

ENHANCING EFL WRITING THROUGH PEER ASSESSMENT INTEGRATED WITH AI-BASED AUTOMATED CORRECTION: A BLENDED LEARNING INTERVENTION

QIPIN CHENG¹, RENCHUAN ZHANG², YUJIE LIU³,

¹SCHOOL OF HUMANITIES AND SOCIAL SCIENCES, SHANGHAI LIDA UNIVERSITY, SHANGHAI,
201609, CHINA, Email: chengqipin0109@lidapoly.edu.cn

²SCHOOL OF HUMANITIES AND SOCIAL SCIENCES, SHANGHAI LIDA UNIVERSITY, SHANGHAI,
201609, CHINA, Email: zhangrenchuan@lidapoly.edu.cn

³SCHOOL OF NURSING, SHANGHAI LIDA UNIVERSITY, SHANGHAI, 201609, CHINA, Email:
liuyujie0218@lidapoly.edu.cn

Abstract: While automated writing evaluation (AWE) systems provide granular feedback on linguistic features, their capacity to address thematic depth and organizational coherence remains limited. This quasi-experimental study proposed a triadic feedback model integrating AWE (iWrite platform), structured peer assessment, and teacher scaffolding. Seventy-one Chinese university students were assigned to an experimental group (AWE + peer assessment) or control group (teacher feedback only). Pre/post-tests assessed writing proficiency using CET-4 rubrics, supplemented by Likert-scale questionnaires and textual analysis. Results indicated: (1) Significant gains in the experimental group's overall writing scores ($d^* = 1.24$, $p^* < .001$); (2) Improved student engagement in revision cycles (97.2% implemented peer suggestions); (3) Enhanced metacognitive awareness of text organization. The intervention demonstrates how technology-mediated peer assessment can augment cognitive and social dimensions of writing development.

Keywords: Peer assessment; EFL writing; automated writing evaluation; blended learning

Funding:

This work was supported by Shanghai Educational Science Research Project: Research on Peer Assessment Design based on Intelligent Writing Marking Technology in Blended Learning Environment (Grant No.C2025282) .

INTRODUCTION

Peer evaluation, as a vital pedagogical tool, has garnered significant attention and research in recent years, particularly within English writing education. Numerous scholars have conducted empirical studies on the effectiveness of peer review mechanisms. He Jiajia (2019) utilized Peerceptiv, an online peer review

system, to investigate its impact on English writing instruction among 53 college students. The study objectively assessed peer evaluation methods and proposed targeted recommendations, such as combining online peer reviews with offline teacher guidance to enhance students' writing proficiency. Zhang Jing (2018) demonstrated that teacher-supervised peer evaluations significantly improved students' English writing skills, with teachers serving as scaffolding agents. Zhao Jingui (2021) focused on a "self-evaluation report + group evaluation report" model for peer review implementation.

In recent years, computer science and information technology-based essay evaluation systems have become increasingly vital in English writing instruction. While these systems provide detailed feedback on linguistic expression, they notably lack comprehensive assessment of ideological content and structural coherence. To address this gap, this study investigates how peer review mechanisms in online essay evaluation systems influence learners' English writing proficiency and capabilities through analyzing feedback regarding ideological substance and textual organization. The findings aim to offer practical insights for enhancing college-level English writing pedagogy.

Research Purpose

This study aims to verify the effectiveness of a blended teaching model for college English writing based on peer review mechanism through empirical comparative research. Specific research questions: Can the peer review mechanism fully leverage students' initiative? Can the peer review mechanism effectively improve students' writing proficiency?

Research subject

This study focuses on two classes (Class 1 and Class 2) of the Big Data Science and Technology program at Shanghai Lida University, with a total of 71 students evenly split between male and female students. Both classes share comparable English proficiency levels upon enrollment in the same major. As the English instructor for both classes, the researcher maintained consistent teaching materials, content, schedule, and methods. The key distinction was the experimental design: one class was designated as an "experimental group" for writing instruction using a peer review system combined with teacher feedback, while the other served as a control group following the traditional single-phase evaluation method.

Research process

The experimental period of this study was 16 weeks (From February 28 to June 18, 2025), which included a week of pre-test, training, post-test and questionnaire survey, as well as three specific writing processes lasting 12 weeks.

Pre-experiment Weekly Assessment. During the pre-experiment week, teachers selected the 2023 June National College English Test Band 4 essay topic "How Will Our Lives Continue Without the Internet" as a classroom-based pre-test composition. Students were required to write an essay of approximately 120 words within half an hour. The essays were manually graded according to the Band 4 College English Test scoring criteria, with a full score of 15 points. Each essay's final score was calculated as the average of two graders' evaluations. Results showed that the experimental class scored an average of 7.24 points, while the control class averaged 7.31 points, with an internal consistency coefficient of 0.899. There was no significant difference in English writing proficiency between the two classes ($P > 0.05$). Detailed scoring criteria and peer review standards are shown in Table 1.

