The impact of automated writing evaluation (AWE) on EFL learners' peer and self-editing

Fatima Abdullah Yahya Al-Inbari[1] · Baleigh Qassim Mohammed Al-Wasy[2]

Abstract

Automated Writing Evaluation (AWE) is one of the machine techniques used for assessing learners' writing. Recently, this technique has been widely implemented for improving learners' editing strategies. Several studies have been conducted to compare self-editing with peer editing. However, only a few studies have compared automated peer and self-editing. To fill this research gap, the present study implements AWE software, WRITER, for peer and self-editing. For this purpose, a pre-post quasi-experimental research design with convenience sampling is done for automated and non-automated editing of cause-effect essay writing. Arab, EFL learners of English, 44 in number, have been assigned to four groups: two peer and self-editing control groups and two automated peer and self-editing experimental groups. There is a triangulation of the quasi-experimental design with qualitative data from retrospective notes and questionnaire responses of the participants during and after automated editing. The quantitative data have been analyzed using non-parametric tests. The qualitative data have undergone thematic and content analysis. The results reveal that the AWE software has positively affected both the peer and self-editing experimental groups. However, no significant difference is detected between them. The analysis of the qualitative data reflects participants' positive evaluation of both the software and the automated peer and self-editing experience.

✉ Baleigh Qassim Mohammed Al-Wasy
  baleigh5112@gmail.com

  Fatima Abdullah Yahya Al-Inbari
  faalnbari@nu.edu.sa; fatimainbari77@gmail.com

1 English Department, College of Languages and Translation, Yadma Branch, Najran University, King AbdulAziz Road, P.O. Box. 1988, Najran, Saudi Arabia

2 Department of English, College of Education and Human Sciences, Sana'a University, Sana'a, Yemen

1 Introduction

Automated writing evaluation (AWE) is not a newly invented technique in this century. It had been in use since the 1960s (Chen & Cheng, 2008). AWE systems "recognize certain types of errors and offer automated feedback on correcting these errors, in addition to providing global feedback on content and development" (Weigle, 2013, p. 47). This technology is originally developed to enlighten the large amounts of grading of student essays that is time consuming. Later, writing instructors and researchers implement it in improving and investigating the editing processes of students.

Advocates in writing theory and research adopt this technique in the process of editing and feedback. Teachers encourage students to write several drafts of their papers, substantially revise them, and give students formative feedback (Ferris, 2003). Editing is defined as "manipulating a text in such a way that it yields a product which is as correct as possible and thus contains the fewest errors possible," (DePoel, Carstens, & John, 2012, p. 6). However, the view of editing varies from considering it as correction of formal aspects of writing such as correcting grammar and spelling errors to correction of form, content, and organization. Mahendran (2012) sums up the purpose of editing in decreasing "ambiguities and anomalies" in a written text and improving its "readability and acceptability in terms of the writer's goals and intentions". Peer editing and self-editing, whatever be the names given to these processes of writing by different researchers, have their theoretical basis in the Noticing Hypothesis. Schmidt (1990) and Schmidt (1993) indicate that the foundation of Noticing Hypothesis is that if L2 writers identify errors, this helps them realize gaps in their interlanguage and enhances positive learning of the L2. As L2 writers go through their written drafts or through their peers' edits of their written drafts, they realize the mistakes. If editing occurs on a regular basis, it is likely to lead to skill acquisition (i.e., the transfer of declarative knowledge into automatic use of editing for improving writing).

Peer editing is defined as an activity "for peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others of similar status," (Topping, Smith, Swanson, & Elliot, 2000, p. 150). Self-editing is a similar process to peer editing though it is done by the writer himself/herself. Certainly, there are many differences between the two in the aims and the outcome, but they are similar in nature. However, discussing these differences is out of the scope of this study.

There is a large body of research on automated feedback. However, very few studies examined the impact of automated feedback on peer versus self-editing. They followed different research procedures, focused on different aspects of writing ability, used different writing genres, or evaluated the impact of automated editing on learning or learners' attitudes. Consequently, their results are varying and non-confirmatory. Therefore, there is a growing need to strengthen results reported by this growing body of research. The present study from a process-oriented approach attempts to fill in this research gap. It investigates the effect of an AWE program on students' peer and self-editing. It also examines how students perceive and evaluate the software as well as their automated editing experience.

2 Literature review

2.1 Peer editing and self-editing

Several studies explored the impact of peer and self-editing on writing. Abadikhah and Yasami (2014), Prabasiwi and Warsono (2017), and Winarto (2018) examined the effect of peer and self-editing strategies on learners' writing skills. All the afore-mentioned studies confirmed a positive impact of both the editing strategies on linguistic accuracy as well as writing performance. Other studies had the aim to compare non-automated editing perspectives (peer, self-, and teacher editing) with each other; and examined their impact on writing. Diab (2010) and Khaki and Biria (2016) compared the role of peer and self-editing strategies, whereas Hemati (2012) compared the impact of the three strategies on EFL learners' writing. The first two studies concluded that self-editors showed a better improvement and were able to detect more errors. On the other hand, the results of the third study showed that teacher editing was the best among the three editing techniques, and peer editors showed better improvement in grammatical accuracy, when compared with self-editors. To conclude, most of the studies have confirmed the efficacy of non-automated peer and self-editing for improving writing. However, a final decision about which has better impact on writing is yet to be reached.

Within the growing recent awareness of the merits of automated corrective feed-back in writing classroom, many researchers investigated automated peer editing as well as self-editing. These were applied in writing research within the general frame-works of CAI (computer-assisted instruction), CALL (computer-assisted language learning), and TELL (technology-enhanced language learning). The use of technology was not new in the field of language learning. Thouësny & Bradley (2011) traced its introduction to Burn's work in 1979. Another important framework for the integration of technology in language learning is CM (computer-mediation) for language learning. Within this framework, the concept of affordance, multimodality, and multiliteracies were utilized for providing theoretical explanations for the role of technology in enhancing skill acquisition. (Lamy & Hampel, 2007).

