Research Article

The Impact of Automated Writing Evaluation on Iranian EFL Learners' Essay Writing: A Mixed-Methods Study

Reza Bagheri Nevisi[1*] and Roya Mohammadi Yeganeh[1]

[1]University of Qom, Department of English Language and Literature, Faculty of Humanities, PO Box No. 37185-396, Qom, Iran

[*]Corresponding author: ✉ re.baghery@gmail.com ⓘ 0000-0002-8582-8119

Abstract

While writing skill is extensively studied in EFL contexts, more in-depth research is needed to explore how technology can assist its pedagogy. The present study aimed to investigate the impact of using an automated writing evaluation on Iranian EFL learners' essay writing. Learning how to reduce errors (in an EFL context) by being corrected at the moment and being exposed to different examples regarding that error in the learners' new texts through automated writing evaluation (AWE) tends to be the significance of this study. To this end, 50 Iranian EFL learners who were studying at the University of Qom, were randomly chosen. The sample included 25 females and 25 males, whose ages ranged from 19 to 25. The participants were given a pre-test before using AWE software. They were given a topic to write about as a pre-test. After the treatment, an IELTS Task 2 was utilized as a posttest. The IELTS writing band descriptors were used to evaluate the writings. The ANCOVA results showed a remarkable improvement in the essay writing of the EFL learners using an AWE software (i.e., Grammarly). The analysis of interview data revealed that the learners were more enthusiastic about using the AWE feedback because they were corrected while they were writing their essays. Since AWE is discovered to be a helpful device to promote learners' writing skills, students would also be inspired to become associated with such online learning environments and utilize them earnestly and productively. This research also discovered the learners who got feedback from the AWE device got more prosperous but they also started to ask their teacher to provide more feedback to have AWE feedback and traditional feedback combined. The findings have implications for language teachers, material developers, and curriculum designers.

Keywords: automated writing evaluation, essay writing, mixed-methods study, process writing

1. Introduction

In the past decade, a highly increasing Interest in the field of L2 writing has been noticed by automated writing evaluation (AWE) (e.g., Al-Inbari & Al-Wasy, 2023; Fan, 2023). Disputably, the most propitious point of contact between the areas of AWE and L2 writing tends to be automated feedback (Shermis & Burstein, 2003; Warschauer & Ware, 2006). This feedback tends to play a crucial role in learners' writing ability as it is cost-effective, practical, and helpful. Receiving the feedback, the learners would be aware of their errors immediately, and they'd be given the information they require on the spot, which can exert a good impact on the writing process of language, and some perceive it as a menace.

Those who support AWE use in the classroom dispute this point that the tremendous advantages of AWE are their ability to assess and respond to student writing as well as humans do (Attali & Burstein, 2006) and doing so in a much more time- and cost-effective way. Hypothetically, AWE can motivate and guide student revision and the learners' autonomy would be enhanced too (Chen & Cheng, 2008). It is meant to support process writing approaches that the value of multiple drafting is emphasized through scaffolding suggestions and explanations. The assimilation of AWE into the curriculum is highly believed to be consistent with the drive toward individualized assessment and instruction (Burstein & Marcu, 2003).

As mentioned previously, there have been various arguments over the beneficial or deleterious impact of the feedback given by AWE software on learners' writing ability. In fact, the experiential evidence on corrective feedback tends to be contradictory and away from being definite, and they frequently challenged Truscott's claims against corrective feedback (Ferris, 2004; Hyland & Hyland, 2006). Following the mentioned points, Ferris (2004) stated that "positive impacts are predicted by existing research for written error correction" (p. 50).

Correcting learners' essays and giving feedbacks on them, by using applications which fulfill this task automatically, provide learners with the correct form of their errors at the moment that can exert a good impact on their writing ability. This automated system is known as AWE (Automated Writing Evaluation). Despite the recent development of AWE technology and the increasing interest in utilizing this technology in language classrooms, the effects of using AWE on reducing grammatical errors in L2 writing have been considered by a few studies (e.g., Liao, 2016).

