

Research on College English Writing Based on Peerceptiv Peer Review System

DU Jie^{[a],*}

^[a] School of International Studies, Jingdezhen Ceramic University, Jingdezhen, Jiangxi, China.

* Corresponding author.

Supported by Jingdezhen Ceramic University Teaching Reform Program: “Research on Online and Offline College English Writing Teaching Based on Peerceptiv System”, No. TDJG-21-Y39; also supported by “Research on Strategies of Enhancing the City Image of Jingdezhen through Official Short Video Platforms”, No. 20224GY008-24.

Received 12 July 2024; accepted 26 September 2024

Published online 26 October 2024

Abstract

This research conducted a writing experiment based on the Peerceptiv peer review system, collecting the first and second drafts of 47 non-English majors based on one topic, as well as the peer review comments and grades, the reliability of students' evaluations, and average grades on each dimension from the Peerceptiv platform. Through a comparative analysis of the mean grades of students' first and second drafts in various dimensions, it was noticed that students' writing skills showed a slow upward trend through the multi-peer reviewing mode. Students' second draft showed better quality than the first one. However, the overall rating reliability of the grades in various dimensions of the two drafts was low. Compared with the first draft, the credibility of the second draft's peer review improved significantly in several dimensions. The reason lies in the students' limited English proficiency and their negative attitudes resulting from relying solely on peer evaluation and grading mechanisms. Therefore, the combination of “teacher's personalized guidance + online peer review + teacher evaluation” can maximize the advantages of the Peerceptiv platform and help writing teaching.

Key words: Peerceptiv; Peer Review; Rating Reliability; College English Writing

Du, J. (2024). Research on College English Writing Based on Peerceptiv Peer Review System. *Studies in Literature and Language*, 29(2), 29-36. Available from: <http://www.cscanada.net/index.php/sll/article/view/13632>
DOI: <http://dx.doi.org/10.3968/13632>

1. INTRODUCTION

Peerceptiv is a computer-assisted automated assessment intelligent system initially developed by the University of Pittsburgh in the United States in 2002 and first applied to writing teaching. Currently, this system has been utilized in thousands of English writing courses in countries such as the United States, Canada, Australia, South Korea, Singapore, and Estonia (Zhang, Di, & Schunn, 2016). Xu Jinfen conducted an empirical study on the application of the Peerceptiv system in writing from aspects including research subjects and environments, research content, peer evaluation results, influencing factors of peer evaluation, and attitudes and perceptions of teachers and learners. She proposed that Peerceptiv has great application prospects based on expanding research environments and subjects, improving research methods, and broadening research perspectives (Xu & Zhu, 2019). Gao Ying conducted research on the effectiveness of peer feedback in writing using Peerceptiv (Gao, et al, 2018), He Jiajia conducted research on personalized tutoring in writing with this system (He, 2019). Zhang Fuhui conducted a comparative study on the writing learning effects of Peerceptiv, iWrite, and the QQ platform (Zhang, Li, Long, & Gao, 2019). All these empirical studies have concluded that Peerceptiv significantly enhances writing efficiency.

Schunn believes that Peerceptiv's effective application is attributed to its unique four core design elements: anonymous peer evaluation, structured evaluation dimensions, feedback on evaluations (re-evaluation), and evaluation accuracy assessment (Schunn et al, 2016).

