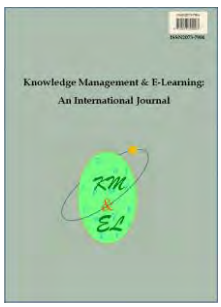
Knowledge Management & E-Learning, Vol.17, No.2. Jun 2025

Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance

Ali Al Ghaithi
Sohar University, Oman
Behnam Behforouz
University of Technology and Applied Sciences, Shinas, Oman



Knowledge Management & E-Learning: An International Journal (KM&EL)
ISSN 2073-7904

Recommended citation:
Al Ghaithi, A., & Behforouz, B. (2025). Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance. *Knowledge Management & E-Learning, 17*(2), 206–224. https://doi.org/10.34105/j.kmel.2025.17.009

Knowledge Management & E-Learning, 17(2), 206–224

Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance

Ali Al Ghaithi*

Foundation Department
Sohar University, Oman
E-mail: AGhaithi@su.edu.om

Behnam Behforouz

Preparatory Studies Center
University of Technology and Applied Sciences, Shinas, Oman
E-mail: Behnam.Behforouz@utas.edu.om

*Corresponding author

Abstract: This study aimed to measure a newly developed automated written evaluation feedback instrument, EditGPT, on the writing performance of 30 Omani EFL learners and their perceptions of this tool's functionality in learning English as a foreign language (EFL). The learners were divided into three groups: control (receiving general feedback on writing), experimental group A (receiving writing feedback from researchers), and experimental group (receiving writing feedback from EditGPT). To collect the required data, a pretest and posttest of writing based on narration and compare-and-contrast were conducted to measure students' progress. A questionnaire initially developed by Huang and Renandya (2020) was adopted, translated, and used thoroughly to elicit students' perceptions. The results showed that experimental group A performed much better than the other two groups; experimental group B, which received feedback only from EditGPT, showed more progress than the control group. The results of this study demonstrate the usability of EditGPT and the role of the teacher in such learning contexts.

Keywords: Automated writing evaluation; Writing performance; Perceptions; Functionality; EFL learners

Biographical notes: Ali Al Ghaithi is an English Lecturer at the Foundation Department of Sohar University in Oman. Currently, he is a Ph.D. candidate focusing on Applied Linguistics. Ali got his master's degree from the University of Putra Malaysia. Ali is interested in research studies that mainly implement Artificial Intelligence in teaching and learning processes.

Behnam Behforouz (Ph.D.) is an English Lecturer in the Preparatory Studies Center at the University of Technology and Applied Sciences, Shinas, Oman. His main areas of interest are TESOL, Applied Linguistics, Language Education, and Educational Technologies.

1. Introduction

In educational curricula, writing is a fundamental element that fosters growth in the literacy of students who learn languages and encourages academic and personal progress, such as preparing proposals and publishing in scientific journals and conferences (Ahmed et al., 2021). Writing skills encourage students to consider, organize, and generate ideas through written work to communicate with the author and audience. Therefore, to prevent confusion among readers, a text author must create accurate language, effectively organize written work, and identify and eliminate errors (Bitchener & Ferris, 2012; Khadawardi, 2020). Some scholars (Banaruee, 2016; Richards & Renandya, 2002) have unequivocally declared that developing writing abilities constitutes one of the most arduous aspects of language instruction and acquisition. Additionally, writing duties require students to possess exceptionally complex skills if their level of proficiency is low. It is customary to offer feedback on students' written work as an instructional method to enhance their writing ability. Measuring feedback efficiency in writing ability is an emerging area of study emphasized by numerous scholars (Ferris, 2004, 2007; Karim & Nassaji, 2018; Lee, 2009). According to Ferris (2004), feedback functions importantly in narrowing students' current knowledge, revealing potential areas for growth and the target language they are required to attain. Teachers must invest substantial energy and time to provide students with constructive criticism of written work (Zhang, 2017). In addition to time constraints, excessive duties, and large class sizes, contextual factors exacerbate the burden of criticism. Subsequently, automated writing evaluation (AWE) instruments were implemented in writing courses to supplement instructor feedback (Wilson & Czik, 2016). Writing in L2 studies advocates incorporating automated tools into lessons to enhance instructor feedback effectiveness. This integration would enable instructors to consider significant issues, such as content and organization, rather than lower-order problems, such as mechanics and grammatical structures (Koltovskaia, 2020; Li et al., 2015; Ranalli, 2018).

By providing automated corrective feedback, technological tools assist foreign language students in producing more precise writing, that is, fewer syntax, mechanics, organization, and vocabulary errors. AWE programs have the potential to offer not only assistance to L2 learners but also a time-saving benefit to language teachers. Rather than providing feedback on writing techniques at a lower level, lecturers can allocate additional time to the various facets of L2 writing (Grimes & Warschauer, 2010). This technology provides writers with time-word-choice alternatives by analyzing the phrase's initial letter input and clause context. Novice L2 learners can enhance their output while reducing grammatical and lexical errors. Furthermore, progress in machine learning and natural language processing has led to the development of intelligent writing advisors (Frankenberg-Garcia, 2020). These assistants provide simultaneous critiques of more intricate facets of writing in contrast to text editors. In addition to generating an automated score, contemporary AWE tools can offer feedback on content and language, owing to progress in statistical methodologies and language processing technologies. Hence, conducting studies on the impact of automatic corrective feedback delivered by AWE is of considerable importance (Wang et al., 2022). Although incorporating Web 2.0 tools and AWE programs in L2 studies has been extensively studied, there seems to be insufficient research examining the utility of predictive texts to assist L2 writing. In addition, study data on text prediction and accompanying intelligent writing aids are scarce (Hammoda, 2024). Hence, examining the potential applications of automated written corrective feedback as a supplementary instrument within and beyond the classroom environment for students' writing ability is critical (Frankenberg-Garcia, 2020). Therefore, this study aims

to investigate the role of EditGPT, an automated correcting platform, on learners' writing performance. This study focuses on learning English as a foreign language (EFL) among Omani students. To this end, the following questions were analyzed:

RQ1: Does the use of EditGPT assist Omani EFL learners in improving their writing performance?