As a matter of fact, nowadays most language learners have been inundated with technology, so teachers can utilize this golden opportunity to

assist them develop their writing skill. In other words, since they are facing lots of time-consuming and abundant commenting on student drafts and inspired by assurance of computerized writing assessment, AWE is considered as a silver bullet for language and literacy development (Warschauer & Ware, 2006).

2. Literature Review
2.1. Automated Writing Evaluation

Various researchers have been dealing with improvement of language programs that what they do is to deal with grading and presenting feedback on the writing skills. Burstein and Marco (2003) maintained that writing tends to be a particular language capability which possibly tends to be best developed by writing constantly, suitable and constant feedback. Due to new technological inventions in this field as the AWE computer program, these checking processes have been automated which is believed to be a supporter to teachers and a tool that freedom and planning time to students in increasing their level of motivation is provided by that. (Shim et al., 2013).

The utilization of AWE has been rising as a teacher assistant tool which high-level of feedback and writing quality is provided by it. This computer program is believed to assist learners` writing quality be improved because of its fast and individualized feedback that is accompanied by explanations of grammar, spelling, sentence and word usage which learners 'autonomy can be contribute to (Wang, et al., 2013). It would also provide the learners with the amount of clarity, coherence and cohesion their text possess.

The advantages of this tool to improve writing is confirmed by several studies that the use of AWE is promoted in three extents. Initially, word processing facilitates the aspects of editing and revising grammar and spelling which contributes to learners 'awareness in their writing (Wang et al., 2013). Second, through an error correction program the students are given the opportunity to recognize their errors immediately and teachers have the chance of interacting with their learners in specific error correction and feedback (Shim et al., 2013). Moreover, computerized feedback concentrates the learners' attention on sentence-level error which encourage them to enhance inaccurate usage and their capability to identify and reformulate errors when no human support is available which can encourage autonomous learning (Wang et al., 2013). Third, artificial technology systems claim to be more objective and accurate when grading standardized essay tests, as human markers in the typical test score diverge by some points needing a third marker to have a final grade agreement (Warschauer & Grimes, 2008). In addition, the

feedback produced by the people tends to be flexible and restricted according to student´s background and needs. However, the AWE is able to check large number of essays and score them immediately and accurately due to the Latent Sematic Analysis technique.

2.2. Relevant Empirical Studies

It has been said that former studies have investigated English writing development due to studying in the ESL context if it is long-term development (e.g., Knoch et al., 2015) or short-term development (e.g., Storch, 2009). A thorough inspection of AWE and its related research studies was presented by Warschauer and Ware (2006) last decade and it might be disputed that their wide categorization of various types of AWE research still holds true, with some studies which the validity of AWE and comparison of the machine scoring with the human scoring is concerned by it (e.g., Deane, 2013); others on the use of AWE in developing students' standardized writing test scores (Attali, 2004; Tang & Rich, 2017). However, the factor which is considered as a significant point is for more process-product research on the utility of AWE to disclose the process of AWE particular application and how it affects instructing the writing (Warschauer & Ware, 2006).

Actually, after their call for classroom research on AWE (Warschauer & Ware, 2006), the last ten years observed an ascending body of studies, which were published in international peer-reviewed journals investigating the use of AWE in the classroom (e.g., Chen & Cheng, 2008; Grimes & Warschauer, 2010; Li et al., 2014; Li et al., 2015; Wang et al., 2013; Warschauer & Grimes, 2008) and even a particular Calico issue on AWE released in 2016 (cf. Li et al., 2016), whose findings appeared to support Grimes and Warschauer's recommendation of AWE's "utility in a fallible tool" if deployed effectively (Grimes & Warschauer, 2010, p. 4).