Firstly, anonymous peer evaluation refers to the process where all writers are evaluators in an anonymous state during the peer evaluation phase after submitting writing tasks online. This allows evaluators to be free from interpersonal pressure and to provide the most objective and authentic evaluations of their peers' compositions. Secondly, structured evaluation dimensions of the Peerceptiv system are various detailed and structured evaluation dimension sheets embedded with it, based on which evaluators can provide targeted revision suggestions, thereby enhancing the effectiveness of peer evaluation. Additionally, teachers can set their own evaluation dimensions before assigning tasks, taking into account students' language proficiency, majors, and topic types. Student evaluators provide detailed revision suggestions and assign scores based on the preset dimensions. Therefore, this system offers great flexibility. Thirdly, feedback on evaluations (re-evaluation) refers to the process where writers provide evaluative feedback and scores in response to their peers' revision suggestions and scores for their own writing. This involves whether the writer believes the suggestions given by their peers are reasonable, pertinent, objective, and useful. This aspect is an important indicator for evaluating whether peer evaluations are pertinent and useful. Fourth one is evaluation accuracy assessment. The system backend comprehensively calculates the average score for each composition across all evaluators for each evaluation dimension. Evaluators whose scores deviate significantly from the average receive low credibility scores, while those whose scores are close to the average receive high credibility scores. Moreover, evaluation accuracy is also used as an evaluation indicator and is one of the factors considered in calculating each student's final score. Therefore, to increase their scores in this section, students will also carefully and objectively evaluate the tasks assigned to them on the platform. The four elements of the Peerceptiv system work together to enhance the reliability of peer evaluation.

2. THE WRITING EXPERIMENT PROCESS AND EVALUATION RUBRICS BASED ON PEERCEPTIV SYSTEM

The number of students participating in the writing experiment based on the Peerceptiv system this time is 47. They are second-year university students majoring in science and engineering but not in English. They possess good reading abilities, yet they have not received specialized writing training. Writing skills instruction is generally integrated into reading and writing courses. However, due to the limited total class hours, the teaching arrangement for writing is correspondingly

restricted in time. Therefore, students' writing and expression abilities are slightly inferior to their reading abilities, but they have strong comprehensive learning abilities. Students in the experiment have no experience using peer review system. Before the writing experiment, they displayed a proactive and positive attitude, expressing their willingness to participate fully and complete tasks at every stage. The writing experiment process is as follows:

Phase One: Teachers train students to use the Peerceptiv platform (the Peerceptiv system interface is entirely in English) and create individual student accounts.

Phase Two: Teachers explain writing techniques and elaborate on the detailed scoring criteria based on the current writing task, then assign the essay topic "Family Education" on the platform. Students accept the task, complete their first draft, and submit it to the platform.

Phase Three: The platform randomly assigns each essay to three other students' accounts. Peer reviewers read and evaluate the three assigned essays, write comments according to the evaluation criteria, and grade the draft. The platform automatically feeds back the review content to the writers' accounts.

Phase Four: Based on the suggestions given by the three peers and what they have learned from reviewing three of their peers' essays, writers revise their first draft and submit the second draft to the platform. At the same time, writers evaluate whether the three peers who reviewed their work provided fair, objective, and useful suggestions and reevaluate the viewers work. The evaluation also includes both commenting and scoring.

Phase Five: The platform continues to assign each second draft to three other students' accounts (different from the first draft reviewers). Peer reviewers read and evaluate the three assigned second drafts, provide comments, and grade the second draft. Phase Six: Students evaluate the second drafts in return. Phase Seven: Teachers compare and evaluate students' first and second drafts, providing feedback.

Teachers interview students with abnormal data to understand their thought and practice during the writing. Teachers download comprehensive data from the backend to review and analyze the accuracy of student scoring and the final scores of each student. The specific process is illustrated in Figure 1.

The writing experiment lasts for approximately 2 months, with each phased task taking approximately 10 days. Students evaluate and score their peers' work based on four dimensions: Content, Organization, Language, and Mechanics. The first three dimensions are further subdivided into three aspects each, with a scoring range of 1-7 for each aspect. Additionally, students objectively describe issues in their peers' essays based on problem prompts for each dimension. The specific descriptions and weights of the criteria are outlined in the Table 1:

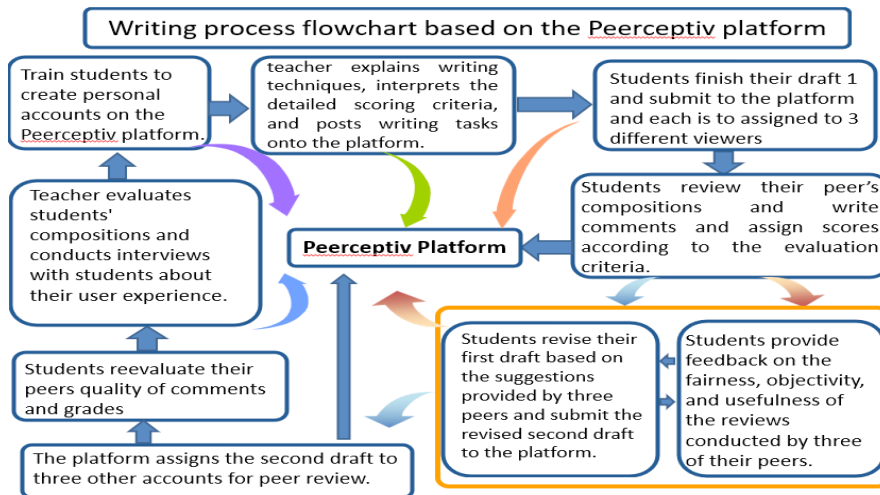


Figure 1
Writing process flowchart based on the Peerceptiv platform

Table 1
Review Rubrics

No.	Dimensions	Description of dimension
1	Content	<p>Is there a thesis statement on The Topic in the essay? Is the thesis statement well supported? Are there sentences or details that do not support the thesis statement and therefore should be eliminated or rewritten? Be specific about your suggestion. Rate ABC from 1-7 A.Thesis: Is the thesis statement about The Topic stated clearly? #weight=1 B.Thesis: Is the thesis adequately supported in each paragraph? #weight=1 C.Relevance: Are relevant personal experiences developed regarding family education? #weight=1</p> <hr/> <p>COHERENCE:Do transitional words and phrases help make the sentences and paragraphs coherent? Please be specific about your suggestion.Rate ABC from 1-7</p>
2	Organization	<p>A.Coherence: Are the paragraphs organized coherently? #weight=1 B.Coherence: Are the sentences in each paragraph organized coherently? #weight=1 C.Cohesion: How well are transitional words and phrases used? #weight=1</p> <hr/> <p>Is the essay concise? Are there a variety of sentence structures in the essay? Are there any errors regarding gender, tense, agreement, comparative construction, non-predicate verb, parallel structure, inconsistent point of view in the essay?Please describe the type of problem including the location of one instance and how they should be fixed. Rate ABC from 1-7</p>
3	Language	<p>Word Choice: Words convey the intended message in a specific, precise, and engaging manner. #weight=1 B.Word Count: Does the essay meet the requirement of word tokens? #weight=1 C.Sentence Structure: Are there a variety of sentence structures in the essay? #weight=1</p> <hr/> <p>Are there any common problems in terms of punctuation, spelling, capitalization, paragraphing in the essay? Please describe the type of problem including the location of one instance and indicate how it should be fixed.</p>
4	Mechanics	<p>Writing Convention: Standard writing conventions (e.g., spelling, punctuation, capitalization, paragraphing) are used effectively to enhance readability in the essay. #weight=1 Rate from 1-7</p>

3. RESULTS AND DISCUSSION

3.1 Analysis of average scores on various dimensions for first and second drafts presented in bar chart

After two rounds of peer evaluation, bar charts showing the average peer rating for each dimension of the first and second drafts and overall rating reliability of the two drafts were obtained from the platform.

