ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 16, No. 3, pp. 899-910, May 2025
DOI: https://doi.org/10.17507/jltr.1603.20

The Effects of Automated Writing Evaluation (AWE)+Human Feedback on the Quality of Argumentative Writing by Chinese EFL Learners

Yan Zhang
Department of English Language, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Malaysia;
Department of Foreign Languages, Yuncheng University, Yuncheng, China

Ilyana Binti Jalaluddin*
Department of English Language, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Malaysia

Roslina Mamat
Department of Foreign Languages, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Malaysia

Tingting Zhao
International Education School, Yuncheng University, Yuncheng, China

***Abstract*—Nowadays, Chinese EFL learners are increasingly using automated writing evaluation (AWE) to provide feedback on their writing. However, AWE is relatively inadequate for providing elaborate feedback on the global aspects of writing that require human judgment. Thus, how to combine AWE with human feedback is a significant issue worth exploring. Regarding the effects of AWE+human feedback, cohesion and coherence are rarely studied. Thus, this study aimed to address that gap. This research employed a quasi-experimental study with a quantitative method to address AWE+human feedback. It also aimed to compare the effects of traditional feedback with AWE+human feedback modes, that is, teacher-only, AWE+teacher (AT), and AWE+peer+teacher (APT), on EFL learners' writing quality in terms of holistic score, cohesion, and coherence. A total of 90 EFL learners from three intact classes of English major were randomly assigned to the control and experimental groups. The control group received only teacher feedback. The experimental groups received AT feedback and APT feedback in their writing process, respectively. Two instruments, iWrite and Coh-Metrix, were used to collect the data. Results showed that all feedback types improved holistic scores, coherence, and cohesion, with the APT model producing the most significant improvements across these dimensions. The APT group demonstrated particularly high holistic scores. It enhanced sentence-level coherence and lexical cohesion, suggesting that integrating AWE, peer, and teacher feedback provides a comprehensive and effective approach to developing writing proficiency.**

***Index Terms*—AWE+human feedback, writing quality, holistic scores, cohesion, coherence**

I. Introduction

A. Research Problem Statement

Feedback has long been regarded as an effective method for improving students' writing quality (Guo et al., 2024; Cheng et al., 2021; Hyland, 2006; Ferris, 1997). In particular, teacher feedback is vital in improving students' writing abilities (Ferris, 1997). However, providing individualized feedback is often both prolonged and arduous, especially for teachers working with large classes (Xu et al., 2022; Zhai & Ma, 2022; Yu, 2021). This overwhelming workload leads to teachers either avoiding assigning writing tasks (Han & Sari, 2024; Chen et al., 2009; Weigle, 2007) or providing less individualized feedback on students' writing content in order not to shoulder this workload (Chen et al., 2009; Han & Sari, 2024). As a result, students may not receive the comprehensive support they need to improve their writing effectively.

To address this problem, automated writing evaluation (AWE) systems, such as Criterion, e-rater, MyAccess, Write & Improve, Intelligent Essay Assessor (IEA), BETSY, and Write To Learn (Dikli, 2006; Fu et al., 2024), have been developed. These tools offer immediate and personalized feedback on students' written compositions, significantly reducing the teacher's burden of revising and grading assignments (Huang & Chen, 2024; Ranalli, 2018). As a result, AWE systems are increasingly being integrated into English writing classrooms worldwide, demonstrating their potential

* Corresponding Author.

to improve the efficiency of writing instruction (Alharbi, 2023; Hassanzadeh & Fotoohnejad, 2021; Liao, 2016; Stevenson & Phakiti, 2014).

However, despite the advantages that AWE systems provide in terms of efficiency (Barrot, 2023; Wang et al., 2022), they fall short in offering detailed feedback on global aspects of writing, such as organization, coherence, and content—areas that require human judgment (Wang et al., 2022). Consequently, there is a growing consensus that AWE tools should not replace human feedback but should supplement it (Link, 2022).

B. Research Gap

In recent decades, the effectiveness of combining AWE with human feedback has been a hot topic in writing feedback research (Liu & Shao, 2024). Many studies have explored the combination of AWE with either teacher or peer feedback (Fu et al., 2024; Liao, 2022; Liu, 2022) on writing quality. Results showed that the combination of AWE+teacher/peer feedback enhanced writing quality compared to using AWE alone (Han & Sari, 2024; Tan et al., 2023). Furthermore, limited research has explored the impact of blending AWE with teacher and peer feedback. Interestingly, this combination of AWE, peer, and teacher feedback has been proven feasible in writing classrooms (Chen & Gong, 2022; Huang & Zhang, 2014). However, the effect of this feedback on students' writing quality needs to be further explored, especially from the perspectives of cohesion and coherence.

In English as a Foreign Language (EFL) contexts, cohesion and coherence are critical aspects of writing, especially for establishing clear discourse linkages that enhance readability and comprehensibility (Plakans & Gebril, 2017; Crossley et al., 2016; Lee, 2002). Despite the importance of these elements, few studies have examined the effects of different feedback approaches on cohesion and coherence specifically. Therefore, this study aims to investigate the impact of AWE+human feedback on the writing quality of Chinese EFL learners, focusing on holistic scores, cohesion, and coherence. By comparing traditional teacher feedback, AWE+teacher feedback, and AWE+peer+teacher feedback modes, this study seeks to determine which feedback approach most effectively supports students' writing development. To achieve these aims, the following research questions are posed:

RQ1: What are the effects of teacher-only, AWE+teacher, and AWE+peer+teacher feedback on the writing quality of Chinese EFL learners at the tertiary level in terms of holistic scores, cohesion, and coherence?