Table 1 Scoring criteria and rules

word number	Does the article meet the word count?
-------------	---------------------------------------

content	Is the article on topic and full of content? Does the article have a prominent central idea, a clear theme, and is it reasonable?
structure	Is the structure of the article reasonable and logical? Is the context of the article coherent and is the use of conjunctions appropriate?
words and phrases	Are the words correctly spelled, capitalized and used correctly? Is the sentence pattern diversified in the article? Are the words and collocations appropriate and accurate?
comprehensive evaluation	merit not enough propose

Peer Evaluation Training Week. First, teachers introduce the peer evaluation mechanism and its advantages to experimental classes, emphasizing how collective participation, mutual review, learning from each other, and leveraging strengths can fully engage students in academic tasks. Second, based on established scoring criteria, teachers select exemplary, good, and poor-grade model essays for classroom analysis, guidance, and discussion, providing practical feedback on peer evaluation methods. Third, teachers conduct mock peer evaluations using pre-test materials, addressing students' specific questions through detailed explanations. Finally, one class is randomly selected as an experimental group, organized according to the principles of inter-group homogeneity, intra-group heterogeneity, and complementary strengths.

Experimental Process. Throughout the experiment, teachers assigned three writing tasks to two classes with identical themes, all extending from College English 3 classroom content. Both classes maintained equal writing frequency, requiring students to produce a second draft of the same text. The experimental class was required to conduct peer review after completing the first draft, revise based on collected feedback, and submit a second draft. In contrast, the control class wrote only one initial draft, received immediate peer feedback, revised according to teacher guidance, and submitted their final draft. Both groups' final compositions were uniformly evaluated by two instructors.

Post-experiment assessment. After the experiment, the teacher asked students again to write a class essay titled "How Will Our Life Go On Without Internet" within 30 minutes. By comparing the pre-test and post-test scores, the teacher summarized the effectiveness of the peer evaluation mechanism model.

Post-experiment questionnaire survey. To further understand students' attitudes and perceptions of peer evaluation mechanisms, the research team had students in the experimental class complete questionnaires after participating in peer evaluation activities. The questionnaire adopted a Likert five-point scale with "1-5" representing responses from "strongly disagree" to "strongly agree". The questionnaire achieved a 100% response rate, with an effective completion rate reaching 97%.

Data collection and analysis

After a semester of experiments, the specific data collected and analyzed by this study are shown in Table 2. Table 2 summarizes the statistical data of the pre-test scores of the two classes, and the specific analysis is as follows.

Table 2 Comparison of pre-test scores between control class and experimental class

word number	content	structure	words and phrases	total points
-------------	---------	-----------	-------------------	--------------

	CC	EC	CC	EC	CC	EC	CC	EC	CC	EC
group										
M	51. 87	52. 23	46. 36	48. 25	50. 37	51. 14	53. 65	55. 57	51. 13	51. 76
S.D	0. 94	0. 96	0. 74	0. 98	0. 55	0. 69	1. 03	0. 60	0. 70	0. 79
t	-1. 436		-2. 372		1. 369		2. 184		2. 0326	
Sig.	0. 048		0. 000		0. 019		0. 118		0. 073	

In terms of word count, the experimental class scored slightly higher than the control class by 0.3484 points, with the Sig. value indicating this difference is statistically "highly significant" ($p < 0.01$). This suggests that students in both classes had nearly identical word counts. While the experimental class outperformed the control class in other dimensions like "content", "structure", and "word choice" compared to word count metrics, the Sig. values showed these differences didn't reach statistical "highly significant" levels ($p < 0.01$). This demonstrates that both classes showed comparable performance outside the word count dimension. Therefore, this study concludes that the overall academic performance of both classes is relatively close, making them suitable subjects for this experiment. The semester-long experimental results analysis between the two classes is shown in Table 3.

Table 3 Comparison of comprehensive scores between control class and experimental class

Test order	First test				Second test	
Third test						
Group	EC	CC		EC	CC	EC
CC						
average value	85. 3	80. 8	86. 4	81. 1	86. 2	78. 3
standard error	2. 76	2. 54	2. 42	2. 24	2. 65	2. 63
Sig.	0. 0012			0. 0013		0. 0013
(bilateral)						

In Table 3, the experimental class demonstrated outstanding performance under peer evaluation incentives, achieving an average score of 85.9 points across three compositions—a 7.3 percentage point improvement over the control class's average score of 80.0 points, indicating significant effectiveness. This phenomenon is further reflected in the Sig. values of the three tests, all reaching the statistically "highly significant" level ($p < 0.01$). These findings demonstrate that the flipped classroom model based on peer evaluation proposed in this study has effectively stimulated students' learning motivation and achieved notable academic outcomes.