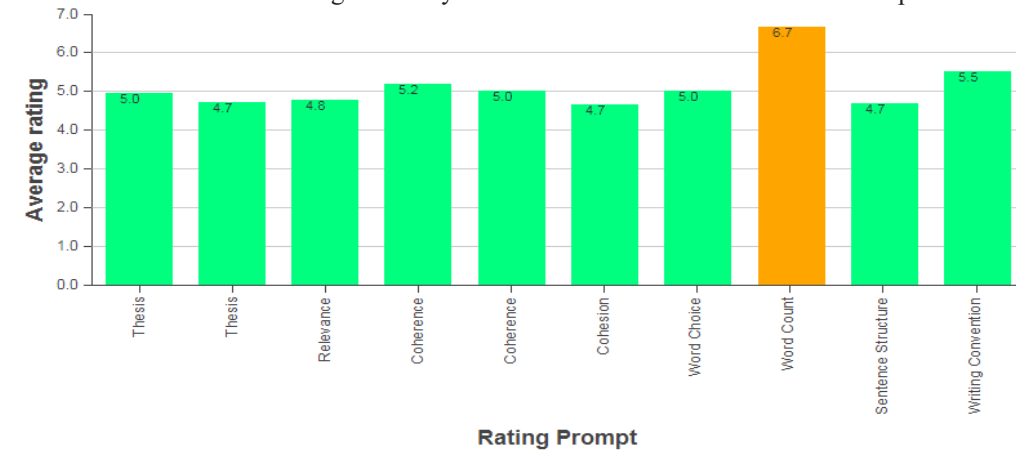


Figure 2
Average peer rating for each dimension in draft 1

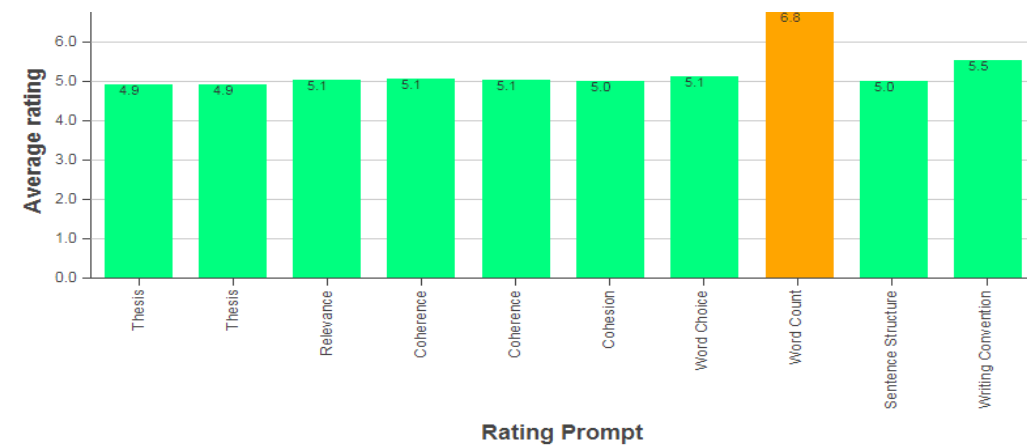


Figure 3
Average peer rating for each dimension in draft 2

Table 2
Comparison of average score in each dimension from the two drafts and percentage of points obtained

Dim.	Thes.	Relev.	Coher.1	Coher.2	Cohes.	Word choice	Word count	Sentence structure	Writing tradition
Dra. 1	5.0	4.8	5.2	5.0	4.7	5.0	6.7	4.7	5.5
Dra. 2	4.9	5.1	5.1	5.1	5.0	5.1	6.8	5.0	5.5
Average score	4.95	4.95	5.15	5.05	4.85	5.05	6.75	4.85	5.5
Scoring rate of the average %	70.7	70.7	73.6	72.1	69.3	72.1	96.4	69.3	78.6

From analyzing the overall data trend, there is a relatively small difference in scores between the first and second drafts across various dimensions, but there is an overall upward trend. Except for a slight decrease

of 0.1 in the average scores for thesis statement clarity (“Is the thesis statement about the topic stated clearly?”) and paragraph coherence (“Are the paragraphs organized coherently?”), both of which remained at 5.5, with

writing conventions remaining unchanged, there were improvements in every other dimension. Among them, four items showed an increase of more than 0.2 points. The largest increases were observed in “relevance,” “cohesion,” and “sentence structure,” all of which rose by 0.3 points. This indicates that the quality of the second draft, formed through two rounds of peer evaluation and revision, has improved. Students have selected reasonable and appropriate argumentative materials to effectively support their thesis statements, used cohesive words more accurately, and attempted to use complex and diverse long sentences to replace the simpler sentences that appeared frequently in the first draft.