RQ2: What are the perceptions of Omani EFL learners about using Edit GPT to improve writing performance?

2. Literature review

2.1. Automated writing evaluation

It is believed that AWE instruments or infrastructure, which have been in use since the 1960s (Chen & Cheng, 2008), can assist in addressing these obstacles and, as a result, reduce the workload of educators by providing students with continuous, immediate feedback beyond the classroom settings (Woodworth & Barkaoui, 2020). AWE instruments are endorsed for their capacity to deliver detailed and thorough feedback encompassing various writing features such as content, syntax, grammar, and word selection. They also facilitate scoring functionalities that aid instructors in assessing and grading students' essays, emphasizing the linguistic components. These tools provide students access to various composing and writing resources, such as thesauri (Chen & Cheng, 2008). Additionally, they include features that enable instructors to generate direct writing assignments (Ranalli et al., 2016). Woodworth and Barkaoui (2020) identified Criterion, Project Essay Grade, My Access, and Piagi as the most commonly utilized AWE tools. These applications facilitate instant feedback on various writing aspects, permit multiple drafts, and enable instructors and students to track the accomplished tasks.

Burstein et al. (2003) suggested that automatic correcting feedback mechanisms could simplify the recurrent writing procedure by relieving instructors in delivering formative personalized feedback to learners. A computerized system can provide learners with a diagnostic evaluation that addresses crucial writing elements, including sentence structure, word choice, and organization. By applying this knowledge, students are empowered to revise their papers individually, thereby participating autonomously in the writing, feedback, and revision processes (Wang et al., 2022).

By leveraging statistical methods, machine learning, and artificial intelligence, AWE instruments assess written material in a reduced timeframe compared with traditional evaluations (Grimes & Warschauer, 2010). AWE instruments minimize the workforce, making them indispensable for formative assessments (Bennett, 2011). Through the provision of repetitive formative feedback, these tools assist students in enhancing their writing ability. Contemporary AWE programs involving Criterion, MyAccess, and Writing Roadmap 2.0 offer evaluations and numerical results that assess the total writing score. The score is derived from a piece of student work and assists students in improving writing in multiple domains, including grammar, word selection, sentence construction, organization, and concept development (Gao et al., 2020; Li et al., 2017; McNamara et al., 2015; Rich, 2012).

When discussing automatic textual assessments, machine judgment quality frequently hinges on the degree of concurrence with human judgments. Numerous studies

have yielded favorable outcomes concerning validity and reliability (Shermis, 2014; Rupp et al., 2019; Latifi & Gierl, 2021). In every text evaluation facet, human assessors fail to perform better than technological advances. Coding errors are more prevalent among specialists in text segmentation and analysis. However, regarding conceptual association recognition, human assessors have demonstrated superiority over technology (Burkhart et al., 2020). Furthermore, opinion distortion can impact human and computer evaluations by excessively including particular textual characteristics (Perelman, 2014; Fleckenstein et al., 2020a, 2020b).

In contrast to feedback given by peers, automated feedback offers several benefits. Automated feedback eliminates the impediment that hinders the human ability to deliver valuable feedback. AWE can identify writing mistakes and evaluate dozens of textual indicators within seconds. AWE systems produce one of the following outcomes according to this assessment: (a) a singular score that symbolizes the quality of a written assignment (similar to holistic marking), (b) distinct scores that correspond to particular writing traits (similar to analytic scoring), or (c) trait scores along with a total quality metric that is calculated by averaging or totaling the individual trait scores. Learners receive automated feedback that addresses their writing errors and offers suggestions to enhance their writing. Unlike human raters, AWE is an entirely uniform scoring system. AWE constantly identifies patterns of mistakes in student-written work, and a similar grade is assigned to the same manuscript, regardless of the number of repetitions submitted for assessment. In conclusion, automated feedback has the potential to expedite the feedback-practice cycle, which is critical for the development of writing abilities because of its ability to deliver immediate feedback (Kellogg & Whiteford, 2009).

Few scholarly investigations have explored the effect of automated feedback on writing quality, centered on two automated feedback systems: the Summary Street system by Pearson and ETS's E-rater® and Criterion® (Cheung, 2015). In their study, Kellogg et al. (2010) divided undergraduates into three groups based on their interaction with automated feedback generated by the E-rater and Criterion: those who received no input, occasional input, and constant input. Following automated feedback pertaining to grammar, usage, mechanics, style, organization, and growth, the students authored and amended the compositions. Furthermore, each student was assigned an e-rater aggregate quality score ranging from 1 to 6. The findings revealed that undergraduates with ongoing automated feedback significantly reduced grammar, use, mechanics, and style errors in their final manuscripts compared with those who obtained irregular input. Nonetheless, no significant variations in the aggregate quality score from the e-rater were observed among the feedback circumstances of the various versions.

Shermis et al. (2008) investigated the impacts of E-rater and Criterion automated feedback. Employing hierarchical linear modeling, this study assessed whether consistent instruction and exposure to automatic input were correlated with improvement in the standards of the final versions of seven subsequent papers composed of students in grades 6-8 and 10. Development was assessed through evaluations of various metrics, such as the general quality rating assigned by e-raters, duration of the texts, total number of distinguished words, and defects in grammar, usage, mechanics, and style. Students exhibited growth in all dependent metrics throughout the seven final revisions, with eighth graders demonstrating the highest level of achievement compared to their peers. While the results indicated an increase in the overall rating of e-rater quality, this may result in obscuring practice and general growth effects. Given the extended duration of the study (11 months) and the absence of control for practice effects or in-class instructions, it is

plausible that the progress made from the initial draft to the final draft was minimal. Furthermore, the increase in e-rater overall quality across essays could have been accounted for by performance and class instruction rather than automated feedback.