RQ2: Are there significant differences among the effects of the three feedback modes on the writing performance of Chinese EFL learners at the tertiary level in terms of holistic scores, cohesion, and coherence?

II. Literature Review

Feedback has been a concern of various researchers for centuries (Hyland, 2006). It plays a significant role in writing, and much research has been conducted in this domain (Xu et al., 2022). This section traces and analyzes the development of feedback, automated writing evaluation (AWE), and empirical studies on the integration of AWE and human feedback.

A. Feedback

The term "feedback" was first used in biology to refer to the process in which the output information is returned to the organism to play a role in the re-output of its information. Ramaprasad (1983) defines feedback as prompt information of the gap between the actual level and expected level in system parameters from the perspective of human performance. In writing, feedback is defined as input from the reader to the author (Hyland & Hyland, 2006). With this kind of input, readers can provide the author with information to modify the composition to help the author accurately express their opinions and easily communicate with others. From the perspective of process writing theory, Keh (1990) defines feedback as the information provided by readers to the author for revision, including evaluation, queries, and suggestions.

While different scholars have varied interpretations of feedback, most of them agree that feedback is information that helps learners rewrite and change their work. In the present study, feedback refers to a kind of written information input provided by AWE, peers, or teachers to students' written content, including correction, evaluation, suggesting, and scoring, which aims to help students accurately express their thoughts and opinions to effectively improve their writing proficiency.

B. AWE Feedback

As artificial intelligence continues to advance, there has been a rise in the development of web-based tools designed to provide prompt feedback on student writing. These tools are called Automated Writing Evaluation (AWE) systems. Burstein (2003) defines Automated Writing Evaluation (AWE) as a technology that provides scores (assigning a number to an essay) or evaluation (providing diagnostic feedback) for written works.

Since the first AWE called Project Essay Grade (PEG) was introduced in 1966, AWE has become increasingly popular in education around the world (Fu et al., 2022). The most widely used AWE tools across the globe include Criterion, e-rater, MyAccess, Write & Improve, Intelligent Essay Assessor (IEA), BETSY, and WriteToLearn (Fu et al., 2022). Similarly, Chinese experts have developed their own automated assessment systems in response to the practical needs of English writing instruction in China, such as iWrite and the Juku Pigai website.

Automated writing evaluation (AWE) has been debated among scholars, with varying perspectives on its benefits and limitations. Advocates of AWE argue that it can save teachers time by providing feedback on standard features, freeing them up to focus on higher-level writing qualities such as argumentation, organization, and development (Stevenson,

2014). They also claim that AWE overcomes knowledge barriers and provides a consistent evaluation based on predetermined criteria, ensuring students receive accurate feedback regardless of individual biases (Cotos et al., 2017). Additionally, some students may prefer AWE feedback over feedback from teachers or peers, as it can be less intimidating and more impersonal, improving their affective response to the evaluation process (Fu et al., 2024).

Despite AWE's benefits, critics argue that it may not be able to evaluate some aspects of writing, such as creativity or critical thinking skills that require human judgment (Wilson & Roscoe, 2020). Others point out that AWE may overemphasize grammatical errors at the expense of content and coherence (Warschauer & Grimes, 2008). Some scholars also express concerns over the potential misuse of AWE, such as relying solely on automated feedback without incorporating certain human evaluations (Ranalli, 2021).

Overall, while scholars hold different perspectives on AWE, it has been acknowledged that AWE tools can be used as a supplement (rather than a replacement) for teacher feedback (Link, 2022).

C. AWE+Human Feedback

As previously mentioned, writing feedback research has focused heavily on the efficacy of integrating AWE with human input in recent decades (Liu & Shao, 2024). Some studies have explored the combination of AWE+ teacher mode in the context of writing instruction (Han & Sari, 2024; Liao, 2022; Liu, 2022; Xu, 2020). For example, Liao (2022) found that integrating AWE (Criterion) and teacher feedback resulted in more positive attitudes towards feedback than just using one or the other. Similarly, Xu (2020) compared the effect of AWE (iWrite) and AWE (iWrite)+teacher feedback on the English writing of non-English major college students and found that AWE+teacher was more effective in improving students' writing proficiency in all four aspects: language, content, text structure, and technical specification.

Peer feedback is recognized as another valuable form of human feedback (Shang, 2022). Chen and Cui (2022) compared the effectiveness of two feedback modes—AWE (iWrite) versus peer feedback—regarding cohesion and coherence in a continuation writing task. The results proved that students given peer feedback had greater success in increasing the use of cohesive devices. The cohesive devices were automatically generated by Coh-Metrix 3.0. Several studies have explored the potential benefits of integrating AWE and peer feedback to enhance students' writing quality (Fu et al., 2024; Tan et al., 2023). For example, Tan et al. (2023) compared three different feedback modes—AWE, peer, and AWE+peer—on Chinese undergraduates' writing in terms of complexity, accuracy, and fluency (CAF) and found that AWE+peer feedback outperformed the others. Shang (2022) compared the effect of online peer feedback (OPF) versus AWE on the writing performance of learners in complexity, grammatical accuracy, and lexical density. The study showed that OPF is potentially more helpful than AWE in improving sentence writing, resulting in fewer grammatical errors and the production of more lexical terms. Although AWE+peer feedback can have an impact on students' writing performance, they still expected to receive teacher feedback (Fu et al., 2024) and hoped to get it along with feedback from their AWE and their peers (Cheng et al., 2021).