However, looking at the percentage obtained from average score in each dimension of both the first and second drafts, except for the word count scoring item, which had a score rate of 96.4%, the overall score rates were above the passing mark, falling into the middle-to-upper range.

One of the reasons for this is that in this writing experiment, the first and second drafts of students' essays on a given topic were evaluated and provided feedback by six different reviewers over two rounds. Specific and detailed suggestions for revision were given across various dimensions. For some deep-level issues in the essays, different peer reviewers often provided different interpretations and modification suggestions. These numerous comments from different perspectives could stimulate the writers' cognitive level and broaden their limited horizons, thereby benefiting the revision of the texts. Gao Ying's research (2018) found that “there is a positive correlation between the improvement rate of text revision and the total number of comments received. The more comments a writer receives, the higher their revision rate.” In other words, there is a positive correlation between the number of feedbacks on students' essays and the improvement rate of text revision. Therefore, these multiple peer feedbacks are crucial for text revision and can have a positive impact. Especially when the authors and reviewers are in an asymmetrical state of English proficiency, where the authors have lower English proficiency and the reviewers have higher English proficiency, the number of comments is the largest, the most detailed, and the suggestions for revision are the most constructive, thus most beneficial for text revision. Pathan's research on the impact of language proficiency on feedback and text revision also pointed out that high-level authors can raise valuable questions, provide solutions, and focus on high-level writing issues (content, logic, etc.) when providing peer feedback (Patchan, 2013). Therefore, for low-level authors, they benefit more when paired with high-level authors for peer review. The specific reason is that the feedback they receive is comprehensive and in-depth. Thus, the asymmetrical combination of language proficiency is optimal.

The following is evaluation examples of four high-level reviewers towards low-level authors from platform data for this writing experiment:

Example 1, comment from content dimension: “First of all, there is no necessary relationship between the title of the article “a happy family” and the topic assigned “family education”. Secondly, your thesis is how to get a happy family but supporting details are too limited in your body part, one more point needs to be added to the part. Thirdly, the relationship between good family and the influence of family education on children's development is not stated in your article. The last phrase “Thank you for your listening” should be eliminated because this is a composition but not a speech”

Example 2, comment from content dimension: “In the last paragraph, the example is irrelevant to your argument that the ideal education is to run with love, and is too brief without details. You need to restate your ideal education and fully justify it.”

Example 3, comment from language dimension: There are several basic grammatical mistakes in the essay, for example, in Sentence 3, Paragraph 1, the expression “the three views” should be replaced with “Three outlooks” (with first letter capitalized). In Sentence 4, Paragraph 2, “because “is not properly used. Replace it with if/when.”

Example 4, comment from organization dimension: “Limited transitional words and phrases are used in the essay, there are obvious jumps between sentences. In order to make the sentence (In my imagination) in paragraph 4 more formal, I think it can be changed to (In my opinion or In my view)”.

In all four examples, the reviewers pointed out specific issues and provided concrete suggestions for revision. When teachers reviewed the original texts written by the authors, they also identified the described problems, and the revision suggestions given by the reviewers were basically consistent with those given by the teachers. Therefore, if a low-level writer can benefit from the evaluations of six high-level reviewers in two rounds of peer review for writing, their final level of performance on the writing task will surely improve significantly. The upward trend shown in the chart should be attributed to the facilitation of effective multi-dimensional and multi-reviewer evaluations.