Research examining the efficacy of AWE has four advantages. Automated feedback has the potential to boost learning, foster student independence and inspiration to write, alleviate instructors' workload, and elevate learners' caliber of written work (Benali, 2021). AWE systems have the capability to offer individuals formative feedback along with summative feedback. This is achieved by providing various composing prospects, which can enhance students' recognition of mistakes and areas for improvement, ultimately leading to improved writing results. Indeed, these systems serve as writing aids by incorporating various online writing resources (e.g., thesaurus and word libraries) and editing functionalities (spelling and grammar checkers). Furthermore, one could contend that these systems foster autonomy among the learners by allowing student writers to engage in self-regulated learning environments in which they can practice, compose, assess, and modify a substantial quantity of essays (Chen & Cheng, 2008). Moreover, empirical studies have demonstrated that the promptness of automated feedback can enhance the caliber of writing students (Stevenson & Phakiti, 2014). In conjunction with the feedback and grades produced by these computers, prompt replies can inspire and motivate students to improve their writing skills (Cheng, 2017). AWE applications are also considered beneficial for instructors, as they enable them to devote more time to instructing students on various facets of writing rather than primarily grading essays and providing feedback (Stevenson & Phakiti, 2014).

Li et al. (2014) investigated the impacts of implementing the AWE system on written correctness in two academic ESL writing courses at a midwestern university in the United States over a 15-week semester. Qualitative data were collected by conducting interviews with ESL students and instructors, while quantitative data were gathered by analyzing AWE error reports and the number of submissions for each paper, as recorded by the system. Seventy ESL students and four instructors were involved in the study. The students were allowed to utilize AWE beyond the classroom and were obligated to use the system to revise and submit the final draft. Criterion App, an AWE system, highlights errors, offers explanations or suggestions for corrections, and delivers corrective feedback (on grammar, usage, mechanics, and style). The findings from the data analysis indicated that teachers and students were optimistic about the effectiveness of AWE instruments in enhancing essay standards. However, neither the teachers nor the students were pleased with the content or structure of the feedback provided by the system. Additionally, the results showed that implementing the Criterion App resulted in a more significant number of revisions, while automated feedback assisted students in enhancing the precision of their writing as they progressed from initial to final manuscripts.

Wang et al. (2020) completed a research investigation to determine whether comments received from AWE systems affected the level of students' utilization of textual proof as they progressed from the first version to the amended draft. This research also investigated how learners interpreted and applied feedback messages. The research was conducted in Louisiana, involving 143 students in the 3rd through 8th grades and seven educators from seven distinct institutions. Students were expected to demonstrate their comprehension of the texts through a multi-paragraph response incorporating illustrative details. Furthermore, the literary analysis responses of students were evaluated for their ability to employ well-founded logic by pertinent text-based evidence in the development of the topic. Qualitative data were collected by investigating the written tasks in their initial

and revised manuscripts. In addition, students were required to complete a survey regarding their evaluation of the efficacy of the eRevise AWE system when they submitted their revised drafts. eRevise was developed to evaluate students' writing by providing feedback on response-to-text assessment, rating responses, and assessing their capacity to formulate arguments about texts. Students were instructed to provide arguments supporting their responses to the instructions. eRevise employed a three-tiered feedback system to evaluate students' capacity to reason about texts in their written work: fullness (constraints on students to supplement proof and furnish additional information for each component of evidence), explanation (guides students to elucidate evidence), and connection (directs students to establish a correlation between their evidence and the overarching argument). The findings from the data analysis indicated that students maintained favorable attitudes toward the utilization of eRevise. The results also indicated that most students refined their papers, as evidenced by the increased incorporation of textual evidence from the initial to the revised manuscript. This improvement was consistent with the received feedback from eRevise. However, 18% of students demonstrated a significant improvement in essay quality.

2.2. Students' perceptions toward AWE

Some studies have discussed students' perceptions of automated written corrective feedback instruments (Fan, 2023; Thi & Nikolov, 2022; Fahmi & Cahyono, 2021; ONeill & Russell, 2019). In their recent study, Fan (2023) measured the effect of Grammarly's automated written corrections on the writing proficiency of students learning the English language. The participants of this study consisted of sixty-seven students enrolled in two university courses. The experimental group received the instructor and Grammarly feedback, whereas the control group received only teacher feedback. Written pretests and posttests, observations, interviews, and surveys were used to collect the data. As part of this method, Grammarly provided feedback on two manuscripts for the experimental group, whereas only instructor feedback was provided for the control group. Students viewed Grammarly, an automated written corrective feedback instrument for grammar assistance, favourably, according to a subsequent survey.

Thi and Nikolov (2022) investigated the integration of Grammarly's automated feedback with the instructor to enhance writing instruction. 27 intermediate-level EFL students represented a university in Myanmar. A pretest and a posttest were administered to the six writing assignments gathered over 13 weeks. Instructional feedback was provided to the students for each task using Grammarly, Track Changes, or both. The comments were assessed and classified. The revisions made by students were also examined. Teaching and learning outcomes were evaluated using an assessment instrument. Students' opinions were obtained through a survey. The results revealed a substantial enhancement in scores, and students generally held positive perceptions of the feedback provided by both the teacher and Grammarly. This implies that Grammarly had the potential to alleviate instructors' workload by offering complementary feedback.

In English writing courses for particular purposes, Fahmi and Cahyono (2021) investigated EFL students' perspectives regarding written corrective feedback by combining instructor feedback with the automatic writing evaluator software Grammarly. The research included twenty-six undergraduates enrolled in an ESP writing course. Teachers provided feedback to students who used Grammarly to compose numerous cause-and-effect passages on various subjects as a course component. To gauge the students'

opinions, the researchers distributed a 16-item survey to gauge their contentment with Grammarly's functionality, feedback provision, and the efficacy of integrating it with instructor remarks. According to the findings, students considered the combination of instructor feedback and Grammarly beneficial in enhancing their writing abilities.