Chen and Cheng (2008) examined the utility of an AWE program with three similar classes of three teachers during one term. It may be discussed that the most significant contribution of their research to the field tends to be their thoughts and understanding of the AWE usage in the revising process of writing instruction, pursued by teacher and peer feedback in the later process. In addition, it is them who initially suggested the potential usefulness for constructing a minimum score demand as a prerequisite for submission to AWE. For instance, AWE score and feedback were used as a reference in her scoring by one teacher in their study, who wanted her students to revise their essay in the system till they had attained a minimum score of 4 out of 6 before they handed it in to teacher assessment and peer review.

Warschauer and Grimes (2008)'s mixed-methods exploratory case study of four schools in their use of two AWE programs divulged that although the program encouraged students to revise more, the revision was limited to language forms only, few on content or organization. In addition, teachers' use of AWE varied from school to school and was ascertained most by teachers' prior beliefs about writing pedagogy, which debatably called for the inevitability of teacher training on writing pedagogy if AWE was to be successfully applied in the classroom.

Grimes and Warschauer (2010) did a 3-year longitudinal study about using AWE in eight schools in California and came to the conclusion that AWE motivated students to write and revise more and promoted learner autonomy. The successful use of AWE was attributed fairly to the maturity of the AWE programs in the study, but more crucially to the local social factors such as technical, administrative and teacher support, which seemed but to confirm the assertion that the key to technology use might be neither hardware nor software, rather human.

In the EFL context, Wang et al. (2013) the effect and role of applying AWE was probed on freshmen writing with a group of 57 students from a university. The quasi-experimental pretest/posttest research design was applied and a vivid difference was displayed by outcomes between the experimental group and the control group in writing accuracy, that the experimental group showed a clear writing attainment regarding writing accuracy and learner autonomy perception. In discussing the pedagogical implications, they suggested that teachers be involved more energetically in teaching models of writing to students so that students know how their language accuracy can be developed and how their writing content and structure can be improved. In examining the impact accuracy with 70 nonnatives.

Li et al. (2015) discovered that corrective feedback had enhanced the number of revisions and improve writing accuracy. Their study seemed to support the claim of the utility of the practice suggested by Chen and Cheng (2008) of requiring a minimum score before submission to AWE. Additionally, similar to the previous studies (e.g., Grimes & Warschauer, 2010; Warschuaer & Grimes, 2008; Wang, et al., 2013), their study strengthened the significant role of teachers, and it was suggested that the instructor's ways of implementing AWE might impact how students involved themselves in revising in AWE. Al-Inbari and Al-Wasy (2023) conducted a mixed-methods study to examine the impact of an AWE program on the peer and self-editing of cause-and-effect essays. The results of qualitative and quantitative data analysis revealed that students who used the AWE tool

thought that the AWE feedback was very helpful with their editing had improved significantly. Fan (2023) investigated how AWE feedback through Grammarly affected EFL students' writing using a mixed-methods design. The results revealed that there were not any significant differences between the experimental and control groups. Moreover, the analysis of the qualitative data (fixed-response and open-ended questionnaire data) supported the quantitative results.

Although the significance of teacher pedagogical roles has been implied or proposed in some of the studies (e.g., Li et al., 2015; Wang et al., 2013; Warschauer & Grimes, 2008), no systematic training was presented to teachers regarding the writing pedagogy in those studies reviewed. Moreover, none of these AWE studies seemed to suggest a conjectural procedure of utilizing AWE efficiently in the classroom so far, in most cases, the ways of utilizing AWE merely depended on teachers (e.g., Link et al., 2014). To achieve the above-mentioned objectives, the following research questions of the present study were formulated:

RQ1: Does the learners' writing quality change using AWE process writing program?

RQ2: Do the learners find it helpful to be evaluated by AWE software and does it help them in writing new essays?