Hoang (2019) reasonably classified research on AWE into two main areas: learner-centric and system-centric. Studies that focused on features of feedback or mechanism of scoring of AWE programs were named system-centric by him whereas studies that concentrated on learners' attitudes, evaluation, or response to automated feedback were called student-centric. With respect to this classification, the present study belongs to the student-centric wing of AWE studies. An insightful classification of the directions of student-centric studies was that of Lai's (2010) study. Lai divided the effects of AWE on EFL writers into (1) the product perspective (i.e., the impact of AWE on learners' final written products); (2) the process perspective (i.e., the impact of AWE on the teaching, editing, and learning processes); and (3) the perception perspective (i.e., the students' and/or teachers' attitudes towards AWE). This could serve as a meaningful classification of student-centric studies of AWE followed by authors like Stevenson and Phakiti (2014).

2.2 Automated writing evaluation (AWE)

Automated Writing Evaluation (AWE) has been considered one type of machine scoring. It has been used to score learners' writing, with regard to particular criteria. Though it has been under development since the 1960s, its use in assessment and instruction remained controversial. According to Warschauer and Ware (2006), there were three main programs for AWE. These three programs were: the Intelligent Essay Assessor (IEA), MY Access, and Criterion. These programs did not only have automated scoring and feedback, but they also included "model essays, scoring rubrics, graphic organizers, dictionaries and thesauri," (Warschauer & Ware, 2006, p. 162). Besides, some of these programs provided facilities for both teachers and learners. These facilities included providing immediate feedback, highlighting particular sections, allowing students to access samples of writing and web-based dictionaries, uploading work and creating portfolios, and allowing teachers to add additional feedback comments and track a students' writing progress and grades. (Hockly, 2019)

Several studies have been carried out to explore the effect of AWE on the learners' writing performance. Some of these studies investigated the effect of a using particular AWE mode on the writing performance (Mohsen & Alshahrani, 2019; Wang, Shang, & Briody, 2013; Wang & Wang, 2012). These studies showed a significant improvement in the performance of learners who used AWE. Some other studies compared teacher-only-feedback with teacher feedback when joined with AWE. Link, Mehrzad, and Rahimi (2020) indicated a clearer improvement with teacher feedback accompanied with AWE. On the other hand, other studies reported no benefit of using AWE in writing. For example, Huang and Renandya (2020) investigated the effect of automated feedback on university learners with lower-language level in China. The results of this study revealed that the integration of automated feedback did not always have a positive effect on the learners' final draft.

2.3 Automated peer and self-editing in writing classrooms

The literature reviewed below with relevance to the use of technology in assistance of self- and peer editing would be presented as two main research trends. The first research trend evaluated the impact of one automated editing strategy on writing in comparison to a non-automated editing strategy. The second research trend compared different automated editing strategies with each other in terms of their impact on writing.

Researchers of the first trend used various technologies such as Google documents (Daweli, 2018; Saricaoglu & Bilki, 2021), web-based feedback (Yang & Meng, 2013; Wang, 2013), and phone applications (Li & Hegelheimer, 2013; Al-Wasy & Mahdi, 2016) to assist either peer or self-editing in writing classrooms. The findings of these studies proved the learners' progress in the two editing strategies, and consequently in the learners' writing performance. A considerable number of studies have been conducted to investigate the factors that affect leaners' use of AWE tools and how the use of these tools may improve the learners' final product (Li, 2021; Chen et al., 2022; Liu & Yu, 2022). These studies showed that the nature of the used device and the ease of use were the most important factors behind learners' satisfaction with

AWE. They also assured the overall improvement in learners' final product. To conclude, the majority of studies in the first trend focused on different groups of learners, different aspects of writing abilities, or the merits of the used program. They reported improvements in writing due to the use of automated self- or peer editing.

The second trend of studies compared cooperatively and individually presented automated corrective feedback. Elola and Oskoz (2010) held a comparison between performances of peer-writers and individual writers using wikis and chats. They detected no statistically significant differences in aspects of accuracy, complexity, and fluency when comparing the individuals' and peer-writers' assignments. Hojeij and Hurley (2017) used a triangulated research design to gauge the impact of apps (namely Edmodo, Notability, and Powtoon) on peer and self-editing processes as well as learners' motivation and engagement. Collected data were from a questionnaire, unstructured interviews, narrative practices, and written performance before and after the treatment. The study revealed an overall improvement in writing quality. Students were motivated and the majority had high opinions of using the three-flip technology for editing their writings. Cautions were directed by researchers that future use of the three-flip technology should be accompanied by proper training on technology use as well as proper guidance for the editing processes. Tavşanlı and Kara (2021) compared self-editing to peer editing of 60 Turkish, fourth-grade students with relevance to achievement in following spelling and punctuation rules. Their research design was a mixture of qualitative and quantitative aspects. The qualitative aspect was implemented for selection criteria and analyzing written texts whereas the qualitative aspect was used for evaluating participants' attitudes and overall evaluation of the editing experience. Experimental and control groups were assigned four writing topics; the first was considered a pre-test and the last a post-test. The experimental group was trained in detecting and correcting spelling and punctuation errors while the control group did not receive similar training. The experimental group texts were much better than texts produced by the control group in terms of spelling and punctuation. Feng and Chukharev-Hudilainen (2022) adopted a genre-based approach to gauge the efficacy of a genre-based AWE system in improving graduate engineers' rhetorical moves in their research abstracts. Data collected from pre- and post-drafts and interviews with participants confirmed the efficacy of the AWE in enhancing the fore-mentioned genre-features. To conclude this review of the comparative trend, it is clear that the majority of these studies have not reached a decisive conclusion about which editing procedure was better for writing.