In contrast to the conventional grammar feedback provided by an academic learning consultant, ONeill and Russell (2019) examined how students perceived using Grammarly, an automated feedback program. This study aimed to determine whether perceptions varied according to cohort, language proficiency, or mode of delivery when comparing the two feedback approaches. The participants comprised ninety-six individuals; forty-four received conventional feedback, and fifty-four received Grammarly feedback. Regarding the feedback experience, students responded to a questionnaire containing both a Likert scale and open-ended items. Generally, students who received feedback from Grammarly and guidance from their advisors expressed more favorable opinions and higher satisfaction levels than those who received only conventional feedback.

Automated written corrections and instructor corrective criticism in English as a Foreign Language (EFL) assessments were investigated by Ghufron (2019) from the learners' perspective. In addition, the merits and disadvantages of these techniques were assessed. A total of 120 students from an Indonesian university participated in this study. Twenty students participated in the interviews, and their responses to the student surveys were gathered. In addition to the open- and closed-ended questionnaires, the research was conducted using both methods. The results indicated that the students held favorable opinions regarding the two feedback forms. Teacher feedback enhanced the organization and content of grammar, spelling, and words more than Grammarly did. Grammarly was inferior in terms of organization and content, whereas teacher feedback was more effective in terms of mechanics.

3. Method

3.1. Participants

Sixty Omani students were selected using random sampling. They were studying introductory courses in the General Foundation Program (GFP) at a higher education institution in Oman. The ages of the students varied between 18 and 19 years, and all of them spoke Arabic as their native language. The study consisted of three groups of 20 students each. The first group was a control group; the other two were experimental groups A and B. Students enrolled in pre-intermediate levels during the second semester of 2024-2025 either based on the university's placement test or their assessments of the previous semester.

3.2. Instruments

To collect the essential data to answer the research questions, the following instruments were used.

3.2.1. Writing tasks

Pretests and posttests of writing on familiar topics were conducted for the students. The ecological validity of the assignments was established through retrieval from the course of this study. Because of its prevalence in university courses, the genre encompasses narrative and compare-and-contrast essays. The educational program at the institution utilized the second edition of English in the 21st Century. The book was used for instruction during the entirety of the second academic term of 2024, encompassing a range of skills. This specific literary work functioned as a fundamental textbook for pre-intermediate level undergraduates. The writing assignments incorporated four guiding cues that mirrored the structure illustrated in Fig. 1. These instructions were similarly created to mitigate potential linguistic variations.

Write an essay in about 150 words. Your writing is going to be evaluated according to the:

- The degree to which you have achieved the task
- The organization and coherence of the paragraph
- The range and accuracy of the grammatical structures
- The range and accuracy of the vocabulary

Your essay should have an introduction, two supporting paragraphs, and a concluding paragraph. Compare and contrast between e-books and paper books.

Fig. 1. Sample writing task instructions

A researcher-made assessment rubric was developed and piloted with 10 Omani EFL students to evaluate their writing performance progress. The reliability of the rubrics was 0.71, which is close to 1 and shows a relatively reliable source of assessment.

3.2.2. Perception questionnaire

To collect information regarding students' perceptions of using EditGPT, the first part of the questionnaire was adapted from Huang and Renandya (2020) and included 10 questions using a 5-point Likert Scale. On this scale, 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree. The main reason for adopting this questionnaire was that Huang and Renandya (2020) administered it to EFL students with positions similar to those of the study population. The second reason was that this questionnaire could be considered comprehensive based on the existing literature on automated writing corrective feedback to elicit EFL students' perceptions regarding the use of computerized tools in receiving feedback on their tasks. The questionnaire was divided into three sections. The first part of the questionnaire included three items on perceived comprehensibility. The second part, which included another three items, was related to students' perceptions of the advantages of receiving feedback for revision. The last four items in the last part were related to students' perceptions of the benefits of receiving feedback on their writing performance. Based on Fan's (2023) study, the reliability of the first part was calculated as 0.75, the second part as 0.81, and the last part as 0.72, which showed the questionnaire's high internal consistency and reliability.

To implement the questionnaire among the Omani EFL learners, the statements above were translated into Arabic by native speakers of Arabic who hold a Ph.D. in Applied Linguistics within the contextual teaching experience in Oman. After the translation, a pilot test was administered to ten EFL pre-intermediate learners to assess internal consistency. The final Arabic iteration of the survey demonstrated sufficient reliability, as evidenced by Cronbach's alpha coefficients of 0.80, 0.81, and 0.83, respectively.

3.2.3. EditGPT tool

Anthropic Company designed its artificial intelligence tool EditGPT to be beneficial, trustworthy, and ad-free. Constitutional AI is a method used to guarantee that the model functions securely and in the users' best interests. EditGPT can be distinguished by its ability to modify documents and provide enhanced tips. EditGPT evaluates the written materials and contexts of the user's submitted Word, PDF, and text files and offers critical feedback. The software can identify and correct orthographic and grammatical mistakes, propose alternative terminology and phrasing, highlight inconsistencies or repetition, and suggest methods to enhance the overall coherence and organization of the text. Highlighting EditGPT's recommendations within the written content facilitates evaluation and acceptance or rejection of modifications by the user. This enables an interactive revision experience in which an AI assistant assists in refining text creation and writing. Fig. 2. shows an example of feedback students received from the EditGPT application.

