15 minutes.	January 5th, 2024
6	The survey: Questionnaires were distributed to the participants, and questionnaires were collected after they were filled out.	January 8th, 2024
7	Interview: A total of 8 interviews , including 2-3 representative students of different English proficiency levels respectively (high, medium, and low) participated the interview. Every student finished the interview within 3 mins.	January 9th, 2024

3.5 Data Collection and Analysis

In this study, participants' writing errors were classified into ten categories such as grammatical errors, spelling errors, word choice, punctuation, sentence structure, tense inconsistency, clarity, content, passage structure and wrong error detection according to Algaraaby and Mahyoob's (2023) classification. The researcher used manual counting of errors identified by ChatGPT. To calculate the types of students' writing errors detected by ChatGPT, the method of frequency analysis were mainly used.. Base on the carefully calculated frequency of each error type, a clear picture of the types of errors commonly made by students in English writing and their distribution was shown. In this process, the task of identifying the number of errors was done manually by the author and confirmed by the supervisor. The correctness of the judgement of each error type was determined by the author and the supervisor to ensure the authority and reliability of the results of the analysis.

For the questionnaire results, data were analyzed by using SPSS27.0 software. Firstly, descriptive statistical analysis was done to describe the overall acceptability and satisfaction with the ChatGPT-generating end commentary by calculating the mean, standard deviation and other indicators. As mentioned earlier, the overall situation regarding students' acceptance and

satisfaction with ChatGPT is reflected in a total of ten items in the questionnaire. These items correspond to five different dimensions: perceived ease of use (items 1-2), perceived usefulness (items 3-4), attitude (items 5-6), behavioral intention (items 7-8), and actual use (items 9-10). Following the methodology of Liu and Ma (2023), this study calculates the average score for each pair of items, integrating them into the five core components of the TAM for further analysis. This provided a visualization of the students' attitudes. Secondly, in order to test the stability and reliability of the results of the questionnaire, reliability and validity analyses were conducted. In addition, correlation analysis was applied to explore the relationship between students' acceptability of the ChatGPT-generating end commentary and influencing factors (e.g., gender, English proficiency level). Finally, differences in acceptability among different groups (e.g., gender, English proficiency level) were compared by ANOVA or Chi-square test.

For the interview results, content analysis was mainly used. Content analysis extracted key messages and themes by coding and categorizing the interviews sentence-by-sentence or paragraph-by-paragraph, helping us to gain insight into students' perceptions and feelings about the ChatGPT-generating end commentary.

4. RESULTS AND DISCUSSIONS

4.1 ChatGPT's Capability of Error Detection

The types of errors indicated by the ChatGPT provide us with a rich data base when analyzing student essays in depth. Based on the descriptive data, a comprehensive and in-depth understanding

of the types of errors in student compositions is provided.

Based on the classification method of English writing errors proposed by Algaraady and Mahyoob (2023), this study conducted a detailed categorization of the error types pointed out by ChatGPT, totaling ten categories. "Table 2" shows the number of errors detected by ChatGPT and revised by participants.

Table 2. Writing errors detected by ChatGPT and revised by students

Categories	Examples	Detection and revision	Times
grammatical errors	When we see litter, we ought to pick up it and throw it into the dustbin.	detected by ChatGPT	18
		revised by participants	18
spelling errors	Don't walk on the gress and pick the flowers.	detected by ChatGPT	63
		revised by participants	63
word choice	In other word,you ought not to hurt them.	detected by ChatGPT	11
		revised by participants	11
punctuation	Now. Let's me tell you something about it.	detected by ChatGPT	2
		revised by participants	2
sentence structure	We'd better save energy, ride a bike instead of driving a car.	detected by ChatGPT	3
		revised by participants	2
tense inconsistency	We ought not to throw litter here and there. We had better put it into the dustbin.	detected by ChatGPT	1
		revised by participants	1
clarity errors	We'd better take care of the trees and the flowers.	detected by ChatGPT	9
		revised by participants	8
content	In each action, further explanation of why these actions are helpful in protecting the environment can be added.	detected by ChatGPT	64
		revised by participants	0
passage structure	Adding transition sentences between each paragraph makes the essay tighter and smoother.	detected by ChatGPT	49
		revised by participants	2
wrong error detection	We shouldn't throw rubbish here and there, and we don't spit anywhere in school.	detected by ChatGPT	9
		revised by participants	3

Firstly, in terms of basic language rules, students' errors in grammar, sentence structure, punctuation and tense inconsistency are relatively few. This indicates that students performed well in mastering basic language rules and were able to construct well-structured and logical sentences. However, this does not mean that the existence of these errors can be ignored, especially in the pursuit of high-quality writing where precise mastery of language rules is crucial.