To conclude, the majority of the studies that focused on automated self- and peer editing identified improvements in some aspects of writing due to automated self- or peer editing. However, the comparative studies that targeted automated self- vis-à-vis peer editing are rather few. More studies are needed to accelerate trends in this direction and make a sound contribution to research on comparative automated editing. It is still unconfirmed which automated editing strategy is better for which learners or for which writing activity. The present study is an attempt to fill this gap with relevance to EFL Saudi context by using an app that has not been used before in comparative automated editing research. The study attempts to answer the following research questions:

1. Is there a significant difference in the quality of writing of the self-editing control and experimental groups after the employment of AWE?
2. Is there a significant difference in the quality of writing of the peer editing control and experimental groups after the employment of AWE?
3. a) Is there a significant difference in the quality of writing of the self-editing experimental group before and after the employment of AWE?
   b) Is there a significant difference in the quality of writing of the peer editing experimental group before and after the employment of AWE?
4. Is there a significant difference in the quality of writing of the self-editing and peer editing experimental groups after the employment of AWE?
5. How do experimental groups perceive and evaluate the software as well as their automated editing experiences?

3 Research methodology

To investigate the impact of AWE software on EFL learners' peer editing and self-editing, the researchers used a quasi-experimental, comparative design with convenience sampling. There was a triangulation for this investigation with qualitative data from participants' responses to a questionnaire and a retrospective note question. The experiment entailed doing pre-test and post-test cause-effect essay writing. The retrospective note question was to be answered during editing. The questionnaire was to be answered after completing editing by the experimental groups.

3.1 Participants

The participants in this study consisted of forty-four EFL Saudi students (9 male and 35 female) selected, under their consent, from the students of level 8, Department of English in two universities: Najran University and University of Bisha. The adoption of level 8 was due to the students' completion of five writing courses. Therefore, it was assumed that they would be more familiar with the stages of writing and the editing process. All participants studied English for seven semesters in the English department, in addition to six years of studying English at intermediate and secondary schools. It is worth mentioning that none of them has ever been to any of the English-speaking countries.

3.2 The software

Based on the researchers' guidance, the software used by the research subjects was WRITER. They used it to revise and edit their peers' essays or their own ones. The researchers recommend the software for the excellent editing affordances it provides. It corrects various types of errors including grammatical, spelling, and punctuation errors. It uses deep grammar error correction when dealing with these errors. For example, when dealing with grammatical errors, it does not only correct errors but also provides learners with many examples of the correct use of a particular grammatical rule. Moreover, it underlines an error and presents various corrections for this

error. One must only click on the most appropriate correction to apply it in the text. After one completes correcting all errors, the software goes through the whole text for the purpose of proofreading before submission.

The software also has other important features, such as clarity, readability, terminology, writing style, tone, and uniqueness. It also includes simple additional tools such as a plagiarism checker, style guide builder, tone detector, etc. Moreover, it allows sharing team content. Therefore, it is possible to use it individually, in pairs, or in teamwork. It also focuses on style, whether formal or informal, and it allows users to save different styles and terminology that can be re-used later. Consequently, using an editing tool with the above features may help learners improve their writing skills. Learners can edit their essay using online feedback given by WRITER.

3.3 Data collection tools

To answer research questions, researchers used three tools for data collection: a test (writing an essay), a questionnaire, and a retrospective note question. The test consisted of only one question, asking students to write a cause-effect essay on the effects of coronavirus on people in Saudi Arabia. Researchers opted cause-effect essay genre because it was one of the required essay types in the writing syllabuses at the two universities from which participants were selected. Students were familiar with this essay type. They were required to write a five-paragraph, cause-effect essay. The introduction paragraph introduces background of Corona Virus and its major effects on Saudi people's lives. The three supporting paragraphs detail the effects it induces on Saudis' lives. The concluding paragraph summarizes the main idea and the major effects explained in the supporting paragraphs. Two other research tools were used to enrich the study with more analytical directions: a questionnaire (administered after editing) and a retrospective notes' question (directed during editing). Two other research tools were used to enrich the study with more analytical directions: a questionnaire which was administered after editing and a retrospective notes' question which was directed during editing. These two tools served as complimentary, qualitative devices for the quantitative aspect of the study.

The questionnaire was an open-ended questionnaire. It consisted of two parts: A and B. Part A included 4 questions about whether participants liked writing; their average grade in writing courses; if they edited for self or others and the frequency of previous editing experiences; and whether they had previously used an editing program. Part B listed five questions. Question 1 asked about the nature of help participants got from the program. Question 2 inquired about the features they found most helpful in the app. Question 3 interrogated the most difficult aspect(s) of the program. Question 4 asked participants to opt one of two descriptions for their electronic editing experience and to justify their choice. Question 5 requested them to add any further remarks they feel like adding. The questionnaire was administered at the end of the treatment.

The retrospective notes, on the other hand, answered one question. It requested participants to describe their electronic editing experience while editing.

3.4 Procedures

To achieve the aims of this study, the following procedures were followed:

1. The participants were divided into four groups: one control group for peer editing (n=8), one experimental group for peer editing (n=14), one control group for self-editing (n=9), and one experimental group for self-editing (n=13). Before commencing the experiment, all the groups were asked to write an essay about the effect of Corona virus on people in Saudi Arabia. This task was used as a pre-test to check their writing performance.
2. The essays of the pre-test were scored according to rubrics developed by Klimova (2011). These rubrics were developed according to Bacha's model (2001) – Jacobs (1981). In evaluating essays, these rubrics adopt five elements with percentages: content 30%, organization 20%, vocabulary 20%, language use 25%, and mechanics 5%.
3. A virtual classroom was created for the two experimental group members to explain the features of the software (WRITER), how to use it, how to deal with the program feedback, and the different functions it has.
4. The members of peer editing experimental group were asked to upload their peers' essays to the software (WRITER) and members of the self-editing experimental group were directed to upload their own ones. Then the two groups had to receive the program feedback, make all necessary corrections, and finally copy the final drafts of the edited essays as they appear in the program.
5. The retrospective note question was introduced in this stage to the experimental groups.
6. The peer editing control group members were asked to edit their peers' essays and the self-editing control group members were asked to edit their own ones.
7. The final drafts of the essays of all control and experimental groups were considered as a post-test of the treatment, and they were scored according to rubrics developed by Klimova (2011).
8. When the post-test essays were submitted, the questionnaire was introduced to the participants in the experimental groups.