Supported by previous studies, researchers have begun to focus on multiple feedback, that is, the combination of teacher feedback, peer feedback, and machine feedback. Huang and Zhang (2014) explored the AWE+peer+teacher feedback mode on revising college students' English essays. The result showed that this multi-mode feedback can increase students' amendments, allowing them to make multiple revisions. Chen and Gong (2022) conducted two rounds of action research in the English writing class of freshmen majoring in English. Through in-depth observation of teaching and detailed analysis, it was found that the AWE+peer+teacher feedback model can promote students' in-depth learning. These studies didn't explore the effect of AWE+peer+teacher on writing quality. Therefore, there is a need to continue to study the effect of this multi-feedback mode.

Cohesion and coherence are crucial aspects of EFL writing, specifically in establishing discourse linkages (Plakans & Gebril, 2017; Crossley et al., 2016; Lee, 2002). These qualities are indispensable for effective writing (Witte & Faigley, 1981). Coherence refers to the semantic and pragmatic relations between text segments that are perceived by a reader. On the other hand, cohesion relates to linguistic representations of these relationships (Plakans & Gebril, 2017; Lee, 2002). Coherence has been described primarily from a functional standpoint as consistency in subject development (Todd et al., 2007) and examined holistically in writing assessment (Plakans & Gebril, 2017) as a reader's internal, process-oriented concept. Cohesion has been systematically classified as a text-based quality and more objectively studied by manual or automated analysis (Crossley et al., 2016).

The majority of research on cohesion in written discourse focuses on grammatical cohesion (the usage of reference phrases and conjunctions) and lexical cohesion (the link between lexical elements that are morphologically or semantically related). According to studies, these two types of cohesiveness are the most common forms of discourse linkage and have a positive relationship with writing quality (Liu & Braine, 2005; Reynolds, 2001). Coherence has been investigated and classified into two types: Sentential linkages of the text have been used to describe local coherence (McNamara & Kintsch, 1996, p. 252). The microstructure of a text is frequently referred to by its local coherence. In contrast, global coherence refers to the underlying relationships between a text's ideas (McNamara & Kintsch, 1996). Because global coherence refers to the macro-structure of a text and is related to the entire discourse of a document, a text's macrostructure "can be cued directly in the text via topic headers and topic sentences" (McNamara & Kintsch, 1996, p. 252).

Despite the importance of these elements, few studies have examined the effects of different feedback approaches on cohesion and coherence specifically. Therefore, this study aims to investigate AWE+human feedback's impact on the writing quality of Chinese EFL learners, focusing on holistic scores, cohesion, and coherence.

III. Methodology

A. Research Design

This study employed a quasi-experimental research design. Three intact classes were randomly assigned as one control group and two experimental groups. The control group received only teacher (TF) feedback, while the first experimental group was given AWE+Teacher (AT) feedback, and the second experimental group was given AWE+Peer+Teacher (APT) feedback. A quantitative research method was employed to examine the changes in students' writing quality regarding holistic scores, cohesion and coherence after the intervention of different feedback modes through pretest and posttest writing tasks.

B. Setting and Participants

This study was conducted in a Foreign Language Department at a local Chinese university. In this department, the sophomores are offered 1.5-credit writing courses (*Basic English Writing*) in their third term. Data for the current study were collected in the third semester when the students began to learn basic writing skills and to improve English writing progressively. The participants of this study were 90 sophomores from three intact classes of Business English Major. Three intact classes were allocated randomly to serve as experimental and control groups. Each class followed their assigned feedback mode throughout the study. The same instructor will teach three classes, which was necessary to prevent the instructor's experience and manner from functioning as an extraneous variable affecting the study's validity.

In order to ensure that the three classes were identical in terms of writing proficiency, all participants in the control group and the two experimental groups were administered a pretest writing task before the study started. The pretest scores of the three groups were compared through an ANOVA analysis. The results of the ANOVA test (see Table1) showed that there was no significant difference (F=.587 p>.05) among the pretest scores of the experimental groups and the control group.

TABLE 1
ANOVA TEST OF THE PRETEST FOR THREE FEEDBACK MODES

Pretest					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	64.286	2	32.143	.587	.558
Within Groups	4767.834	87	54.803		
Total	4832.120	89			

$\alpha = 0.05$

C. Data Collection

(a). Instrument

1. iWrite

The AWE tool used in the present study was iWrite 3.0. iWrite (https://iwrite.unipus.cn/) is an AWE system developed explicitly for Chinese English learners by the Foreign Language Teaching and Research Press and the National Research Center of Foreign Language Education in 2015. The system has undergone several versions, and the research used the latest version, iWrite 3.0. iWrite is a web-based writing tool designed to support high school and college students by offering a variety of writing tasks and comprehensive evaluations. It assesses students' writing from four critical perspectives: language (e.g., grammatical accuracy and vocabulary complexity), content (e.g., topic relevance and coherence), organization (e.g., use of cohesive devices), and mechanics (e.g., punctuation and spelling). The system generates a holistic score for each draft and assigns a 1 to 5-star rating for each evaluated aspect. To facilitate revisions, iWrite 3.0 also provides general feedback on these areas, with specific attention to coherence. Coherence is assessed through a percentage score that reflects both the relevance to the topic and the overall coherence of the text, offering students clear guidance for improving the logical flow and cohesion of their writing. This feature makes iWrite 3.0 a suitable tool for evaluating both surface-level and deeper structural aspects of student writing in this study (see Figure 1).