RESULTS AND DISCUSSION

Based on the data collected in one semester, this study conducted a questionnaire survey in the 16th week. The results will be discussed from three aspects: students' attitude towards peer evaluation, the effectiveness of peer evaluation and reflection on blended classroom.

1. Analysis of students' attitude towards peer evaluation

In traditional English writing classrooms, students are passive recipients isolated from social interactions, lacking autonomy in composition, reader awareness, and writing motivation. Their compositions tend to be monotonous with repetitive vocabulary. Student feedback indicates that peer review mechanisms provide opportunities for collaborative communication and highly interactive exchanges, establishing a two-way feedback dialogue model. Through comparative analysis of these three writing tasks, this study found that 97.2% of students actively revised their essays based on peer feedback, 95.7% effectively improved structural issues in their compositions, and 94.3%

comprehensively addressed language problems by incorporating peer suggestions while consulting dictionaries and reference materials.

In addition, in order to better understand students' attitudes and perceptions of the effects of precision teaching under peer evaluation, teachers asked students in the experimental class to fill in questionnaires before and after peer evaluation activities, as shown in Table 4.

Table 4 Student attitude questionnaire survey

project	Pre-test	Post-test	Standard error values	t	Sig.
Do you want to correct your own articles for your classmates?	2. 93	3. 53	0. 16	-3. 01	0. 001
Do you think the score given by your classmates is fair?	3. 38	3. 83	0. 17	-5. 57	0. 000
Do you have the confidence to give correct marks to your classmates' essays?	3. 04	3. 78	0. 15	-4. 62	0. 000
Do you think the grades you gave your classmates are fair?	3. 48	3. 91	0. 16	-4. 23	0. 000
Can peer review solve your writing problems?	3. 03	3. 82	0. 15	-3. 52	0. 001

As shown in Table 4, students in the experimental class initially expressed concerns about peer evaluation (pre-test mean 2.93, below the median of 3.00) and lacked confidence in self-assessment (pre-test mean 3.04, close to the median). Through teacher-guided training and practical exercises, peer evaluation recognition significantly improved. Moreover, students demonstrated positive changes in their self-assessment confidence ($p < 0.05$).

2. Students' analysis of the effectiveness of peer evaluation

The effectiveness analysis of peer evaluation by students is shown in Table 5.

Table 5 Effectiveness analysis of peer evaluation

project	mean	standard deviation
Did peer review teach you how to properly capture the main points of a topic?	3. 63	0. 64
Did peer review teach you how to write?	3. 68	0. 54
Does peer review help you better understand what makes a good essay structure?	3. 75	0. 62
Did peer review help you understand which sentences were more appropriate?	3. 78	0. 56
Can peer review help you improve your score quickly?	3. 90	0. 71
Does peer evaluation effectively find out the shortcomings of students' compositions?	3. 71	0. 58
Can peer evaluation help strengthen students' sense of participation?	4. 18	0. 67
Does peer evaluation help to enhance communication and cooperation between students?	4. 13	0. 65

The results of the one-sample t-test in Table 5 indicate that students' average scores across all items in the questionnaire exceeded 3.00 ($p < 0.05$), demonstrating strong confidence in peer evaluation's effectiveness in English writing. Peer review enables comprehensive improvement in various aspects

including essay themes, content, structure, and sentence composition. Notably, it significantly enhances student engagement and collaborative communication, with mean scores of $M=4.18$ and $M=4.13$ respectively. Overall, peer evaluation positively impacts students' English writing by fostering initiative and teaching them to learn from each other's strengths, identify weaknesses, and recognize others' merits. Simultaneously, it helps build rapport, trust, and respect among peers. Throughout the learning process, students become acutely aware of their responsibilities: to lead by example and actively engage in English writing development.

CONCLUSION

This study reveals that peer review through an essay auto-correction system significantly enhances students' English writing proficiency, refines their writing processes, and develops critical thinking skills. Students generally endorse peer evaluation, noting it boosts writing motivation, strengthens confidence, and cultivates reader awareness. When implemented with teachers' guidance, this system compensates for online tools' limitations in assessing content coherence and structural refinement. It guides learners to shift focus from superficial language issues to deeper aspects like text organization and semantic optimization, effectively nurturing critical analysis and analytical reasoning abilities. However, the research also identifies challenges including inconsistent feedback quality and lack of authoritative authority in peer reviews.