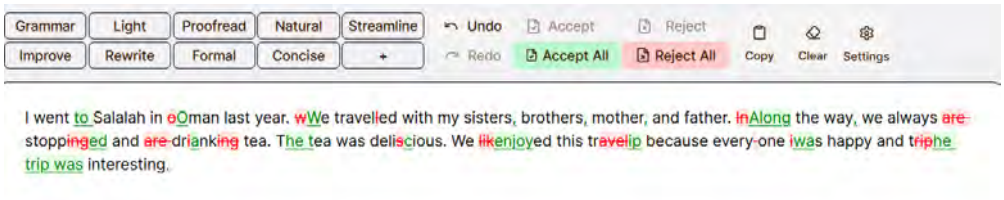


Fig. 2. Sample feedback from the EditGPT tool

3.3. Procedures

This study was conducted during the second semester of the 2024-2025 academic year in a higher education institution in Oman. Since these students were busy studying in the foundation program, following ethical considerations and the university's policies, permission to use the participants in extracurricular activities was received. Data were collected over four weeks. In the first week, the researchers conducted a writing pretest to ensure the homogeneity of students in all three groups on writing proficiency; after that, individual accounts were signed for one of the experimental groups to use EditGPT. Explanations regarding the format and style of narrative essays were given to the students of all groups, including some ready-made examples and explanations. In the second week, the students were asked to write an essay on a unit topic using narrative writing strategies and deliver it to the researchers. They were given three days to complete the task. Within the week, the researchers examined the writing tasks of the control group and underlined

their mistakes using abbreviations to show the issue areas without further explanation. It is worth mentioning that the topics of the pretests, in-class writing assignments, and posttests of all the groups were the same, and students could not select their topics freely. In experimental group A, the students received detailed feedback from the researchers on every utterance and sentence in their writing tasks in a one-to-one, face-to-face situation. They received feedback on task achievement, organization, lexicon, grammar, punctuation, spelling, and mechanics. Experimental group B was informed to use EditGPT to copy and paste the text or upload their files on the application. They were not supposed to receive feedback from the researchers. In week 3, the researchers followed the same process with comparison and contrast-based writing. Students in the experimental group exposed to EditGPT could use the features of this application. In the last week of the treatment, students were given a posttest of writing based on comparison and contrast topics. In total, the students in all groups completed two tasks during the treatment period. While the control group received general feedback twice from the researchers, the experimental groups received detailed feedback throughout the experiment.

3.4. Ethical considerations

The significance of ethical considerations in research is heightened because conducting research violating legal regulations would be challenging without proper authorization. Approval was obtained after establishing suitable communication with the organization in question. Before the data collection phase, the participants were given reading instructions during the introductory stages of the study. After receiving explicit confirmation that any data they had contributed would be kept confidential, every participant agreed to participate in the study.

4. Data analysis and results

4.1. Writing performance

To collect the required data to answer the research questions, SPSS version 27.0 was used. The analysis process was divided into three stages. Since the study aimed to measure the effect of EditGPT on the writing performance of EFL students, in the first stage, the pretests and posttests of learners were compared. As most applications changed the grammar and sentence structures of the written tasks, a grammar-focused analysis was conducted on the students' pretests and posttests. Finally, in stage three, the perceptions of students in experimental group B toward using AWE were thoroughly elicited.

The first research question aimed to determine the effect of EditGPT on Omani EFL learners' writing performance. To select an appropriate test for statistical analysis, it was necessary to measure the normal distribution of the data. Table 1 shows the Shapiro-Wilk test of normality results among the three groups in the pretest and posttest.

As can be observed in Table 1, all data among the groups are distributed normally because the *p*-value is greater than 0.5. Therefore, an ANOVA test was conducted to obtain more comprehensive information about the groups. Table 2 shows the results of ANOVA for the three groups in pretest and posttest.

Table 1
The results of the Shapiro-Wilk Test of normality

	Groups	Shapiro-Wilk		
		Statistic	*df*	Sig.
Pretest	Control	.942	10	.575
	Experiment A	.953	10	.703
	Experiment B	.882	10	.137
Posttest	Control	.916	10	.328
	Experiment A	.890	10	.172
	Experiment B	.882	10	.138

Table 2
The results of ANOVA for the three groups in the pretests and posttests

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Pretest	3.467	2	1.733	.914	.413
Posttest	242.867	2	121.433	84.940	.000

As can be observed in Table 2, the results of ANOVA show that the writing performance of students in the pretest was similar in all groups because at $F = 0.914$, Sig. is 0.413, which is larger than the alpha level (usually 0.05); therefore, it can be stated that there was no significant difference among the three groups in the pretest. This was predictable because the students were all at the same level of English proficiency and had not received any specific instruction before the treatment. In contrast, the results of the posttest showed a significant difference between the three groups in the posttest because, at $F = 84.940$, the Sig. is 0.000, which is lower than the *p*-value; therefore, it could be concluded that there were significant differences among the groups in the posttest. The mean differences in Table 3 provide helpful information for determining which group performed better.

Table 3
Descriptive statistics of three groups in the pretest and posttest

Groups		Pretest	Posttest
Control	Mean	7.8000	8.7000
	N	20	20
	Std. Deviation	1.31656	1.15950
Experiment A	Mean	7.0000	15.6000
	N	20	20
	Std. Deviation	1.15470	.84327
Experiment B	Mean	7.2000	11.3000
	N	20	20
	Std. Deviation	1.61933	1.49443

As can be observed in Table 3, all the groups showed progress from the pretest to the posttest, which is definitely due to the exposure to any training in the classroom. However, a quick look at the mean scores of control and experimental groups A and B revealed mean scores of 8.7, 15.60, and 11.30, respectively. These mean scores showed

that experimental group A, who received individual feedback on their writing from the researchers, performed far better than the learners of the other groups. In experimental group B, those exposed to EditGPT alone without assistance showed more progress than the control group. This revealed that EditGPT could be helpful in the process of writing improvement and make the learners more independent than the teacher.

4.2. Student perceptions

The second research question of this study was an attempt to understand students' perceptions of using EditGPT in the language learning process, especially in writing skills. The analysis of the perception questionnaire revealed the following results.