3. Method

3.1. Participants

Due to the difficulties of randomization, available sampling was used. The sample of the study, who were at an intermediate level of English proficiency, included 50 Iranian EFL learners at the University of Qom. They included 25 females and 25 males, and their ages ranged from 19 to 25. They were assigned to an experimental group and a control group. The experimental group used an AWE program to receive feedback on their writing.

3.2. Instrumentation

3.2.1. Oxford Placement Test

The Oxford Online Placement Test was used to determine the level of the participants before applying the AWE approach. It helps to place students into the appropriate level class for a language course. The Oxford Placement Test is computer-adaptive, which means that the test adjusts the difficulty of questions based on the student's responses. This makes it more motivating and ensures that it gives a more precise measurement than traditional placement

tests. Answers are automatically marked after each task, giving you an instant result once the test is complete. It is also used as a quick measure of a student's general language ability. This test is different from most other placement tests. The Oxford Placement Test has two sections: Use of English and Listening. The Use of English section assesses students' knowledge of grammatical form and vocabulary. The Listening section assesses students' general listening ability. Both sections test how well students understand the meaning of what is being communicated, which is an excellent indicator of general language ability.

3.2.2. Pretest and Posttest

The participants were given a topic from Cambridge IELTS tasks 2 to write about as a pre-test to evaluate their writing. After the administration of the treatment, the participants' improvement was checked through another topic from Cambridge IELTS tasks 2 as a post-test.

3.2.3. Automated Writing Evaluation Tool

The software "Grammarly" was used to see the impacts of it on learners' writing development. Grammarly is a popular software available as a browser add-on for Google Chrome, Firefox, and even Microsoft Edge. It checks for grammar and spelling errors as you write something. It will show an indicator at the bottom right corner of the writing area. Clicking on the indicator will show you the number of errors. Spelling, grammar, and contextual errors will be highlighted with an underline as you write. This software assessed the following: Correctness, clarity, engagement, and delivery in the participants' writings. In fact, the data were gathered and the participants' writing (which was accomplished using the software Grammarly) was checked and evaluated.

3.3.4. Open-ended Questionnaire

Cambridge IELTS writing band descriptors are the criteria based on which the IELTS test takers' writing tasks are evaluated and assessed. Each criterion was awarded a band score from 0 to 9. The criteria are weighted equally and the overall band score is the average of the four component scores, rounded to the nearest whole or half band.

3.3.5. Open-ended Questionnaire

The participants were asked some questions about the effectiveness of the AWE software. In this regard, an interview was conducted which included

open-ended questions. The validity and reliability of the questions were examined by an expert. The questions of the interview were as follows:

1. Were you satisfied with Grammarly feedback?
2. What kind of feedback was helpful the most to you?
3. How did you use the feedback in terms of revising?
4. What kind of strategies did you use to achieve your best score?
5. Was it easy to correct the highlighted errors in Grammarly based on the feedback?
6. Are you confident in using Grammarly?
7. Was it easier to find/identify errors by yourself after using Grammarly?
8. What kind of errors do you usually make in writing?
9. Can you identify your writing weakness from the feedback in Grammarly?

3.4. Procedure

To conduct this study, two groups of participants were utilized which include an experimental group and a control group. The participants were given a pre-test before using AWE software. In fact, they were administered the Oxford Placement Test to determine their English proficiency levels. So, all the participants, whether in the experimental group or the control group, took the test to determine their English level. Thus, after taking this test, the participants' proficiency level (i.e., intermediate) was revealed. Then, the participants were given a topic to write about as a pre-test. Cambridge IELTS writing band descriptors were used to evaluate and assess the IELTS test takers' writing tasks.

Afterward, the control group received feedback from their teacher. However, the AWE software was utilized to monitor its impact on experimental group's writing. The software named Grammarly was applied to check learners' writing. This software includes some items: correctness, clarity, engagement, and delivery. In the correctness part, the software will check and improve spelling, grammar, and punctuation. *Clarity* helps the learners make their writing easier to understand, which can play a crucial role in having clear writing with high coherence. "Engagement" makes the writing more interesting and effective, and *Delivery* helps to make the right impression on the reader. These options, available in this software, will help the learners improve their writing ability and since they'll see their mistakes being corrected right away, with suitable explanations and examples given by the software, they'll probably have better writings later, and this improvement was checked through the post-test.