The second reason lies in the fact that the process of students reviewing their peers' compositions is also a learning process, which helps students produce high-quality texts in their second drafts. Each student will read compositions from six different peers in two rounds of peer review, and since these compositions have the same topic but different authors, there will inevitably be significant differences in their themes, word choices, arguments, and sentence structures. Students need to read closely the texts in order to grade and review them, so the review process is a critical and deep learning process. Especially when the manuscript being reviewed

is from a high-level student and the reviewer is a low-level student, forming an asymmetric pair, the low-level reviewer will naturally draw inspiration from certain advanced vocabulary and the usage of complex, expressive long sentences in the reviewed manuscript.

They may also imitate a unique argumentative perspective. These imitations and inspirations are part of the process of internalizing learning, which can help students, especially low-level students, develop high-quality second drafts.

3.2 Analysis of the bar chart for the mean value of overall credibility scores for the first and second drafts across various dimensions

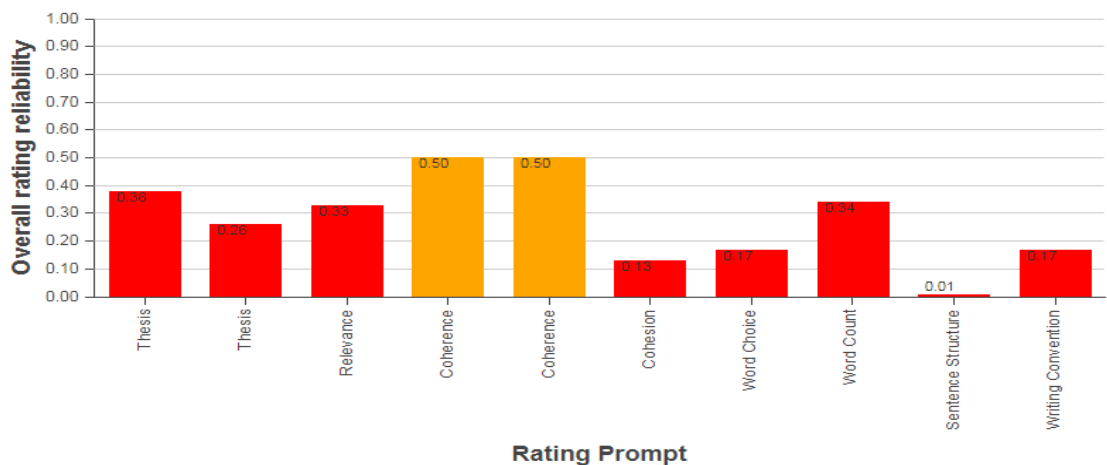


Figure 4 Overall rating reliability in draft 1

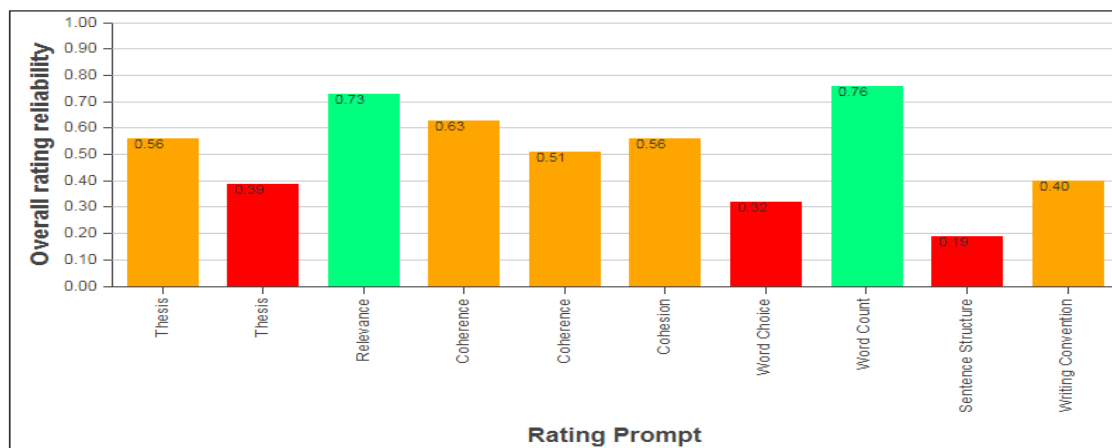


Figure 5 Overall rating reliability in draft 2

Table 3 Comparison of average scores, differences, mean values, and mean score rates across various dimensions in reliability between the first and second drafts