Table 4
The descriptive analysis of the perception questionnaire within experimental group B

	N	Min.	Max.	Mean	Std. Dev.
Perceived comprehensibility of EditGPT feedback					
Feedback is understandable by EditGPT	20	2	4	3.10	.876
Able to revise writing based on EditGPT feedback	20	4	5	4.30	.483
Feedback from EditGPT is clear	20	2	4	3.50	.707
Perceived advantages of EditGPT feedback for revision					
Feedback is useful in correcting grammatical mistakes	20	3	5	3.90	.568
EditGPT helps get higher marks in writing	20	2	5	3.80	.789
EditGPT helps increase the writing quality	20	3	5	4.10	.568
Perceived usefulness of EditGPT feedback to improve writing performance					
Feedback helps understand the writing issues	20	2	4	3.50	.707
EditGPT helps improve grammar	20	3	5	4.10	.568
EditGPT helps in vocabulary increase	20	2	2	2.00	.000
EditGPT helps the writing performance	20	3	5	3.90	.568

As shown in Table 4, the perceptions of experimental group B were thoroughly elicited. Based on Table 4, most students stated that they could modify their writing tasks based on the comments they received from EditGPT, as this statement had the highest mean score of 4.30 compared to the others. The second highest mean score was 4.10 for the statement, which shows students' optimistic belief that EditGPT could improve the quality of their writing tasks. The students also stated that receiving feedback from EditGPT helped them become aware of their writing problems (M = 3.50). In contrast, the two statements received the least attention from students. The first, with a mean score of 3.50, questioned the comprehensibility of the feedback students received from EditGPT; that is, students required further explanation of their writing tasks. Finally, the students believed that EditGPT did not assist them in increasing their vocabulary size, as this statement received the lowest mean score of 2.00 among many others.

5. Discussion

This study aimed to assess the feasibility of integrating an automated corrective feedback system into writing courses as an adjunct to teacher evaluations. This study investigated the concurrent impact of the EditGPT AI tool and instructor feedback on students' writing. Furthermore, this study aimed to determine learners' opinions on utilizing the EditGPT AI tool. While all experimental and control groups were provided with similar in-person tutorials on writing, experimental group A, which received individual feedback on their writing tasks, performed better than the other two groups. Simultaneously, experimental group B, which was exposed to EditGPT as the only source of feedback, performed better than the control group. Participants in the control group received general feedback on their tasks using abbreviations. Although the control group progressed from pretest to posttest, the results were not statistically significant. Therefore, it can be concluded that using EditGPT beside the teacher in class will greatly affect the performance of Omani EFL learners. Furthermore, students using EditGPT had favorable perceptions of using this AI tool. The overall instruction delivered to the students during class affected the outcome. This suggests that an automated corrective feedback system such as the EditGPT AI system implemented in this study could be a practical instrument in writing acquisition settings.

There could be some reasons for the better performance of those students who received higher grades after the human feedback sessions. Feedback from a human can be individualized and compassionate. The teacher could provide feedback based on the students' strengths and weaknesses in completing writing tasks. This could help the learners better understand the feedback, and this led to better retention of details in the posttest. This point could be considered as the weakness of automated corrective feedback by the application, as the tool was not able to distinguish the cognitive level of understanding among the students. Face-to-face interaction between the teacher and students was considered a reason for the better performance of students who received human feedback. Knowing that the written work was evaluated deeply and detailed feedback sessions were provided, students put more effort into writing the following written tasks during treatment. Another reason for the better performance of learners could be found in the immediate feedback and clarification that the teacher provided for the learners. Teachers and students could engage together via dialogue for further inquiries and explanations, which was missing while receiving feedback from EditGPT. Emotional intelligence could be another valid reason for better results in experimental group A. This could be explained by the teacher providing feedback in a specific way to control the emotional circumstances of students. He could be claimed to provide constructive feedback in a motivating and encouraging way rather than simply underlying the mistake.

The findings of this study align with those of several other studies (Wang et al., 2013; Li et al., 2015; Liao, 2016; Ranalli, 2018; Barrot, 2021) that have similarly documented the beneficial impacts of automated corrective feedback tools on students' writing. According to Wang et al. (2013), the implementation of AWE significantly improved both writing accuracy and comprehension of learner autonomy. Similarly, Li et al. (2015) developed the AWCF for students enrolled in ESL writing courses using the criterion method. The findings obtained by implementing a mixed-methods strategy indicated that the students maintained a favorable attitude towards Criterion, which could have contributed to their improved linguistic precision. In light of the scarcity of empirical research examining the effect of AWE tools on writing accuracy, Liao (2016) determined the effect of AWE tools on reducing different types of mistakes in corrections and new compositions using a time-series investigation methodology. Her research indicated that

Criterion could significantly reduce the frequency of L2 learners' writing errors. Barrot (2021) investigated the effects of AWCF on the writing accuracy of college L2 learners. The research discovered that the ability of AWCF to facilitate observation, offer adaptive metalinguistic clarification, and involve students in self-directed learning were the primary factors contributing to these results.

Several studies have arrived at findings that align with the second research objective (Fan, 2023; Thi & Nikolov, 2022; Fahmi & Cahyono, 2021; Tsalgini, 2019; Ayan & Erdemir, 2023; Ghufron, 2019; Wilson & Czik, 2016), namely that students and teachers have a positive perception toward the automatic corrective feedback tool. Fan (2023) investigated the impact of Grammarly's automated written corrections (AWCF) on the writing proficiency of students studying English as a Foreign Language. The findings revealed that students viewed Grammarly, an automated written corrective feedback (AWCF) instrument for grammar assistance, favorably, according to a subsequent survey. Another study by Thi and Nikolov (2022) investigated the integration of Grammarly's automated feedback with that of an instructor. 27 low intermediate-level EFL students represented a university in Myanmar. The results revealed a substantial enhancement in scores, and students generally held positive perceptions of the feedback provided by both the teacher and Grammarly. Similarly, Fahmi and Cahyono (2021) investigated EFL students' perspectives regarding written corrective feedback, which combines instructor feedback with the automatic writing evaluator software Grammarly. The study revealed that students positively perceived automated written corrective feedback (Grammarly) and considered combining instructor feedback and Grammarly to enhance their writing abilities. O'Neill and Russell (2019) found that students who received feedback from Grammarly and guidance from their advisors expressed their opinions more favorably.