Secondly, According to "Table 2", of the ten errors, ChatGPT identifies errors of grammar, spelling, content, and passage structure more frequently. For example, there were 64 suggestions about content, 49 about structure, and 63 about spelling. Spelling errors and word choice errors are more prominent in students' compositions.

Moreover, as is shown in "Table 3", the high mean (1.37) and large standard deviation (1.34) of the spelling errors indicate that students had a general struggle with spelling and that the errors were spread out. This may be due to students' poor memory of vocabulary or lack of adequate proofreading and checking during the writing process. The mean value of word choice errors, although not high (0.24), reflects the fact that students still need to improve their use of precise and appropriate vocabulary.

In addition, content errors and passage structure errors are also issues that need to be focused on in students' essays. The high mean (1.39) value of content errors indicates that students are deficient in understanding the topic, expressing the central idea and constructing reasonable content. This may be

related to students' knowledge base, thinking skills, writing experience and limited revising time. The mean (1.07) and standard deviation (0.39) of passage structure errors also indicate that students need to strengthen in constructing coherent and logical paragraph.

Finally, it is worth noting that the ChatGPT also has some problems in detecting errors. The mean value of wrong error detection, although not high (0.20), suggests that ChatGPT is not entirely accurate in identifying errors. This may be due to limitations of the model itself, or difficulties in dealing with specific types of errors. Therefore,

when using ChatGPT for composition error identification, it is necessary to maintain a cautious and objective attitude and combine other methods and tools for comprehensive evaluation. In summary, the students' compositions performed well in terms of basic language rules, but there were obvious deficiencies in spelling, word choice, content and passage structure. Also, ChatGPT needs to be improved in detecting errors. To address these issues, the researchers need to take targeted measures to strengthen students' writing ability and optimize the recognition performance of ChatGPT to better serve students' writing learning and improvement.

Table 3. Descriptive analysis of writing error types identified by ChatGPT

	Mean	Standard Deviation	Sample Variance
Grammatical errors	.39	.493	.243
Spelling errors	1.37	1.339	1.794
Word Choice	.24	.431	.186
Punctuation	.04	.206	.043
Sentence structure	.07	.250	.062
Tense inconsistency	.02	.147	.022
Clarity Errors	.20	.453	.205
Content	1.39	.577	.332
Passage Structure	1.07	.389	.151
wrong error detection	.20	.453	.205

a Note: the total number of samples is 46.

4.2 Results of Questionnaires

4.2.1 Overall Perception

This study conducted an in-depth data analysis based on SPSS27.0 software on the acceptability of 46 junior high school students regarding the

English writing end commentary generated by ChatGPT. Based on the TAM theoretical model, the participants' perceptions of ChatGPT were investigated. "Table 4" is the results of the descriptive statistical analyses of the dimensions of perceived ease of use, perceived usefulness, attitude, behavioral intention, and actual use.

Table 4. Descriptive analysis of acceptance of ChatGPT-generating end commentary

	N	Minimum	Maximum	Mean	Standard Deviation
Perceived Ease of Use (PEU)	46	1.5	5.0	3.859	.7864
Perceived Usefulness (PU)	46	1.5	5.0	4.000	.7888
Attitude (AT)	46	3.0	5.0	4.207	.6287
Behavioral Intention (BI)	46	3.0	5.0	4.120	.7901
Actual Use (AU)	46	3.0	5.0	4.000	.6912

a Note: the full score is 5.