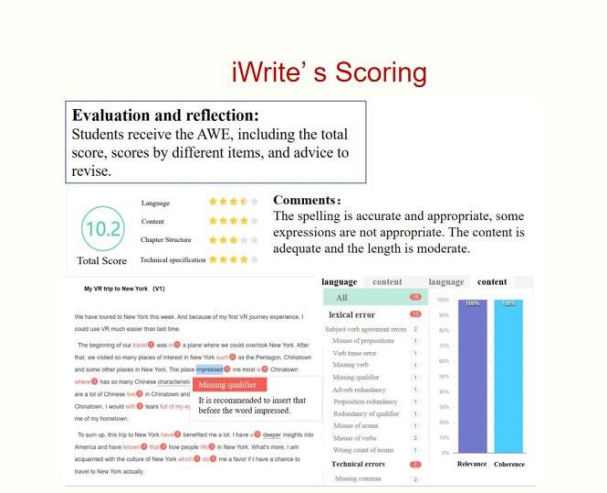

Figure 1. iWrite's Scoring

2. Coh-Metrix 3.0

Although iWrite provides feedback on cohesion and coherence, it only provides results about coherence through a percentage score; a more nuanced analysis requires additional tools. Therefore, cohesion and coherence were further explored using Coh-Metrix. Coh-Metrix (http://cohmetrix.memphis.edu/cohmetrixhome/) is a computational tool (see Figure 2) that produces 108 indices of a text's linguistic and discourse representations. These values can be used in many ways to investigate the writing quality from the vocabulary, syntactic, and coherence perspectives. To comprehensively analyze the writing quality regarding coherence and cohesion, the author will employ six indices (see Table 2) automatically generated from Coh-Metrix 3.0. Previous studies validated these indices (Crossley et al., 2016).

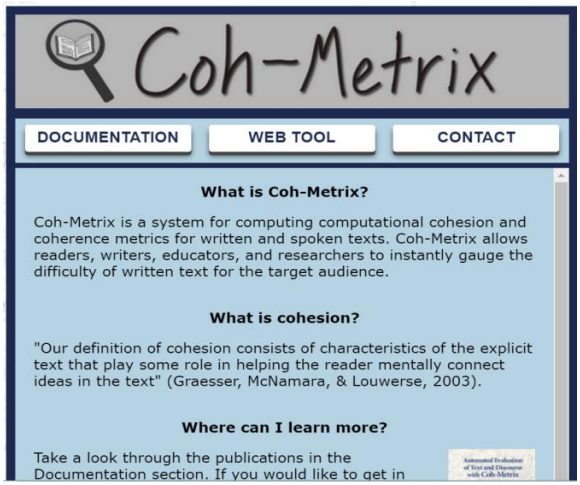

Figure 2. Coh-Metrix

This study selected six indices to measure cohesion and coherence (see Table 2). Cohesion indicators evaluate surface-level features like lexical items and connectives, reflecting connections between sentences and paragraphs (Crossley & McNamara, 2011). Two referential indices (CRFAOa and CRFCWOa) and two connective indices (CNCAll and CNCCaus) were selected. In contrast, coherence indicators assess deeper semantic consistency and the logical flow of ideas (McNamara et al., 2014). Here, two Latent Semantic Analysis indices (LSASSp and LSAPP1) were used. These indices were chosen because they measure cohesive devices at both local and global levels and have been validated in many studies (e.g., Crossley et al., 2016).

TABLE 2
SIX INDICES OF COHESION AND COHERENCE

Writing Quality	Indices	Description	Category
Cohesion	CRFAOa	Argument overlap, all sentences, binary, mean	Referential Cohesion
	CRFCWOa	Content word overlap, all sentences, proportional, mean	Referential Cohesion
	CNCAll,	All connectives incidence	Connectives
	CNCCaus	Causal connectives incidence	Connectives
Coherence	LSASSp	LSA overlap, all sentences in paragraph, mean	LSA
	LSAPP1	LSA overlap, adjacent paragraphs, mean	LSA

(b). Materials

Writing pretests and posttests were used to examine the effects of the treatments of the study. These tests were selected from the College English Test Band 6 (CET6) writing exercises, and they have the same difficulty level and similar topics. The pretest topic was "My View on the Use of PowerPoint (PPT)," while the posttest topic was "On a Harmonious Dormitory Life".

D. Data Analysis Methods

This study used quantitative data analysis with SPSS 25 and Coh-Metrix to address RQ1, examining the effects of three feedback modes on holistic scores, cohesion, and coherence. Paired-sample t-tests in SPSS 25 were used to compare pretest and posttest scores within each feedback group, assessing improvements in holistic scores. For cohesion and coherence, six indices of data were generated from the pretest and posttest texts of three groups using Coh-Metrix. These indices were then analyzed with paired-sample t-tests in SPSS 25 to measure changes in cohesion and coherence within each group, providing quantitative insights into the effects of each feedback mode. For RQ2, a one-way ANOVA was employed to explore potential significant differences among the three feedback modes regarding holistic writing scores, cohesion, and coherence. Posttest holistic scores and six Coh-Metrix indices representing cohesion and coherence were analyzed across the three groups. Post-hoc tests were then conducted to identify specific differences between the groups.

E. Research Procedure

This study lasted 15 weeks. Students had a writing lesson once a week. Each lesson had two periods, each lasting 45 minutes with a ten-minute break in between.

Week 1

Students were introduced to the course syllabus and objectives in the first period. The second period involved a 30-minute pretest assessing writing proficiency, which resembled CET 6 in difficulty and covered familiar topics. For the pretest, students first wrote their essays in class using pen and paper, then took photos of their handwritten compositions with their phones. They submitted the paper copies to the teacher immediately when the test was over. Once trained on iWrite, students later uploaded digital versions of their essays to iWrite based on the photos they had taken.

Week 2

All students, including the control group, were trained on how to use iWrite 3.0 and submit writing tasks. However, only experimental groups received automated feedback. In the second period, students learned to give peer feedback through iWrite's peer feedback system.