6. Conclusion

The number of colleges worldwide that offer online courses is rising fast. Higher education establishments are constantly implementing online learning trends (Alhumsi & Alshaye, 2021). This study attempted to analyze the implementation of the EditGPT application in a language-learning context, particularly to determine its impact on the writing performance of Omani EFL learners. In addition, the study aimed to elicit students' perceptions of using EditGPT. This study included three groups of students. The control group received general feedback on the writing tasks, experimental group A received detailed individual feedback from the researchers, and experimental group B received feedback from EditGPT. After the statistical analysis, it was found that the group that received feedback from the researchers performed better than the other experimental groups that received feedback from EditGPT. However, both experimental groups performed far better than the control group did. So, it could be stated that EditGPT as a facilitator could function as a complementary application to teachers. This will help students to become more independent of the learning process. The students' perceptions in experimental group B also revealed that using technological instruments, especially in writing skills, helped facilitate autonomy among the students.

The results of this study convey some messages to teachers, students, and institutions. Due to the lack of adequate teaching time and probably the crowdedness of some classes, teachers cannot provide detailed individual feedback for their students. In this case and based on the findings of this study, using tools such as EditGPT indeed plays a vital learning role in the process. This type of technological implementation reduces the

burden on teachers. However, the study's results stated that the teacher's role as the primary source of feedback could not be entirely ignored. Students can use these applications and tools to improve their writing quality. These tools provide syntactical and semantic modifications based on the learners' mistakes and errors. Therefore, students can avoid repeating these mistakes in the following writing tasks by paying attention to various correction areas. In addition, these instruments help learners access a source of feedback without any time and place limitations and any cognitive pressure in the language learning context. Since the results confirmed the usability of EditGPT in the educational context, the institution must be able to facilitate access to similar tools to move toward being a learner-centered institution.

The present study has some concerns, which could be considered limitations and suggestions for further research:

- One of the major limitations of this study was its small sample size. Sixty students were selected for this study based on convenience random sampling; further studies with a larger population are suggested for similar studies.
- This study targeted pre-intermediate English language proficiency candidates at one of the higher education universities in Oman. To better understand Omani EFL contexts and students using technological devices, further research should be conducted with all proficiency-level students in different institutions in Oman.
- Since the students were engaged in this study beyond their regular language learning classes, unseen factors such as tiredness, hunger, and lack of motivation could affect the study's results; therefore, more studies are suggested with groups of students more focused on treatment implementation.
- The primary focus of this study was EditGPT as a correction tool for EFL learners. Since there are various AI tools such as ChatGPT, Grammarly, and PaperPal, more studies could be conducted to measure their effect on writing performance.
- Finally, the present study used the AI tool to investigate the learners' writing performance, while other AI tools helped give feedback on other skills, such as reading, listening, and speaking. Further studies could be conducted to measure the possibility of using such tools in the skills above.

Author Statement

The authors declare that there is no conflict of interest.

ORCID

Ali Al Ghaithi https://orcid.org/0000-0002-9653-508X

Behnam Behforouz https://orcid.org/0000-0002-0078-2757

References

Ahmed, M. M. H., McGahan, P. S., Indurkhya, B., Kaneko, K., & Nakagawa, M. (2021). Effects of synchronized and asynchronized e-feedback interactions on academic writing, achievement motivation and critical thinking. *Knowledge Management & E-*

Learning, 13(3), 290–315. https://doi.org/10.34105/j.kmel.2021.13.016

Alhumsi, M. H., & Alshaye, R. A. (2021). Applying technology acceptance model to Gauge University students' perceptions of using blackboard in learning academic writing. *Knowledge Management & E-Learning, 13*(3), 316–333. https://doi.org/10.34105/j.kmel.2021.13.017

Ayan, A. D., & Erdemir, N. (2023). EFL teachers' perceptions of automated written corrective feedback and Grammarly. *Ahmet Keleşoğlu Faculty of Education Journal, 5*(3), 1183–1198. https://doi.org/10.38151/akef.2023.106

Banaruee, H. (2016). *Recast in writing*. Sana Gostar Publications.

Barrot, J. S. (2021). Using automated written corrective feedback in writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(4), 584–607. https://doi.org/10.1080/09588221.2021.1936071

Benali, A. (2021). The impact of using automated writing feedback in ESL/EFL classroom contexts. *English Language Teaching, 14*(12), 189–195. https://doi.org/10.5539/elt.v14n12p189

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge. https://doi.org/10.4324/9780203832400

Burkhart, C., Lachner, A., & Nückles, M. (2020). Assisting students' writing with computer-based concept map feedback: A validation study of the CohViz feedback system. *Plos One, 15*(6): e0235209. https://doi.org/10.1371/journal.pone.0235209

Burstein, J., Chodorow, M., & Leacock, C. (2003, August). CriterionSM online essay evaluation: An application for automated evaluation of student essays. In *Proceedings of the AAAI Conference on Innovative Applications of Artificial Intelligence Conference*. Association for the Advancement of Artificial Intelligence. Retrieved from https://aaai.org/papers/00003-iaai03-001-criterionsm-online-essay-evaluation-an-application-for-automated-evaluation-of-student-essays/

Chen, C. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112. Retrieved from https://www.lltjournal.org/item/10125-44145/