4.2.1.1 Perceived Ease of Use (PEU)

As is shown in "Table 4", the perceived ease of use of the ChatGPT-generating end commentary by participants shows a more positive trend. The

minimum value was 1.5, indicating that very few students perceived it to be difficult to use, while the maximum value was 5.0, showing that some students perceived it to be very easy to use. The mean value of 3.86 is close to the medium-high

level of 4.0, indicating that the majority of students found the ChatGPT-generating end commentary is easy to understand and follow. The standard deviation was 0.79, indicating that there was some fluctuation in students' ratings of perceived ease of use, but the overall difference was not significant.

4.2.1.2 Perceived Usefulness (PU)

In terms of perceived usefulness, the participants generally agree that the end commentary generated by ChatGPT is useful. As shown in table 4, the minimum value of 1.5 shows that some individual students perceived the ChatGPT-generating end commentary as having limited usefulness in improving English writing, while the maximum value of 5.0 indicates that some students perceived it as very useful. The mean value of 4.0 is moderately high, reflecting that most students believe that the ChatGPT-generating end commentary has a positive impact on their English writing. The standard deviation is 0.79, indicating that there is a slight difference in students' ratings of perceived usefulness, but the overall consistency is high.

4.2.1.3 Attitude (AT)

Participants' attitudes towards the ChatGPT-generating end commentary were positive overall. The minimum value of 3.0 indicates that there are students who have reservations about the ChatGPT-generating end commentary while the maximum value of 5.0 indicates that there are students who strongly approve of it. The mean value of 4.21 is above average, indicating that the majority of students hold a positive view of the ChatGPT-generating end commentary. The standard deviation is 0.63, indicating that students' ratings on attitudes are relatively concentrated with small differences.

4.2.1.4 Behavioral Intention (BI)

The participants' willingness to use the ChatGPT-generating end commentary in the future is generally strong. As it can be seen from "Table 4", the minimum value is 3.0, indicating that a few students have a wait-and-see attitude towards the future use of the ChatGPT-generating end commentary, while the maximum value is 5.0, showing that there are students who are looking forward to it. The mean value is 4.12, which is above average, indicating that most students are willing to continue to use the ChatGPT-generating end commentary in the future to assist their English

writing. The standard deviation is 0.79, indicating that there are some fluctuations in students' ratings of their willingness to use it in the future, but the overall trend is positive.

4.2.1.5 Actual Use (AU)

In terms of practical application, the participants' evaluations were similarly more positive. As is shown in "Table 4", the minimum value is 3.0, indicating that a few students have some difficulties in the practical application of the ChatGPT-generating end commentary; while the maximum value is 5.0, showing that some students think it is very effective in practical application. The mean value is 4.0, which is close to the high range, indicating that most students are able to use the ChatGPT-generating end commentary better in their actual writing. The standard deviation is 0.69, reflecting that the students' evaluation in practical application is relatively stable with little difference.

In summary, the participants' acceptance of ChatGPT-generating English writing end commentary is high overall, showing positive evaluations in perceived ease of use, perceived usefulness, attitude, behavioral intention and actual use. This provides strong support for the application of ChatGPT in English writing teaching. However, it should also be noted that there are certain differences and fluctuations in students' evaluations of the dimensions, so further attention to individual differences and optimization of the experience of using it are needed in future applications in order to better exploit the potential of ChatGPT in English writing teaching.

4.2.2 Influential Factors on Participants' Perceptions

Generally, the participants have a positive attitude towards the end commentary generated by ChatGPT, but there are differences in the needs of different genders.

Analysis shows that in terms of perceived ease of use, female students rated slightly higher (4.05), suggesting that they may be more likely to adapt to and accept the use of the ChatGPT-generating end commentary. In terms of perceived usefulness, females similarly showed a high level of acceptance (4.39) that the ChatGPT-generating end commentary had a positive effect on improving English writing. In terms of attitude, female students generally held a more positive view (4.36) of the ChatGPT-generating end commentary in

practical writing, and anticipation of future use.. In terms of behavioral intention and actual use, female students also showed higher acceptance (4.30), willingness (4.05) to use the ChatGPT-generating end commentary. To further assess whether the gender difference is significant, a paired-sample analysis was further conducted. As a result, the significant gender difference was only found in the dimension of “perceived usefulness” ($p < 0.05$), while no significant differences were found in the rest four dimensions ($P = 0.05 < 0.01$). In other words, female students scored significantly higher than male students. With regard to “usefulness” of end commentary generated by ChatGPT, this suggests that girls generally found them more useful, while boys were relatively more conservative or sceptical.