Weeks 3-14

These weeks were treatment weeks. During these treatment weeks, the teacher delivered her planned lectures to all classes during scheduled class hours. Then, after class, Writing Task 1 was released to students through iWrite, and different classes received different feedback modes. In other words, these three classes received the same teaching and writing tasks except for different feedback modes. Only feedback modes affected their final writing quality.

Each student in the three groups produced essays on four different topics in four "writing cycles." Given the relatively common use and importance of the argumentative style in English composition, the experimental writing tasks all focused on the argumentative genre, which included various topics close to daily life. A writing task would be released every three weeks. In other words, students had to finish one writing task in three weeks. So, every three weeks was a writing cycle. In each writing cycle, all participants wrote an argumentative essay of more than 200 words on the same controversial topic. Three groups received different feedback processes. The control group received only teacher feedback. Students submitted their work through the iWrite system without seeing any automated feedback. After the teacher had reviewed the work, students revised it based on the teacher's comments and resubmitted it for final grading. Experimental Group 1 combined AWE with teacher feedback. Students first submitted their work through the iWrite system, received automated feedback with scores and specific criteria ratings, and revised based on this feedback as many times as desired before resubmission. Teachers then reviewed the iWrite-revised work, provided feedback, and students completed a final revision based on the teacher's comments. Experimental Group 2 received iWrite, peer, and teacher feedback. The initial stage was identical to Experimental Group 1, where students received and incorporated automated feedback through multiple revisions. Next, students participated in peer review, anonymously exchanging comments based on AWE-provided criteria but without scoring. Finally, after revisions based on peer feedback, students received teacher feedback, completed a final revision, and resubmitted the work. There are four writing cycles altogether.

Week 15

In week 15, the posttest was implemented to obtain data on the effectiveness of the treatments throughout the program.

TABLE 3
RESEARCH PROCEDURE

Groups	Group1(n=30)	Group2(n=30)	Group 3(n=30)
Week 1	The course syllabus and objectives introduction Pre-test writing		
Week 2	Training of AWE (iWrite) for all three groups		
Weeks 3-14	Only teacher feedback	Treatment AWE+Teacher (AT) feedback	Treatment AWE+Peer+Teacher (APT) feedback
Week 15	Post-test writing		

IV. Results and Discussion

A. Results

(a). Results of RQ1

To assess the impact of the three feedback modes on students' holistic writing scores, cohesion, and coherence, pre-test and post-test holistic scores—including six indices of data generated from Coh-Metrix—were compared among the three groups. As noted in the setting and participant sections, a one-way ANOVA revealed no significant differences among the three groups ($F(2, 87) = .587$, $p = .558$). Following the intervention, paired sample t-tests were conducted to determine if each group's treatment affected writing scores, cohesion, and coherence indices. The results are summarized below in Table 4.

Table 4
Paired Sample T-Test of Holistic Scores and Six Indices Between Pre- and Post-Test for Three Groups

Item	Group	Pretest Mean	Posttest Mean	Mean Difference	t-value	df	P
Holistic Scores	TF	63.127	76.290	13.163	9.138	29	.000***
	AT	62.117	79.163	17.047	9.738	29	.000***
	APT	61.057	80.246	19.190	10.626	29	.000***
CRFAOa	TF	.375	.40	.016	.684	29	.499
	AT	.344	.412	.068	2.111	29	.044*
	APT	.369	.319	.029	1.035	29	.309
CRFCWOa	TF	.085	.097	.008	1.583	29	.124
	AT	.073	.091	.018	2.568	29	.016
	APT	.083	.084	.001	.048	29	.962
CNCAll	TF	69.549	68.702	-.847	.230	29	.820
	AT	71.163	76.122	4.959	1.102	29	.280
	APT	70.136	81.122	10.986	2.532	29	.017
CNCCaus	TF	69.706	70.612	.906	.226	29	.823
	AT	73.476	76.048	2.572	.597	29	.555
	APT	67.681	76.777	9.096	1.689	29	.102
LSASSp	TF	.138	.184	.045	3.658	29	.001*
	AT	.109	.219	.110	11.116	29	.000***
	APT	.123	.236	.112	7.709	29	.000***
LSAPP1	TF	.175	.297	.122	4.771	29	.000***
	AT	.142	.303	.161	6.202	29	.000***
	APT	.149	.307	.158	9.327	29	.000***

$\alpha = 0.05$

As shown in Table 4, significant improvements in students' holistic writing scores were observed across all three groups: TF ($t(29) = 9.138$, $p < .001$), AT ($t(29) = 9.738$, $p < .001$), and APT ($t(29) = 10.626$, $p < .001$). These findings suggest that each feedback mode effectively enhanced the overall writing quality of students. When comparing the mean scores for indices of coherence, significant improvements were observed in two indices: LSASSp (sentence-level coherence) and LSAPP1 (paragraph-level coherence). For LSASSp, the results were TF [$t(29) = 3.658$, $p = .001$], AT [$t(29) = 11.116$, $p < .001$], and APT [$t(29) = 7.709$, $p < .001$], indicating all groups contributed to better sentence-level coherence. Similarly, for LSAPP1, significant improvements were found in TF [$t(29) = 4.771$, $p < .001$], AT [$t(29) = 6.202$, $p < .001$], and APT [$t(29) = 9.327$, $p < .001$], suggesting enhanced paragraph-level coherence across all groups.