3.5 Data analysis

In this research paper, the data obtained from pre- and post-tests was analyzed statistically using SPSS (version 23). The analysis also involved the data obtained from the questionnaire and the retrospective notes. To achieve coding reliability, the essays of the pre- and post-tests were evaluated by the two researchers and one more colleague by using the previously mentioned rubrics by Klimova (2011). Inter-rater agreement was evaluated according to Cohen (1988). The inter-rater agreement was 88% for the peer editing control group, 92% for the self-editing control group, 93% for the peer editing experimental group, and 90% for the self-editing experimental group. These numbers indicated that the inter-rater agreement was very good. Then the average of the three corrections was calculated for every essay. Only the average of the scores was used in the analysis. To test the homogeneity of the four groups before

conducting the experiment, the pre-test scores of the participants were compared by using the Kruskal-Wallis Test. The selection of this test was due to the nature of the samples. According to Field (2013, p. 415), Kruskal-Wallis Test "assesses the hypothesis that multiple independent groups come from different populations." Moreover, the researchers used non-parametric tests in different statistical analyses. This can be attributed to the small size of the research samples. As mentioned by Field (2013, p. 140) "for small samples, the sampling distribution is not normal; it has a t-distribution." Non-parametric tests can be used with small samples because "they make fewer assumptions than the other tests," (Field, 2013, p.381). The researchers adopted the level of significance at 0.05 for all statistical analyses.

To address the first research question, scores of the self-editing control and experimental groups in the post-test were analyzed. The scores of the self-editing experimental group were compared with the scores of the self-editing control group by using the Mann-Whitney test, one of the non-parametric tests. According to Field (2013, p. 398) the Mann-Whitney test "works by looking at differences in the ranked positions of scores in different groups." To answer the second research question, the scores of the peer editing control and experimental groups in the post-test were analyzed. The scores of the peer editing experimental group were compared with the scores of the peer editing control group by using the Mann-Whitney test. To tackle questions (3a) and (3b), the scores of the pre-test and post-test for self-editing and peer editing experimental groups were compared by using Wilcoxon test. Wilcoxon test "is used in situations in which there are two sets of scores to compare, but these scores come from the same participants," (Field, 2013, p. 403). To address the fourth question, the scores of the peer editing experimental group were compared with the scores of the self-editing experimental group by using the Mann-Whitney test.

To measure the effect size, researchers followed the form of Cohen's model (1988) of measuring the effect size (eta squared $\eta2$). In the form of Cohen's $\eta2$, the effect size can be reported as small if the $\eta2$ value is (0.01), medium with a (0.06) $\eta2$ value, and indicates a large effect size with (0.14) $\eta2$ value (Ellis, 2010, P. 41).

To deal with the fifth research question, thematic content analysis (TCA) was used to gauge participants' responses to both the questionnaire and the retrospective notes. TCA is a qualitative research method used for detecting, analyzing, and reporting patterns within collected data. It involves transcribing data, labeling, or coding them, detecting similar, and/or contrasting patterns, and finally reporting them. (Braun & Clarke, 2006).

4 Results

Table 1 presented the summary statistics for the pre-test. It showed that the level of the participants prior to the experiment was quite similar. By using the Kruskal-Wallis test, it was revealed that the pre-test scores of the participants in the four groups were close to each other: the self-editing experimental group (N=13, M=26.46), the self-editing control group (N=9, M=20.56), the peer editing experimental group (N=14, M=22.96), and the peer editing control group (N=8, M=17.44). Data from this table also showed that there was no significant difference (p=0.433) in the pre-test scores

Table 1 Test of Homogeneity: Kruskal-Wallis Test

Measurement	Groups	N	Mean Rank	K.W (H)	Df	Sig.
Pre-test	Self-editing cont.	9	20.56	2.744	3	0.433
	Peer editing cont.	8	17.44			
	Self-editing exp.	13	26.46			
	Peer editing exp.	14	22.96			
	Total	44				

Table 2 The scores of the control and experimental self-editing groups

Measurement	Groups	N	Mean Rank	Sum of Ranks	Mann-Whitney U	Sig .
Post-Test	Self-editing cont.	9	7.22	65	20	0.009
	Self-editing exp.	13	14.46	188		
	Total	22				

Table 3 The scores of the control and experimental peer editing groups

Measurement	Groups	N	Mean Rank	Sum of Ranks	Mann-Whitney U	Sig .
Post-Test	Peer editing cont.	8	4.88	39	3	0.0001
	Peer editing exp.	14	15.29	214		
	Total	22				

between the two experimental and the two control groups on the variable of edited texts (whether peer or self). This indicated the homogeneity of the participants.

4.1 Q1. Is there a significant difference in the quality of writing of the self-editing control and experimental groups after the employment of AWE?

To answer the first question, the Mann-Whitney test was used to compare post-test scores obtained by participants in the self-editing control and experimental groups. As indicated in Table 2, mean rank of participants in the self-editing control group was (7.22) whereas mean rank of participants in the self-editing experimental group was (14.46). The table also showed that there was a statistically significant difference between the mean scores of the post-test for the self-editing control and experimental groups (U=20, p=0.009), in favor of the experimental group. This indicates that the use of the AWE software had a positive effect on the writing quality of the self-editing experimental group, when compared with the self-editing control group. The effect size was (0.28) which indicated a large effect of using AWE on the self-editing experimental group according to the scale of eta squared.