Moreover, to examine the factor of English proficiency level, the chi-square was conducted. It is found that the English proficiency level did not significantly influence the participants' perceptions

in the five dimensions ($p = 0.312 - 0.564 > 0.05$). In addition, this result may suggest that performance in these five dimensions is more influenced by other factors, such as personal experience, cultural background, or the characteristics of the technology itself, while the effect of English proficiency is not significant. Therefore, this result may suggest that other factors, such as personal experience, cultural background, or the characteristics of the technology itself needed to be further explored.

4.3 Interview Results

After the survey, 8 participants were selected to conduct in-depth semi-structured interviews, aiming to gain comprehensive understandings of gender and English proficiency factors' effects on students' overall attitudes and acceptance of the ChatGPT-generating end commentary. “Table 5” shows the basic information of the interview.

Table 5. Basic information of interview

Interviewees	Gender	Age	English proficiency	Time length	Content length(words)
A	Female	15	Medium	2'26	573
B	Male	14	High	3'15	646
C	Female	14	Low	2'44	544
D	Male	15	Low	2'55	670
E	Female	15	Low	2'13	611
F	Male	14	Medium	2'38	586
G	Female	15	High	3'03	617
H	Male	15	High	3'01	659

Regardless of their English proficiency, students generally held a positive and open attitude towards the ChatGPT-generating end commentary. As interviewee A said, “The end commentary given by ChatGPT are really useful and I can see the deficiencies in my writing from them.” Student B, who had a high level of English proficiency, stated, “I appreciated the depth and breadth of the ChatGPT-generating end commentary; it really inspired me.” In contrast, Student C, who had a lower level of English proficiency, said, “Although my English is not very good, ChatGPT's end commentary are very concise and easy to understand, and I can understand what I need to improve.” Students generally believed that the end commentary could help them better recognize their

shortcomings in English writing, so that they could make targeted improvements.

In terms of gender, although there was no significant difference between male and female students in terms of overall acceptance, female students showed a more delicate emotional experience in the interviews. Interviewee E (female) shared, “I read the end commentary given by ChatGPT several times carefully and think about how to improve my writing.” Their feedback on the ChatGPT-generating end commentary was more detailed, and they were able to describe in greater depth their own journey through the process of receiving the end commentary. While interviewee F (male) commented, “After receiving the ChatGPT-generating end commentary, I skimmed through it

and found the content to be okay. I didn't spend much time reading it in depth or thinking about how to further improve my writing." Although he was also open to the end commentary generated by ChatGPT, in contrast, he did not show the same enthusiasm for in-depth analysis and multiple revisions of the content as the female student. His feedback was shorter and expressed more of an acceptance of the results and a lack of deep engagement and reflection on the process.

In addition, it is found that students generally hoped the ChatGPT-generating end commentary to be more personalized and relevant. Interviewee H said, "I would like ChatGPT to be more aware of my writing style and habits and give advice that better meets my needs." They wished the ChatGPT can identify their own writing characteristics and problems more accurately, and to give advice that was more relevant to their needs.

Based on the findings of the questionnaire, during subsequent interviews, I delved into the reasons why students made relatively minor modifications to their essays despite the detailed content and structural suggestions provided by ChatGPT. Some students fed backed that due to their limited English proficiency, they found it difficult to correct deep-level errors in their essays. Others indicated that they were unable to fully implement the suggestions due to time constraints. Additionally, some students considered the modification process to be burdensome and thus opted for a more convenient approach.

Moreover, students expect the end commentary to be more personalized and precise. In future applications, the functions and performance of ChatGPT-generating end commentary can be further optimized based on this feedback to better meet students' needs.

5. CONCLUSIONS AND IMPLICATIONS

5.1 Conclusions

This study combined a teaching experiment, a questionnaire survey and interview to explore the efficacy of ChatGPT's linguistic error detection and the participants' acceptance of ChatGPT-generating end commentary. The findings are mainly summarized as follows:

First, during this experiment, ChatGPT successfully identified multiple types of errors, and students demonstrated a strong willingness to revise.