Regarding the cohesion indices, only two measures, CRFAOa (argument overlap) and CNCAll (connective overlap), showed significant improvement, and only in specific groups. For CRFAOa, the AT group [$t(29) = 2.111$, $p < .05$] demonstrated a significant difference between the pretest and posttest, indicating that AT feedback effectively improved students' argument overlap. For CNCAll, only the APT group [$t(29) = 2.532$, $p < .05$] showed significant improvement, indicating that combined feedback improved students' use of connective words.

Although other cohesion indices did not exhibit statistically significant improvements, the mean posttest scores were consistently higher than the pretest scores across all groups, except for teacher feedback on CNCAll, where no improvement was observed.

The findings indicate that while all three feedback modes significantly improved students' coherence at the sentence and paragraph levels, the effects on cohesion were more selective. Teacher feedback did not significantly impact cohesion, whereas AT feedback improved argument overlap and APT feedback enhanced connective word overlap. These results suggest that automated feedback, especially when combined with peer input, can be particularly effective in enhancing cohesion-related aspects of writing.

(b). Results of RQ2

After the treatment, another one-way ANOVA test was conducted to compare the post-test scores of the three groups to see whether one group showed a higher improvement in writing proficiency than others, both in total and in each of the indices of cohesion and coherence. The results are given in Table 5.

TABLE 5
ONE-WAY ANOVA TEST COMPARING HOLISTIC SCORES AND SIX INDICES OF THE POST-TEST OF THE THREE GROUPS

Item	Feedback Mode	Mean	Std. Deviation	F	P
Holistic Scores	TF	76.290	5.573	4.305	.016
	AT	79.163	5.545		
	APT	80.246	5.053		
CRFAOa	TF	.391	.126	.176	.839
	AT	.412	.155		
	APT	.399	.130		
CRFCWOa	TF	.097	.036	1.072	.347
	AT	.091	.036		
	APT	.084	.032		
CNCAll	TF	68.702	15.416	3.208	.045
	AT	76.122	20.47		
	APT	81.122	20.947		
CNCCaus	TF	70.612	16.714	1.101	.337
	AT	76.048	20.470		
	APT	76.777	18.582		
LSASSp	TF	.183	.063	5.717	.005*
	AT	.219	.054		
	APT	.236	.065		
LSAPP1	TF	.297	.130	.053	.949
	AT	.303	.105		
	APT	.307	.100		

α = 0.05

Based on Table 5, the results showed significant differences in the holistic writing scores, CNCAll, and LSASSp across the three feedback types. Specifically, there was a statistically significant difference in the holistic writing scores [$F(2,87) = 4.31$, $p = .016$], indicating that the feedback methods had varied impacts on overall writing quality. Similarly, a significant effect was found for CNCAll [$F(2,87) = 3.21$, $p = .045$] and for LSASSp [$F(2,87) = 5.72$, $p = .005$], suggesting that the feedback modes influenced connective cohesion and sentence coherence differently.

To further explore which groups were significantly different, a post-hoc test was conducted (see Table 6).

TABLE 6
POST HOC TESTS OF FEEDBACK TYPES ON WRITING SCORES AND COHESION-COHERENCE IN POST-TEST

Item	Between Groups	Mean Difference (MD) (APT-TF)	Std. Error	P
Holistic Scores	APT vs TF	3.957	1.394	.015*
CNCAll	APT vs TF	12.420	4.934	.036*
LSASSp	APT vs TF	.052	.016	.004*

α = 0.05

The post-hoc analysis revealed significant differences between the TF and APT groups across key writing measures. Specifically, learners in the APT group outperformed those in the TF group in terms of holistic writing scores, with a mean difference of 3.96 ($MD = 3.96$, $SE = 1.39$, $p = .015$). Additionally, a significant difference was found in CNCAll, where the APT group showed a mean difference of 12.42 ($MD = 12.42$, $SE = 4.93$, $p = .036$), indicating stronger cohesion than the TF group. Moreover, the APT group exhibited superior performance in sentence-level coherence (LSASSp), with a mean difference of 0.052 ($MD = 0.052$, $SE = 0.016$, $p = .004$), highlighting the more significant impact of APT feedback on coherence.

B. Discussion

(a). Holistic Scores

The results show that all three feedback modes improved students' overall writing scores, confirming the benefits of feedback for writing skills (Guo et al., 2024; Hyland, 2006). The APT group (AWE+Peer+Teacher feedback) had the highest scores, indicating that combining these feedback types is most effective. This aligns with research showing that AWE can supplement teacher feedback well (Link, 2022) and that peer feedback adds value (Chen & Cui, 2022; Shang, 2022).

The success of the APT model likely comes from the strengths of each feedback source. iWrite's automated feedback provides instant comments on grammar, cohesion, and syntax, encouraging students to practice more while easing their anxieties. Peer feedback supports critical thinking and self-reflection as students review and learn from each other's work. This layered approach maximizes the impact of feedback and promotes active learning.

(b). Coherence

All feedback modes significantly improved sentence-level (LSASSp) and paragraph-level coherence (LSAPP1), suggesting that each feedback method strengthens coherence at different levels. The APT group showed the most improvement at the sentence level, likely because peer feedback emphasizes sentence clarity, while AWE provides

specific feedback on sentence structure. This supports previous research indicating that peer feedback is particularly effective for sentence-level coherence (Shang, 2022).

The combination of AWE's specific sentence-level guidance, peer review's focus on critical thinking, and teacher feedback's overall structure helped the APT group achieve stronger coherence within sentences and across paragraphs.

(c). Cohesion

1. CNCALL (Lexical Cohesion)

The feedback modes had distinct effects on cohesion. With only teacher feedback, the TF group showed no significant improvement in CNCALL, suggesting that teacher feedback alone may not address lexical cohesion effectively. The AT group, with AWE support, showed moderate improvement, while the APT group had the most significant gains. This suggests that peer feedback on word choice and AWE's guidance on connectives optimizes lexical cohesion.