Dim. Dra.	Thes. 1	Thes. 2	Relev.	Cohe.1	Cohe.2	Cohes.	Word choice	Word count	Sentence structure	Writing trad.
Dra. 1	0.38	0.26	0.33	0.50	0.50	0.13	0.17	0.34	0.01	0.17
Dra. 2	0.56	0.39	0.73	0.63	0.50	0.50	0.32	0.76	0.19	0.40
D-value	0.18	0.13	0.40	0.13	0.00	0.37	0.15	0.42	0.18	0.23
Aver. Sco.	0.46	3.25	0.53	0.565	0.5	0.315	0.245	0.55	0.1	0.285
Scoring rate of the average %	46	32.5	53	56.5	50	31.5	24.5	55	10	28.5

From the perspective of the overall reliability of scores for the first and second drafts, as well as the comparison of their values across various dimensions, the mutual evaluation reliability of the first draft shows a downward trend. In particular, the reliability of scores for sentence structure is almost zero, while the two dimensions of coherence are at intermediate values. The overall reliability of scores for the second draft rebounds, with correlation and word count scoring 0.73 and 0.76 respectively. Among them, there are three dimensions with differences exceeding 0.37, indicating an improvement in the effectiveness of peer evaluation across multiple dimensions. However, the overall evaluation effectiveness is below the passing score. Through monitoring the entire experiment and analyzing platform data, the following reasons have been identified:

Firstly, the mechanism of relying solely on peer evaluation led to a relatively passive attitude among students. Prior to this writing experiment, the Peerceptiv system did not include teacher's involvement in evaluating the first and second drafts of students' work, and students' grades and other data were generated solely from interactions between students. Except for initial offline training, the teacher's involvement in the Peerceptiv system was limited to supervising and urging students to accept and submit tasks to the platform at

various stages. The teacher evaluation process took place outside the platform, where teachers downloaded the first and second drafts of students' work and the final scores generated by the platform's statistics for overall evaluation and grading. This lack of teacher involvement in the platform evaluation process led to the emergence of negative emotions among some students. In interviews after the experiment, many students expressed that the language ability of the review partners randomly assigned by the system was insufficient, making them unqualified reviewers who could not provide effective suggestions. Therefore, their comments were not referential, and their grading was unfair. They also did not put their full effort into evaluating their peers compositions. Kaufman & Schunn believed that when teachers are involved in evaluating writing tasks, students are generally more active in completing them, while in situations where only peer evaluation is conducted, students tend to be perfunctory (Kaufman et al., 2011). Lu Lu also emphasized that the self-awareness and initiative of students as the subjects of writing and evaluation play a crucial role (Lu, 2016). Subsequently, when reviewing the platform data, the author indeed identified several students who provided overly simplistic evaluations, as shown in the following figure:

beige-428: content	16612	1	You thesis statement is clear and well argued.
beige-428: content	16612	1	What do you think of the last paragraph as a summary?
beige-428: Organization	16669	2	The sentences in each paragraph are organized coherently.
beige-428: language	16743	3	A wide variety of sentences structures are used in the essay.
beige-428: Mechanics	16793	4	No problems of punctuation, spelling, paragraphing are found in the essay.