Specifically, students completely corrected all identified grammatical and spelling errors, and most of the word choice and punctuation issues were handled appropriately. However, students need to further improve their correction skills in the areas of sentence structure, content, and paragraph structure. In addition, although ChatGPT performed well in error detection, there were still instances of wrong classification that required manual review. Secondly, the questionnaire survey revealed that the participants generally held a positive attitude towards ChatGPT-generating English composition end commentary, believing that they were able to provide timely, objective, and targeted feedback. This result suggests that ChatGPT has some potential and value for application in English writing teaching and can serve as an effective tool for students' writing assistance. Thirdly, the interview results showed that students generally recognized ChatGPT's ability to identify errors in basic language rules, such as grammar and sentence structure. However, they also mentioned the limitations of ChatGPT in identifying advanced errors, such as content logic and paragraph structure. This suggests that when using ChatGPT for essay evaluation, its strengths and weaknesses should be objectively assessed and teachers' professional judgement should be incorporated. In addition, the types of students' writing errors detected by ChatGPT showed that ChatGPT performs well in identifying common language errors, but there is still room for improvement when it comes to deeper language comprehension and expression. This further supports the students' view from the interviews that the role of ChatGPT in essay assessment needs to be treated with caution.

5.2 Implications

The implications are as follows: firstly, ChatGPT can be used as an auxiliary tool for junior high school English writing teaching, providing students with real-time writing guidance and feedback. Through ChatGPT's written feedback, students can understand their shortcomings in writing in time and make targeted improvements. Secondly, the application of ChatGPT helps develop students' independent learning abilities. Students can use ChatGPT for more English writing practice and self-assessment, and gradually improve their writing skills by constantly adjusting and optimizing their writing strategies. In addition, ChatGPT can also be used as an auxiliary means for teachers to assess students' English writing. Teachers can depend on the feedback of ChatGPT

from get a more comprehensive understanding of students' writing level and progress, so as to make more personalized teaching plans. However, it should be noted that the feedback from ChatGPT is only for reference and cannot completely replace teachers' professional assessment.

This is an exploratory study, so there are still some limitations: the sample size was relatively small, which may affect the generalizability of the findings. Future research could expand the sample size to get a more comprehensive picture of the distribution of teacher ratings and ChatGPT scores. In addition, there may have been variables that were not taken into account during the survey, such as essay topic and difficulty, which could have affected students' acceptability of the end commentary. Future research should control these variables more carefully to improve the accuracy of the study.

AUTHORS' CONTRIBUTIONS

Xiaojuan Yin is mainly responsible experimental design, funding acquisition; project administration; writing original draft, review & editing draft of the paper .Yanqing Lin is mainly responsible for conducting the teaching experiment; data search and analysis; the final draft approval.

ACKNOWLEDGMENTS

This paper is funded by "2022 University-Level Public Basic Course Construction Project" of Minjiang University (College English III, No. MJU2022KC102), "University-Level Joint Science and Education Project (Humanities) General Project: Generative AI Technology Aiding the Practice and Research of College English Teaching" (No. MJKJ24024), and "Research and Practice on Digital Transformation of English Course with the Assistance of Digital Human Teachers and ChatGPT" of Minjiang University 2024 Educational Teaching Research and Reform Project (No. MJUJG2024A011).