2. CRFAOa (Argument Cohesion)

The TF group improved the most in CRFAOa, likely because teachers could focus more on argument structure without covering grammar or other basics. In the AT and APT groups, AWE focused on lower-level feedback, so teachers' attention to argument structure was slightly reduced. This highlights that dedicated teacher feedback alone is valuable for improving argument cohesion, especially for high-level writing qualities.

V. Conclusion

A. Findings

This study highlights that different feedback modes have varying impacts on student writing quality. The APT group (AWE+Peer+Teacher feedback) showed the most substantial improvements across multiple dimensions, including holistic scores, coherence, and cohesion. Specifically, the APT group achieved significantly higher holistic scores than the other groups, indicating that integrating AWE, peer, and teacher feedback provides a more effective and comprehensive approach to enhancing overall writing quality.

The APT model's effectiveness can be attributed to its multi-layered feedback structure: AWE provides detailed, automated feedback on specific linguistic and structural features; peer feedback encourages active engagement with sentence-level coherence and lexical cohesion; and teacher feedback reinforces overall content coherence and argument structure. This synergistic approach addresses both global and local aspects of writing, helping students refine their skills in a balanced and focused manner.

B. Theoretical and Practical Implications

This study could enrich feedback theory, and its findings can contribute to the understanding of feedback in writing education by demonstrating that a multifaceted feedback approach (integrating AWE, peer, and teacher feedback) is more effective than single-source feedback. This supports and expands existing research on the roles of peer and automated feedback, suggesting that they can complement traditional teacher feedback to enhance overall writing quality, particularly in holistic scoring, coherence, and cohesion.

Practically, this study has clear implications for writing instruction; teachers can benefit from implementing a combined feedback approach to maximize improvement in their students' writing. Incorporating AWE and peer feedback in the writing classroom improves specific writing proficiency areas and reduces teacher workload, allowing teachers to focus on higher-level feedback areas. This approach could be precious in resource-limited educational settings, where it is important to maximize feedback efficiency and coverage.

C. Limitation

However, the study has certain limitations. First, it was conducted on a relatively small sample of EFL learners from a single local university in China over a single academic term. Extending the study across a longer period and involving a larger sample size could yield more robust results and further significant indicators. Second, while this research focused on holistic quality and coherence, it did not explore other dimensions of writing quality, such as accuracy, fluency, linguistic diversity, and syntactic complexity. Future research could address these aspects to better understand how feedback modes impact various facets of EFL learners' writing.

Acknowledgements

We would like to extend our gratitude to the Shanxi Provincial Department of Education, China, for their funding and support through the project "Shanxi Provincial Education Science 14th Five-Year Plan 2023 Annual Special Project on Educational Digitization and Educational Teaching Reform: Research on Teacher Written Corrective Feedback in English Writing Based on Big Data" (SZH-230017), which has made this research possible.

REFERENCES

[1] Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, *2023*(1), 4253331. https://doi.org/10.1155/2023/4253331

[2] Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, *36*(4), 584-607. https://doi.org/10.1080/09588221.2021.1936071

[3] Burstein, J., Chodorow, M., & Leacock, C. (2003, August). Criterion^{SM} Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. *IAAI*, 3-10. Retrieved October 31, 2024, from Google Scholar.

[4] Chen, G., & Gong, M.Y. (2022). Duo yuan fan kui mo shi cu jin shen du xue xi de xing dong yan jiu [An Action Research on Multiple Feedback Models for Deeper Learning]. *Wai yu jiao xue*, *43*(3), 60-66.

[5] Chen, H. H. J., Chiu, S. T. L., & Liao, P. (2009). Analyzing the grammar feedback of two automated writing evaluation systems: My Access and Criterion. *Ying yu jiao xue qi kan*, *33*(2), 1–43. https://doi.org/10.6330/ETL.2009.33.2.01

[6] Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*, *57*, 100915. https://doi.org/10.1016/j.jslw.2022.100915

[7] Cheng, X., Zhang, L. J., & Yan, Q. (2021). Exploring teacher written feedback in EFL writing classrooms: Beliefs and practices in interaction. *Language Teaching Research*. https://doi.org/10.1177/13621688211057665

[8] Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology*, *21*(3),104 – 130. https://orcid.org/0000-0002-5586-1495

[9] Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*, 1-16. https://doi.org/10.1016/j.jslw.2016.01.003

[10] Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, *5*(1). Retrieved October30, 2024, from http://www.jtla.org

[11] Ferris, D. R., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of second language writing*, *6*(2), 155-182. https://doi.org/10.1016/S1060-3743(97)90032-1

[12] Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, *37*(1-2), 179-221. https://doi.org/10.1080/09588221.2022.2033787.

[13] Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, *63*, 100962. https://doi.org/10.1016/j.iheduc.2024.100962.

[14] Han, T., & Sari, E. (2024). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*, *37*(4), 961-985. https://doi.org/10.1080/09588221.2022.2067179

[15] Hassanzadeh, M., & Fotoohnejad, S. (2021). Implementing an automated feedback program for a Foreign Language writing course: A learner-centric study: Implementing an AWE tool in a L2 class. *Journal of Computer Assisted Learning*, *37*(5), 1494-1507. https://orcid.org/0000-0003-1510-1149

[16] Huang, J., & Zhang, W.X. (2014). Duo yuan fan kui dui da xue sheng ying yu zuo wen xiu gai de ying xiang yan jiu [The Impact of the Integrated Feedback on Students' Writing Revision]. *Zhong guo wai yu*, *11*(1), 51-56.