In the context of "Internet Plus", exploring and optimizing English writing peer evaluation systems based on essay automatic correction technology has become essential for reforming foreign language writing instruction in China. This study offers valuable insights for writing pedagogy. College English writing courses should leverage modern information technologies to fully utilize the "evaluation-driven improvement" advantages of automated correction systems, enhancing students' practical writing skills while ensuring both quantity and quality enhancement. Future English writing instruction requires comprehensive guidance on peer evaluation methods, intensified training in peer feedback practices, and further utilization of teachers' mediating roles. Through proper peer evaluation guidance and practice, students can provide more effective feedback that facilitates meaningful revisions, maximizing peer evaluation's optimal impact. Beyond focusing on students' foundational knowledge like vocabulary and grammar, educators should introduce common writing strategies and techniques, train logical thinking skills, and strengthen content organization and argumentation abilities. Simultaneously, cultivating students' comprehensive English proficiency and refining writing methodologies will fundamentally improve peer evaluation capabilities and quality, enhance acceptance of peer feedback, and better leverage peer evaluation's role in advancing English writing education.

Acknowledgements:

I would like to express our sincere gratitude to all individuals and institutions that contributed to the successful completion of this research.

Availability of data and materials:

The datasets generated and/or analyzed during the current study are not publicly available due to institutional policies but are available from the corresponding author upon reasonable request.

Funding:

This work was supported by Shanghai Educational Science Research Project: Research on Peer Assessment Design based on Intelligent Writing Marking Technology in Blended Learning Environment

(Grant No.C2025282) .

REFERENCES

1. Cai, J. (2011). A comparative study on peer feedback and teacher feedback in college students' English writing online. *Foreign Language World*, (2).
2. Chen, Y., & Tsai, C. (2009). An educational research course facilitated by online peer assessment. *Innovations in Education and Teaching International*, 46(1), 105–117.
3. Chok, C. M. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–82.
4. Deng, P., & Cen, Y. (2010). Effectiveness of peer evaluation feedback mechanism on the development of Chinese students' second language writing ability. *Foreign Language Teaching*, (1).
5. DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretical foundations and new directions. *Review of Educational Research*, 58(2), 119–149.
6. Gao, Y., Wang, Y., & Christian, D. S. (2019). Research on the adoption of peer feedback comments in English writing and its influencing factors. *Foreign Language Audio-Visual Teaching*, (2).
7. Jiang, Y. (2005). The role of online peer evaluation in the development of writing ability. *Foreign Language Teaching and Research*, (3).
8. Jin, X. (2018). Teaching pathways for cultivating critical thinking ability in college English argumentative writing. *Journal of Jiangxi Normal University (Philosophy and Social Sciences Edition)*, (2).
9. Li, Y. (2021). *A study on the consistency between iWrite's automated scoring and manual scoring—Taking the National Talent Examination as an example of email writing* [Master's thesis, Beijing Foreign Studies University].
10. Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275.
11. Mo, J. (2007). Peer evaluation: Improving college students' writing autonomy. *Journal of PLA University of Foreign Languages*, (3).
12. Pol, J. V. D., Berg, B. A. M. V. D., Admiraal, W. F., et al. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education*, 51(4), 1804–1817.
13. Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1(3), 217–233.
14. Tseng, S., & Tsai, C. (2007). Online peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4), 1161–1174.
15. Wu, Y. (2013). The influence of peer evaluation on self-efficacy: An empirical study based on college English writing. *Shandong Foreign Language Teaching*, (6).
16. Xie, S., Wang, W., & Li, J. (2019). Problems of college students' English writing and teaching and learning strategies from a comparative perspective: Taking Guangdong Second Normal University as an example. *Journal of Huzhou Normal University*, (6).
17. Yu, S., & Lee, I. (2016). Understanding the role of learners with low English language proficiency in peer feedback of second language writing. *TESOL Quarterly*, 50(2), 287–313.
18. Zhang, L., & Sheng, Y. (2015). Case study on feedback effect of automatic essay evaluation system. *Foreign Language and Audiovisual Teaching*, (3).
19. Zhang, S., & Xu, J. (2019). The effect of online peer feedback on writing proficiency of English

learners at different levels from the perspective of ZPD. *Foreign Language and Foreign Language Teaching*, (5).

20. Zhong, C. (2015). Research on autonomous writing revision based on online automatic feedback. *Journal of PLA University of Foreign Languages*, (4).

21. Zou, S., & Pan, M. (2018). The definition of writing competence construct in the Chinese English Proficiency Scale. *Contemporary Foreign Language Research*, (5).