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education, 34*, 18–27. https://doi.org/10.1016/j.iheduc.2017.04.002

Cheung, Y. L. (2015). Feedback from automated essay evaluation systems: A review of selected research. *TESL Reporter, 48*(2), 1–17. Retrieved from https://lir.byuh.edu/index.php/Issue1/article/view/263

Fahmi, M. A., & Cahyono, B. Y. (2021). EFL students' perception on the use of Grammarly and teacher feedback. *Journal of English Educators Society, 6*(1), 18–25. https://doi.org/10.21070/jees.v6i1.849

Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open, 13*(2). https://doi.org/10.1177/21582440231181296

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing, 16*(3), 165–193. https://doi.org/10.1016/j.jslw.2007.07.003

Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime…?). *Journal of Second*

Language Writing, 13(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005
Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum R., & Köller, O. (2020a). Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study. *Assessing Writing, 43*: 100420. https://doi.org/10.1016/j.asw.2019.100420
Fleckenstein, J., Meyer, J., Jansen, T., Keller, S., & Köller, O. (2020b). Is a long essay always a good essay? The effect of text length on writing assessment. *Frontiers in Psychology, 11*: 562462. https://doi.org/10.3389/fpsyg.2020.562462
Frankenberg-Garcia, A. (2020). Combining user needs, lexicographic data and digital writing environments. *Language Teaching, 53*(1), 29–43. https://doi.org/10.1017/S0261444818000277
Gao, J., Li, X., Gu, P., & Liu, Z. (2020). An evaluation of China's automated scoring system Bingo English. *International Journal of English Linguistics, 10*(6), 30–39. https://doi.org/10.5539/ijel.v10n6p30
Ghufron, M. (2019, April). Exploring an automated feedback program 'Grammarly' and teacher corrective feedback in EFL writing assessment: Modern vs. traditional assessment. In *Proceedings of the 3rd English Language and Literature International Conference* (pp. 307–315). EAI. https://doi.org/10.4108/eai.27-4-2019.2285308
Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment, 8*(6). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1625
Hammoda, B. (2024). The impact of educational technologies on entrepreneurial competencies: A systematic review of empirical evidence. *Knowledge Management & E-Learning, 16*(2), 309–333. https://doi.org/10.34105/j.kmel.2024.16.015
Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching, 14*(1), 15–26. https://doi.org/10.1080/17501229.2018.1471083
Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research, 24*(4), 519–539. https://doi.org/10.1177/1362168818802469
Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266. https://doi.org/10.1080/00461520903213600
Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173–196. https://doi.org/10.2190/EC.42.2.c
Khadawardi, H. A. (2020). The effect of implicit corrective feedback on English writing of international second language learners. *English Language Teaching, 14*(1), 123–139. https://doi.org/10.5539/elt.v14n1p123
Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*: 100450. https://doi.org/10.1016/j.asw.2020.100450
Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing, 38*(1), 62–85. https://doi.org/10.1177/0265532220929918
Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal, 63*(1), 13–22. https://doi.org/10.1093/elt/ccn010
Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Li, Z., Link, S., Ma, H., Yang, H., & Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System, 44*, 66–78. https://doi.org/10.1016/j.system.2014.02.007

Li, Z., Feng, H. H., & Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *Calico Journal, 34*(3), 355–375. https://doi.org/10.1558/cj.26382

Liao, H. C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal, 70*(3), 308–319. https://doi.org/10.1093/elt/ccv058

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing, 23*, 35–59. https://doi.org/10.1016/j.asw.2014.09.002

ONeill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1), 42–56. https://doi.org/10.14742/ajet.3795

Perelman, E. (2014). When "the state of the art" is counting words. *Assessing Writing, 21*, 104–111. https://doi.org/10.1016/j.asw.2014.05.001

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2016). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology, 37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

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

Rich, C. S. (2012). The impact of online automated writing evaluation: A case study from Dalian. *Chinese Journal of Applied Linguistics, 35*(1), 63–79. https://doi.org/10.1515/cjal-2012-0006

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667190

Rupp, A. A., Casabianca, J. M., Krüger, M., Keller, S., & Köller, O. (2019). Automated essay scoring at scale: A case study in Switzerland and Germany. *ETS Research Report Series, 2019*(1), 1–23. https://doi.org/10.1002/ets2.12249

Shermis M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing, 20*, 53–76. https://doi.org/10.1016/j.asw.2013.04.001

Shermis, M. D., Garvan, C. W., & Diao, Y. (2008, March). The impact of automated essay scoring on writing outcomes. In *Proceeding of The Annual Meetings of the National Council on Measurement in Education (NCME) 2008.* National Council on Measurement in Education. Retrieved from https://files.eric.ed.gov/fulltext/ED501148.pdf

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

Thi, N. K., & Nikolov, M. (2022). How teacher and grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher, 31*(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Tsalgini, C. (2019). *EFL vocabulary acquisition through digital storytelling: The case of Little Prince*. Doctoral dissertation, Aristotle University of Thessaloniki, Greece. Retrieved from https://ikee.lib.auth.gr/record/314198/files/GRI-2020-26662.pdf

Wang, Y. J., Shang, H. F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning, 26*(3), 234–257. https://doi.org/10.1080/09588221.2012.655300

Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing, 44*: 100449. https://doi.org/10.1016/j.asw.2020.100449

Wang, L., Zhang, G., Chen, J., Lu, X., & Song, F. (2022). The territory effect: How awe reduces territoriality and enhances sharing intention. *Journal of Business Research, 148*, 1–11. https://doi.org/10.1016/j.jbusres.2022.04.014

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education, 100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Woodworth, J., & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom. *TESL Canada Journal, 37*(2), 234–247. https://doi.org/10.18806/tesl.v37i2.1340

Zhang, X. (2017). Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research, 21*(2), 217–240. https://doi.org/10.1177/1362168815623291