[17] Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, *39*(2), 83-101. https://doi.org/10.1017/S0261444806003399

[18] Keh, C. L. (1990). Feedback in the Writing Process: A Model and Methods for Implementation. *ELT Journal*, *4*(4), 294-304. https://doi.org/10.1093/elt/44.4.294

[19] Lee, I. (2002). Teaching coherence to ESL students: A classroom inquiry. *Journal of second language writing*, *11*(2), 135-159. https://doi.org/10.1016/S1060-3743 (02)00065-6

[20] Liao, H. C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, *62*, 77-92. https://doi.org/10.1016/j.system.2016.02.007

[21] Liao, L.Y. (2022). *Automated Evaluation or Human Evaluation? Optimal Means of Evaluation in EFL Writing Assessment*. [Unpublished master's thesis]. University of Macau.

[22] Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

[23] Liu, M., & Shao, C. (2024). The Effects of Integrated Feedback based on AWE on English Writing of Chinese EFL Learners. *Scalable Computing: Practice and Experience*, *25*(2), 987-996. https://doi.org/10.12694/scpe.v25i2.2617

[24] Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623-636. https://doi.org/10.1016/j.system.2005.02.002

[25] Liu, S. L. (2022). A Research on the Blended Evaluation Mode in College English Writing Course. *Journal of Language Teaching and Research*, *13*(4), 763-771. https://doi.org/10.17507/jltr.1304.09

[26] McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse processes*, *22*(3), 247-288. https://doi.org/10.1080/01638539609544975

[27] Plakans, L., and Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, *31*, 98-112. https://doi.org/10.1016/j.asw.2016.08.005.

[28] Ramaprasad, A. (1983). On the definition of feedback. *Behavioral science*, *28*(1), 4-13. https://doi.org/10.1002/bs.3830280103

[29] Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653-674. https://doi.org/10.1080/09588221.2018.1428994

[30] Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, *52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816

[31] Reynolds, D. W. (2001). Language in the balance: Lexical repetition as a function of topic, cultural background, and writing development. *Language Learning*, *51*(3), 437-476. https://doi.org/10.1111/0023-8333.00161

[32] Shang, H. F. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, *30*(1), 4-16. https://doi.org/10.1080/10494820.2019.1629601

[33] Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51-65. https://doi.org/10.1016/j.asw.2013.11.007
[34] Tan, S., Cho, Y. W., & Xu, W. (2023). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments*, *31*(10), 7276-7286. https://doi.org/10.1080/10494820.2022.2066137
[35] Todd, R. W., Khongput, S., & Darasawang, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing writing*, *12*(1), 10-25. https://doi.org/10.1016/j.asw.2007.02.002
[36] Wang, Y., Luo, X., Liu, C. C., Tu, Y. F., & Wang, N. (2022). An integrated automatic writing evaluation and SVVR approach to improve students' EFL writing performance. *Sustainability*, *14*(18), 11586. https://doi.org/10.3390/su141811586.
[37] Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22-36. https://doi.org/10.1080/15544800701771580
[38] Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, *16*(3), 194-209. https://doi.org/10.1016/j.jslw.2007.07.004
[39] Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87-125. https:// doi.org/10.1177/0735633119830764
[40] Witte, S.P, & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition & Communication*, *32*(2), 189-204. https://doi.org/10.58680/ccc198115912
[41] Xu, H. Y. (2020). *Ji yu iWrie2.0 xie zuo ping yue xi tong de liang zhong fan kui fang shi dui da xue sheng ying yu xie zuo ying xiang de shi zheng yan jiu* [An Empirical Study on the Influence of Two Feedback Modes on College English Writing ——Based on iWrite2.0 Writing Evaluation System], [Unpublished master's thesis], Jiangxi shi fan da xue.
[42] Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2022). Examining the relative effects of task complexity and cognitive demands on students' writing in a second language. *Studies in Second Language Acquisition*, *44*(2), 483-506. https://doi.org/10.1017/S0272263121000310
[43] Yu, S. (2021). Feedback-giving practice for L2 writing teachers: Friend or foe? *Journal of Second Language Writing*, *52*, 100798. https://doi.org/10.1016/j.jslw.2021.100798
[44] Zhai, N., & Ma, X. (2022). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, *35*(9), 2817-2842. https://doi.org/ 10.1080/09588221.2021.1897019

Yan Zhang is a Ph.D. candidate at Universiti Putra Malaysia and a lecturer in the Department of Foreign Languages at Yuncheng University, China. Her research interests include applied linguistics, ESL writing, and second language acquisition.

Ilyana Binti Jalaluddin earned her Post Doctoral (TESL)from the Institute of Education, University of Reading, UK, in 2016. She is currently a senior lecturer in the Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia. Her research areas include English language studies, literacy studies, ESL writing, TESL, and second language acquisition. She has published articles in indexed research journals in the area of her studies.

Roslina Mamat is an Associate Professor at the Department of Foreign Languages, Faculty of Modern Language and Communication, Universiti Putra Malaysia. Her field of specialization is Japanese language, discourse analysis and popular culture. She has published articles in indexed research journals in the area of her studies. She was Head of the Department of Foreign Languages Faculty of Modern Languages and Communication, Universiti Putra Malaysia, from 2011-2016. She was the Chair of the Malaysia International Conference on Foreign Languages (MICFL) 2024.

Tingting Zhao is an Associate Professor at International Education School of Yuncheng University, China. Her research interests include applied linguistics, ESL writing, and second language learning. She has published articles in indexed research journals in the area of her studies.