4.2 Q2. Is there a significant difference in the quality of writing of the peer editing control and experimental groups after the employment of AWE?

To answer the second question, the Mann-Whitney test was used to compare post-test scores obtained by participants in the peer editing control and experimental groups. As indicated in Table 3, mean rank of participants in the peer editing control group was (4.88) whereas mean rank of participants in the peer editing experimental group

Table 4 The scores of the pretest and post-test of the experimental self-editing group

(Pre-test)-(Post-Test)	N	Mean Ranks	Sum of Ranks	Wilcoxon(Z)	Sig.
Negative Ranks	13[a]	7	91	3.18	0.001**
Positive Ranks	0[b]	0	0		
Ties	0[c]				
Total	13				

Note: a: Pre-test<Post-test

b: Pre-test>Post-test

c: Pre-test=Post-test

Table 5 The scores of the pretest and post-test of the experimental peer editing group

(Pre-test)-(Post-Test)	N	Mean Ranks	Sum of Ranks	Wilcoxon (Z)	Sig.
Negative Ranks	14[a]	7.5	105	3.3	0.001**
Positive Ranks	0[b]	0	0		
Ties	0[c]				
Total	14				

Note: a: Pre-test<Post-test

b: Pre-test>Post-test

c: Pre-test=Post-test

was (15.29). The results revealed that there was a statistically significant difference between mean scores of the post-test for the peer editing control and experimental groups (U=3, p=0.0001), favoring the experimental group. This indicates that the use of the AWE software had a positive effect on the writing quality of the peer editing experimental group, when compared with the peer editing control group. The effect size was (0.69) which indicated a large effect of using AWE on the peer editing experimental group according to the scale of eta squared.

4.3 Q3a. Is there a significant difference in the quality of writing of the self-editing experimental group before and after the employment of AWE?

To answer question 3a, the Wilcoxon test was used to compare pre-test with post-test scores obtained by participants in self-editing experimental group. Wilcoxon test revealed a significant difference between mean pre-test and post-test scores for the self-editing experimental group (z=3.18, p=0.001), in favor of the post-test. As shown in Table 4, pre-test scores were less than post-test scores for all participants in the self-editing experimental group. This indicates that participants' writing performance has improved in the post-test, in comparison with their performance in the pre-test.

4.4 Q3b. Is there a significant difference in the quality of writing of the peer editing experimental group before and after the employment of AWE?

To answer question 3b, the Wilcoxon test was used to compare pre-test with the post-test scores obtained by participants in the peer editing experimental group. Wilcoxon test showed a significant difference between mean pre-test and post-test scores for peer editing experimental group (z=3.3, p=0.001), in favor of the post-test. As shown in Table 5, pre-test scores were less than post-test scores for all participants of the peer editing experimental group. This indicates that participants in peer editing

Table 6 The scores of the post-test of self-editing and peer editing experimental groups

Measurement	C	N	Mean Rank	Sum of Ranks	Mann–Whitney U	Sig .
Post-Test	Self-editing exp.	13	12.50	162.5	71.5	0.35
	Peer editing exp.	14	15.39	215.5		
	Total	27				

experimental group have better writing performance in the post-test, when compared with their performance in the pre-test.

4.5 Q4. Is there a significant difference in the quality of writing of the self-editing and peer editing experimental groups after the employment of AWE?

To answer question 4, the Mann-Whitney test was used to compare post-test scores obtained by participants in self-editing and peer editing experimental groups. As indicated in Table 6, mean rank of participants in the self-editing experimental group was (12.50) whereas mean rank of participants in peer editing experimental group was (15.39). The difference between self-editing experimental group and peer editing experimental group was not significant (U=71.5, p=0.35). This indicates that the use of the AWE software has an approximately equal effect on the writing quality of self-editing and peer editing experimental groups.

4.6 5) How do experimental groups perceive and evaluate the software as well as their automated editing experiences?

Researchers found thematic analysis a useful procedure to analyze the data collected based on the questionnaire and participants' retrospective notes. Braun and Clarke (2006) defined thematic analysis as 'method for identifying, analyzing, and reporting patterns (themes) within data'. To do the thematic and content analysis of the data, researchers followed the following steps based on Braun and Clarke's (2006) framework for doing thematic analysis: becoming familiar with data, generating initial codes, searching for themes, reviewing themes, defining themes, and writing them up (Maguire & Delahunt, 2017, p. 3345; Kiger & Varpio, 2020, p. 4–5).

Responses to the two parts of the questionnaire and those of the retrospective notes question formed the answer of the fifth research question. The responses to questions in part A were as follows. All peer and self-editors except two had responded positively to whether they like writing. To question 2, all of them responded that they had got good grades ranging between excellent and good except 3 participants who had got low grades. Responding to question 3, half of the participants stated that they had previous self- and editing experiences. The other half reported that they did not have previous experience of editing. For question 4, all participants except 4 had previously used electronic programs to edit. To conclude, most participants had a positive attitude towards writing with good achievement; and most of them had previous experiences in manual and electronic editing.

The responses to questions in part B were more elaborate due to the nature of questions. Responding to question 1 about the nature of help they got from the software, participants gave several answers. The thematic content analysis revealed simi-

larity in responses in the following respects: sixteen responses out of the total number of participants in the experimental groups identified the detection and correction of grammar and spelling mistakes. Seven responses praised speed and easy use. In addition, one found it helpful in organizing thought. Answers to question 2 confirmed some of the features mentioned above. Most of the participants expressed their admiration for the speed and easiness of using the program in detecting grammar and spelling mistakes. Two participants added the features of providing error explanation. One participant admired the merit of suggesting synonyms. One participant liked the feature of detecting plagiarism. Moreover, one participant observed the features of correcting errors related to style and punctuation. As for responses to question 3, nineteen participants found no difficulty using the program. Three participants found difficulty logging in and downloading corrected files. One participant observed that the program did not provide correction for ideas. To question 4, all participants except two opted the option "enlightening" as a description for their electronic editing experiences. The two who chose "challenging" as a description for the experience complained about leaving the final decision about applying the corrections suggested by the program to them. For question 5, only two remarks were given. One stated that the program is "more than a grammar checker". The other left one advice for "everyone to use it", i.e. the program. In conclusion, students found many useful features in the program and very few disadvantages.