Figure 6
Screenshot 1 of student's evaluation comment of the peer assigned

gold-7718: content	16612	1	you can probably use more advanced vocabulary and advanced sentences to make your essay look more advanced.
gold-7718: Organization	16669	2	no
gold-7718: language	16743	3	no
gold-7718: Mechanics	16793	4	good

Figure 7
Screenshot 2 of student's evaluation comment of the peer assigned

The comments in Figures 6 and 7 are simple and general, and even pose a question to the writer without providing constructive suggestions. Therefore, unlike other carefully given evaluations, such comments are not referential and reduce the reliability of the grading. An effective solution is for teachers to participate throughout the process, incorporating teacher's evaluations into the dimensions of students' composition grades and increasing their weight. This can effectively change students' passive attitudes to positive ones, encouraging them to put their full effort into both writing and reviewing.

Secondly, students' language proficiency is the fundamental reason that restricts their ability to provide effective peer evaluations. Bai Liru believes that students' expressive abilities, linguistic competencies, and other factors contribute to anxiety in English writing among college students in China, which subsequently has a negative impact on their English writing skills (Bai, 2021). This anxiety stems from two aspects, namely the dual roles students play in writing experiments. When students participate in writing tasks using the Peerceptiv system, they are both writers and reviewers, and the

effectiveness of their reviews is a component of their final grade. Therefore, on the one hand, students doubt their own limited English proficiency and inadequate grammatical skills, believing they lack the ability to objectively, fairly, reasonably, and accurately evaluate their peers' compositions during peer evaluation. On the other hand, students also complain that their peers' proficiency is limited, their evaluations are too superficial, and the suggestions provided are not valuable for reference. Especially when a student of moderate proficiency has high expectations for peer evaluations but is assigned three low-proficiency peers by the system, the effectiveness of their evaluations will naturally be lowered.

Therefore, it is believed that although modern educational technology is advanced and data acquisition and analysis are convenient, without teachers' guidance and explanations, as well as without teachers' purposeful instructional design and practice for students' writing abilities, the shortcomings of the tools become more prominent, and the advantages of the Peerceptiv system are bound to be limited. The optimal combination mode is "individualized teacher guidance + online peer evaluation + teacher evaluation", allowing students to receive systematic writing training and develop good writing abilities first. With teachers' guidance as a prerequisite, using the Peerceptiv system for further training will lead to greater improvements in writing effectiveness.

REFERENCES

Bai, L. R. (2021). Research on the Inducing Factors and Manifestation Characteristics of English Writing Anxiety Among College Learners. *Foreign language and Literature*, 37(5), 129-138.

- Gao, Y., Zhang, F. H., Zhang, S. J., & Schunn, C. D. (2018). Research on the Effectiveness of Peer Feedback in English Writing Based on the Peerceptive Peer Review System. *Technology Enhanced Foreign Languages*, 180(4), 3-9.
- He, J. J. (2019). Research on Personalized Tutoring Mode for Academic English Writing Based on Peerceptive Online Peer Review System. *Technology Enhanced Foreign Languages*, 186(4), 25-33.
- Kaufman J. H., & Schunn C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, (3), 13-19.
- Lu L. (2016). Research on the Second Writing Process Based on Automated Evaluation Systems. *Foreign Language World*, 173(2), 88-96.
- Patchan M. M. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, (2), 381-382.
- Schunn C. D. Godley A., DeMertino S. (2016). The reliability and validity of peer review of writing in high school AP English classes. *Journal of Adolescents and Adult literacy*, 173(1), 112-113.
- Xu, J. F., & Zhu, Q. (2019). Research on English Online Peer Review Based on the Peerceptive Peer Review System. *Technology Enhanced Foreign Languages*, 186(4), 10-16.
- Zhang, F. H., Di, Y. H., & Schunn, C. D. (2016). Research on the Use of Peerceptiv Peer Review System to Assist Students in English Writing. *Technology Enhanced Foreign Languages*, 170(8), 38-44.
- Zhang, F. H., Li, W. T., Long, M. Y., & Gao, Y. (2019). Comparative Study on the Effects of Self-Regulated Learning in Writing Based on Three Technological Platforms. *Technology Enhanced Foreign Languages*, 189(10), 22-26.