REFERENCES

- [1] Al-Emran, M., V. Mezhyuev, & A. Kamaludin. Technology Acceptance Model in M-Learning Context: A Systematic Review. *Computers & Education*, 2018,125: 389-412. <https://doi.org/10.1016/j.compedu.2018.06.008>
- [2] Alfadda, H. A., & H. S. Mahdi. Measuring Students' Use of Zoom Application in Language Course Based on the Technology Acceptance Model (TAM). *Journal of Psycholinguistic Research*, 2021, 50 (4):883-900. <https://doi.org/10.1007/s10936-020-09752-1>
- [3] Algaraady, J., & M. Mahyoob. ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners. *Arab World English Journal (AWEJ) Special Issue on CALL*, 2023, (9):11-13. <https://doi.org/10.24093/awej/call9.1>
- [4] Bailey, D. R., N. Almusharraf, & A. Almusharraf. Video Conferencing in the e-Learning Context: Explaining Learning Outcome with the Technology Acceptance Model. *Education and Information Technologies*, 2022, 27 (6):7679-7698. <https://doi.org/10.1007/s10639-022-10949-1>
- [5] Ba š ić, Z., A.,Banovac, I., Kruzic, & I., Jerkovic. Better by You, Better than Me, ChatGPT-3 as Writing Assistance in Students' Essays. *Humanities and Social Sciences Communications*, 2023, (10) 750:9-14. <https://doi.org/10.1057/s41599-023-02269-7>
- [6] Chen, M. & M. C., Lv. College English Writing Instruction in the ChatGPT Environment. *Contemporary Foreign Language Studies*, 2024, (1):162-166.
- [7] Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *Management Information System Quarterly*, 1989, 13(3):319-340. <https://doi.org/10.2307/249008>
- [8] Guo, Q., R. L. Feng & Y. F. Hua. Using ChatGPT in English Academic Writing: Benefits and Issues. *Technology Enhanced Foreign Language Education*, 2023, (2):18-23. https://lib.cqvip.com/Qikan/Article/Detail?id=7109972288&from=Qikan_Search_Index
- [9] Imran, M., & N. Almusharraf. Analyzing the Role of ChatGPT as a Writing Assistant at Higher Education Level: A Systematic Review of the Literature. *Contemporary Educational Technology*, 2023, 15(4) ep464:1-14. <https://doi.org/10.30935/cedtech/13605>
- [10] Li, J. & W. Wan. The Impact of Teacher Feedback Types on the Effectiveness of

- Students' English Writing Revisions. *Journal of Xi'an International Studies University*, 2018, 26 (3):69-72.
<https://doi.org/10.16362/j.cnki.cn61-1457/h.2018.03.013>
- Potential, Risks, and Governance of ChatGPT. *Continuing Education Research*, 2024, (5): 56-59.
<http://www.qikan.com.cn/article/jxyj20240512.html>
- [11] Li, J. A Case Study on the Effectiveness of Teacher Written Feedback in English Writing Instruction. *Foreign Language World*, 2013, (2):87-89.CNKI:SUN:WYJY.0.2013-02-012
- [12] Li, Z. J. The Impact of Teacher Comments on the Improvement of English Writing Skills: An Experimental Study. *Journal of Shanxi Agriculture University*, 2007, (5): 78-80.
<https://doi.org/10.13842/j.cnki.issn1671-816x.2007.s1.029>
- [13] Liu, B. Chinese University Students' Attitudes and Perceptions in Learning English Using ChatGPT. *International Journal of Education and Humanities*, 2023, 3(2): 132-140. <http://i-jeh.com/index.php/ijeh/index>
- [14] Liu, G. & C. Ma. Measuring EFL Learners' Use of ChatGPT in Informal Digital Learning of English Based on the Technology Acceptance Model [J]. *Innovation in Language Learning and Teaching*, 2023, (1): 1-12.
<https://doi.org/10.1080/17501229.2023.2240316>
- [15] Mao, Y. S., Y. H. Wang & Y. R. Xing. An Empirical Study on the Use of ChatGPT for Providing Feedback on High School English Writing. *Educational Measurement and Evaluation*, 2024, (1): 3-11.
<https://doi.org/10.16518/j.cnki.emae.2024.01.001>
- [16] Sugita, Y. The Impact of Teachers' Comment Types on Students' Revision. *ELT Journal*, 2006, 60(1): 34-41.
<https://doi.org/10.1093/elt/cci079>
- [17] Sun, P. P., & B. Mei. Modeling Preservice Chinese-as-a-Second/Foreign-Language Teachers' Adoption of Educational Technology: A Technology Acceptance Perspective [J]. *Computer Assisted Language Learning*, 2022, (4): 816-839.
<https://doi.org/10.1080/09588221.2020.1750430>
- [18] Ying, Y. H., J. L., Chen, & B. J., Huang. Reshaping the Educational Ecology: The