Connections between responses to questions in part A and those given for part B questions were evaluated. The question posed was: "Would students' poor attitude, poor achievement of writing, and lack of previous manual and/or electronic editing experiences negatively affect students' perception and evaluation of the electronic editing experience?". The responses to this question did not show a negative impact of the four factors on the electronic editing experiences in students' responses. On the contrary, students who had not previously edited or used technology for editing asserted that they would edit and would use this particular program for editing. One of the respondents wrote that "making words and sentences clearer, more precise, and as effective as possible".

Responses to the retrospective note question reflected learners' praise of the experience and the program in different aspects and varied ways. They commented that it helped them identify errors they had not realized they may commit. In addition to the features mentioned in response to question 2 in part B, they added that they learned how to learn from their mistakes as well as from those of others. They asserted that they would use the program to edit and would recommend it to their friends. In conclusion, participants' retrospective notes confirmed that students' overall experiences of electronic peer and self-editing were positive.

The analysis of the responses to the questionnaire and retrospective notes glossed the findings reached in the quantitative analysis. The overall positive evaluation of experience was a reflux for the high achievement got by the participants.

To conclude both the quantitative and the qualitative analysis, the present study found AWE a useful technique for improving self- and peer editing. Students were satisfied with both the editing experience and the software used for editing. Although, results are not generalizable due to the small sample size, they are significant, reliable, and valid.

5 Discussion

This study examines the effects of automated writing evaluation software on writing cause-effect essays as well as the participants' perceptions and evaluation of automated peer and self-editing.

The first and second research questions are about whether there is a significant difference in the quality of writing between the control and the experimental peer and self-editing groups. The results reveal much more improvements in the written products of the automated editing groups. Both the peer and self-editing experimental groups outperform the control groups who have not used AWE software. The cause-effect essays produced by experimental groups reflect higher quality and sophistication. This may be due to the "effective feedback" provided by the software, i.e., tool affordance. The program provides learners with varieties of feedback accompanied by metalinguistic explanations and examples. This reinforces the detailed value of AWE on writing in the literature review. The results also support the studies in which a comparison between automated versus non-automated editing like that of Li and Hegelheimer (2013), Wang (2013), Al-Wasy and Mahdi (2016), Ebadi and Rahimi (2017), Parra and Calero (2019), Law and Baer (2020), and Saricaoglu and Bilki (2021). Like these studies, the results of this study confirm that there is a significant difference in favor of electronic editing.

The third question is about whether there is a significant difference between the mean scores of the pre-test and post-test for self-editing and peer editing experimental groups, taken separately. The results revealed statistically significant differences between mean scores of the pre-test and post-test, in favor of the post-test for both groups. These results can be attributed to the help of the software. These results are supported by different studies such as Wang and Wang's (2012), Wang et al's (2013), Li and Hegelheimer's (2013), and Tavşanlı and Kara's (2021). These studies found improvement in the post-test scores for both the automated editing.

The fourth research question is about whether there is a significant difference between the impacts of AWE on writings of peer and self-editing experimental groups. The results yield no significant difference between the experimental groups. This finding is in tune with the Elola and Oskoz's (2010) study in terms of overall improvement and with Tavşanlı and Kara's (2021) study in terms of spelling and grammar improvements.

To conclude, the present study confirms results reached by previous studies of overall improvement of writing after the use of automated editing. It also confirms the advantages of automated editing over non-automated editing in terms of its impact on several editing aspects such as mechanics, content, and coherence observed in previous studies. Concerning its comparative nature, the present study finds no significant difference in the writing performances of automated peer editing versus automated self-editing ones. The study is unique because it makes several comparisons from different perspectives. First, it compares automated and non-automated editing. Then, it compares automated self-editing with automated peer editing. The researchers acknowledge that results reached in this study are not generalizable due to the small sample size. However, measures for the repeatability of sample selection and

research procedures are strictly followed to make its repeatability and replication possible in other contexts.

The fifth research question is about participants' automated editing experiences and how these participants perceive and evaluate the software. The findings from responses to the questionnaire and the retrospective note question reveal students' positive attitudes and overall high estimate of the automated peer and self-editing experience. These findings are in tune with findings of Parra and Calero (2019) and Hoang (2019) whose students expressed positive attitudes towards automated editing. In addition, the responses in the questionnaire and the retrospective notes confirm the positive impact of automated editing on the development of metalinguistic awareness. Many of the students' comments reflect the emergence of this awareness. For example, one self-editor describes the experience: "it opened my eyes to some mistakes that I never thought I have committed, and it expanded my knowledge about some grammar and punctuation rules that I didn't know about." Another self-editor reckons "I learned from and corrected my mistakes as I began to focus on placing commas and periods." A peer editor states: "it…. makes me notice my mistakes and correct." Another peer editor writes: "it [helps] me know the rules of writing and to avoid common writing mistakes." This confirms that automated editing improves students' ability to notice the gap in the student's linguistic system and leads to skill acquisition. Therefore, the present study is consistent with the Noticing Hypothesis by Schmidt (1990). He stated that real learning "uptake" was what learners "consciously notice". Therefore, the researchers find that the repeated use of the word "notice" in participants' responses signals the increase in students' awareness of the gap in their linguistic system and the necessity to work on improving it. Attention is considered to be a sufficient condition to encode the stimulus into long-term memory (Schmidt 1993). This entails that automated editing experience has a strong learning impact on the acquisition of the sub-skills of writing. The researchers find that AWE leads to awareness as well as consciousness of learners of their language-learning uptake.

The responses to the questionnaire and retrospective note question can also serve as consumer evaluation of the utility of the program for those who intend to use it for automated editing research and instruction.

To conclude, the present study found AWE beneficial for improving self- and peer editing. Students had positive evaluation of both the editing experience and automated feedback. Although, results are not generalizable due to the small sample size, they are significant, reliable, and valid.

6 Conclusions, limitations, and suggestions for further studies

The present study was designed to determine the effect of AWE software on both peer and self-editing. It also aimed to explore whether there was a significant difference between the effect of the AWE on peer editing on one hand, and the effect of the AWE on self-editing on the other hand. The results of this investigation showed that the AWE software had a positive effect on the writing quality of the research subjects. The second major finding was that the effect of AWE on peer editing was

quite similar to its effect on self-editing. There was no significant difference between the post-test scores of the experimental peer editing group and the experimental self-editing group. It was also shown that students had a positive attitude towards using AWE software in the process of editing. They insisted on the great help presented by such programs and their effective role in improving the final product of the essay.

6.1 Pedagogical implications

Taken together, these results suggest that writing courses can be provided with AWE software programs. The syllabus designers are advised to design some activities which demand learners to use such useful programs. Besides, course instructors have to encourage their students to use these useful online writing checkers. They can allocate part of the lecture time to explain the features of these checkers, assign certain marks to online editing activities, or even suggest some of these checkers to be used by their students.

6.2 Limitations and suggestions for future studies

Several limitations of the current study need to be acknowledged. The most important limitation lies in the fact that the present study does not specifically consider the two variables of gender or age of research subjects. Another limitation of this study is that the investigation is limited to the effect of the AWE on the quality of writing, in general; the effect of AWE on each aspect of writing was not included in the analysis. The AWE, for example, may have more effect on language use than organization. Thirdly, the present study is limited to the use of one AWE tool. A comparison of the effect of more AWE tools on learners' writing performance needs to be conducted.

Further work needs to be done to investigate the effect of AWE on the various aspects of writing. Another possible area of future research will be to compare the effect of different AWE tools on peer and self-editing. Besides, the issue of teachers' attitudes towards the use of AWE in their writing classes can be usefully explored in further research.

7 Appendix 1: A Questionnaire for Peer Editors

7.1 Participant's name

Please, answer the following questions after you complete editing your classmate's essay. Feel free to add any further remarks you would like to add at the back of the paper.

1. Do you like writing?
2. What is your average grade in writing courses?
3. Do you edit your own or others' writing? If yes, how often?
4. Have you ever used a computer or mobile program to help you edit writing?

5. How did the program help you edit your classmate's writing?
6. What are the features in the program that have been most helpful to you?
7. What is the most difficult aspect of using the program?
8. How did you find the experience of editing another person's essay using the program: challenging or enlightening? Explain why you choose one rather than the other.
9. If you have any further remarks, feel free to write them down here.

7.2 Retrospective Notes for Peer Editors

Participant's name:
Describe your experience while editing your classmate's essay. Write whatever comes to your mind while editing.

8 Appendix 2: A Questionnaire for Self-Editors

8.1 Participant's name:

Please, answer the following questions after you complete editing your own essay. Feel free to add any further remarks you would like to add at the back of the paper.
1. Do you like writing?
2. What is your average grade in writing courses?
3. Do you edit your own or others' writing? If yes, how often?
4. Have you ever used a computer or mobile program to help you edit writing?
5. How did the program help you edit your writing?
6. What are the features in the program that have been most helpful to you?
7. What is the most difficult aspect of using the program?
8. How did you find the experience of editing your essay using the program: challenging or enlightening? Explain why you choose one rather than the other.
9. If you have any further remarks, feel free to write them down here.

8.2 Retrospective Notes for Self-editors

Participant's name:
Describe your experience while editing your own essay. Write whatever comes to your mind while editing.

9 Appendix 3: The Data of the four groups

No	Self-editing control. Score Pre-test	Post-Test	No	Peer editing control. Score Pre-test	Post-Test
1	65	86	1	64	67
2	56	71	2	54	58
3	65	69	3	62	67
4	57	67	4	57	62
5	58	68	5	62	64
6	55	65	6	55	58
7	62	67	7	56	57
8	64	86	8	65	69
9	60	69	N0	Peer editing exp. Score	

No	Self-editing exp. Score Pre-test	Post-Test		Pre-test	Post-Test
			1	65	75
1	56	73	2	64	92
2	64	79	3	63	81
3	65	84	4	65	91
4	64	80	5	64	87
5	62	72	6	59	86
6	57	73	7	57	77
7	63	80	8	63	85
8	65	86	9	58	83
9	62	92	10	62	84
10	65	74	11	55	65
11	61	87	12	59	74
12	63	86	13	60	82
13	61	81	14	62	88

Data availability The datasets used during the current study are available from the corresponding author on reasonable request.

References

Abadikhah, S., & Yasami, F. (2014). Comparison of the effects of peer versus self-editing on linguistic accuracy of Iranian EFL students. *The Southeast Asian Journal of English Language Studies*, *20*(3), 113–124.

Al-Wasy, B. Q., & Mahdi, H. S. (2016). The effect of mobile phone applications on improving EFL learners' self-editing. *Journal of Education and Human Development*, *5*(3), 149–157.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Chen, C. F. E., & Cheng, W. Y. E. C. (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.

Chen, Z., Chen, W., Jia, J., & Le, H. (2022). Exploring AWE-supported writing process: An activity theory perspective. *Language Learning & Technology*, 26(2), 129–148. https://doi.org/10125/73482

Daweli, T. W. (2018). Engaging Saudi EFL students in online peer review in a Saudi university context. *Arab World English Journal*, *9*(4), 270–280. https://doi.org/10.24093/awej/vol9no4.20.

Diab, N. M. (2010). Peer editing versus self-editing in the ESL classroom. *System*, *38*(1), 85–95. DOI:https://doi.org/10.1016/j.system.2009.12.008.

DePoel, V., Carstens, K., & John, L. (2012). *Text editing: A handbook for students and practitioners*. Bruxelles, BEL: UPA, ProQuest ebrary.

Ebadi, S., & Rahimi, M. (2017). Exploring the impact of online peer editing using Google Docs on EFL learners' academic writing skills: a mixed methods study. *Computer Assisted Language Learning*, *30*(8), 787–815. https://doi.org/10.1080/09588221.2017.1363056.

Ellis, P. D. (2010). *The essential guide to effect size: Statistical power, meta-analysis, and the interpretation of research results*. UK: Cambridge University Press.

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning and Technology*, *14*(3), 51–71.

Feng, H. H., & Chukharev-Hudilainen, E. (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology*, 26(2), 58– 77. https://doi.org/10125/73479

Ferris, D. (2003). *Response to students writing: implications for second-language students*. USA: Lawrence Erlbaum Associates, Inc.

Field, A. (2013). *Discovering statistics using SPSS*. London, UK: Sage.

Hemati, M. (2012). The effect of teacher, peer, and self-editing on improving grammatical accuracy in EFL learners' writing. *Diversité et Identité Culturelle en Europe*, http://www.diacronia.ro/ro/indexing/details/A4026/pdf

Hoang, T. L. G. (2019). *Examining automated corrective feedback in EFL writing classrooms: A case study of criterion* (Doctoral dissertation), University of Melbourne.

Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, *73*(1), 82–88. https://doi.org/10.1093/elt/ccy044.

Hojeij, Z., & Hurley, Z. (2017). The triple flip: Using technology for peer and self-editing of writing. *International Journal for the Scholarship of Teaching and Learning:*, *11*(1), https://doi.org/10.20429/ijsotl.2017.110104.

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.

Khaki, M., & Biria, R. (2016). Effects of self- and peer editing on Iranian TEFL postgraduate students' L2 writing. *Journal of Applied Linguistics and Language Research*, *3*(1), 155–166.

Kiger, & Lara, V. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131, Medical Teacher, https://doi.org/10.1080/0142159X.2020.1755030

Klimova, B. F. (2011). Evaluating writing in English as a second language. *Procedia - Social and Behavioral Sciences*, *28*, 390–394. https://doi.org/10.1016/j.sbspro.2011.11.074.

Lai, Y. H. (2010). Which do students prefer to evaluate their essays: Peers or computer program? *British Journal of Educational Technology*, *41*(3), 432–454. https://doi.org/10.1111/j.1467-8535.2009.00959.x.

Law, S., & Baer, A. (2020). Using technology and structured peer reviews to enhance students' writing. *Active Learning in Higher Education*, *21*(1), 23–38. https://journals.sagepub.com/home/alh.

Lamy, M., & Hampel, R. (2007). *Online communication in language learning and teaching*. England: Macmillan Distribution Ltd.

Li, R. (2021). Modeling the continuance intention to use automated writing evaluation among Chinese EFL learners. *SAGE Open*, 1–13. Available at: https://doi.org/10.1177/21582440211060782

Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78–105. https://doi.org/1012

Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: effects on self-editing in L2 writing. *Language Learning and Technology*, *17*(3), 135–156.

Link, S., Mehrzad, M., & Rahimi, M. (2020). 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.

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *AISHE-J*, *3*, 33501–33514.

Mahendran, R. (2012). Enhancing ESL learners' writing skills. *Language in India*, *12*(3), 206–211.

Mohsen, M. A., & Alshahrani, A. (2019). The effectiveness of using a hybrid mode of automated writing evaluation system on EFL students' writing. *Teaching English with Technology*, *19*(1), 118–131.

Parra, G. L., & Calero, S. X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, *12*(2), 209–226. https://doi.org/10.29333/iji.2019.12214a.

Prabasiwi, E. A., & Warsono (2017). Employing self and peer editing techniques to teach writing recount texts for students with high and low motivation. *English Education Journal*, *7*(3), 220–226.

Saricaoglu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL*, *33*(3), 265–277. https://doi.org/10.1017/S0958344021000021.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. https://doi.org/10.1093/applin/11.2.129.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51–65. DOI:https://doi.org/10.1016/j.asw.2013.11.007.

Tavşanlı, Ö. F., & Kara, Ü. E. (2021). The effect of a peer and self-assessment-based editorial study on students' ability to follow spelling rules and use punctuation marks correctly. *Participatory Educational Research (PER)*, *8*(3), 268–284. https://doi.org/10.17275/per.21.65.8.3.

Thouësny, S., & Bradley, L. (Eds.). (2011). *Second language teaching and learning with technology: Views of emergent researchers*. Research-publishing. net.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*(2), 149–169. https://doi.org/10.1080/713611428.

Wang, F., & Wang, S. (2012). A comparative study on the influence of automated evaluation system and teacher grading on students' English writing. *Procedia Engineering*, *29*, 993–997. https://doi.org/10.1016/j.proeng.2012.01.077.

Wang, P. (2013). Can automated writing evaluation programs help students improve their English writing? *International Journal of Applied Linguistics & English Literature*, *2*(1), 6–12. doi:https://doi.org/10.7575/ijalel.v.2n.1p.6.

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.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa.

Weigle, S. C. (2013). English as a Second Language Writing and Automated Essay Evaluation: Routledge. https://www.routledgehandbooks.com/doi/https://doi.org/10.4324/9780203122761.ch3

Winarto, A., E (2018). Peer and self-editing strategies to improve students' writing skills. *JEELS*, *5*(1), 49–71.

Yang, Y. F., & Meng, W. T. (2013). The effects of online feedback training on students' text revision. *Language Learning and Technology*, *17*(2), 220–238. http://llt.msu.edu/issues/june2013/yangmeng.pdf.