Therefore, the data were gathered and the participants' writings (which was accomplished, using the software Grammarly) were checked and

evaluated. After this process, a qualitative open-ended questionnaire was given to those participants to know their ideas upon utilizing this AWE software and the impact of it on their writing quality. In fact, the interview questions were made into a Google document and distributed among the participants in the experimental group, who received feedback through Grammarly.

3.5. Data Analysis

To analyze the data, ANCOVA was applied to explore the impacts of automated writing evaluation-assisted process approach on Iranian EFL learners' essay writing, to probe the quantitative research question. Finally, the qualitative data were analyzed using the procedure, suggested by Dörnyei (2007).

4. Results

The purpose behind the present study was to investigate the effect of the AWE process writing program on the improvement of the writing quality of Iranian EFL learners. The statistical analysis of one-way ANCOVA was employed to analyze the data collected through this study. Before discussing the results, it should be noted that the assumption of normality was retained. As displayed in Table 4.1, the ratios of skewness and kurtosis over their respective standard errors were lower than +/- 1.96; hence, normality of the data. It should be noted that the ratios of skewness and kurtosis over their respective standard errors are analogous to standardized scores (Z-scores) which can be compared against the critical values of +/- 1.96 at .05 levels.

Table 1
Descriptive Statistics; Testing Normality of Data

Group		Skewness			Kurtosis		
		Statistic	Std. Error	Ratio	Statistic	Std. Error	Ratio
Experimental	Pretest	-.332	.409	-0.81	-1.040	.798	-1.30
	Posttest	-.390	.409	-0.95	-.940	.798	-1.18
Control	Pretest	-.136	.597	-0.23	-1.018	1.154	-0.88
	Posttest	-.143	.597	-0.24	-1.065	1.154	-0.92

4.1. Homogenizing Groups on Pretest of Writing Quality

An independent-samples t-test was run to compare the experimental and control groups' means on the pretest of writing quality in order to prove the two groups were homogenous in terms of their writing quality prior to the administration of the treatment. Table 2 displays the results of the descriptive statistics for the two groups on the pretest of writing quality.

Table 2
Descriptive Statistics for Writing Pretest by Groups

	Group	N	M	SD	Std. Error Mean
Pretest	Experimental	33	51.15	17.136	2.983
	Control	14	45.93	20.656	5.521

The results showed the experimental ($M = 51.15$, $SD = 17.13$) and control ($M = 45.93$, $SD = 20.65$) groups' means on the pretest of writing quality. Table 3 displays the results of the independent-samples t-test.

Before discussing the results, it should be noted that the assumption of homogeneity of variances was retained on the pretest of writing quality. As displayed in Table 2, the non-significant results of Levene's test indicated that the two groups were homogenous in terms of their variances on pretest of writing quality, $F = .403$, $p > .05$. The results of independent samples t-test, which represented a weak effect size, indicated that there was not any significant difference between the two groups' means on the pretest of writing quality, $t(45) = .899$, $p > .05$, $r = .133$. Thus, it can be concluded that the two groups were homogeneous in terms of their writing quality prior to the administration of the treatment.

Table 3
Independent-Samples t-test; Pretest of Writing Quality by Groups

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.403	.529	.899	45	.374	5.223	5.812	-6.483	16.929
Equal variances not assumed			.832	20.973	.415	5.223	6.275	-7.828	18.274

4.2. Intra-Rater Reliability Indices

Table 4 displays the results of the Pearson correlations computed to estimate the intra-rater reliability of the indices for the pretest and the posttest of writing quality.

Table 4
Pearson Correlations; Intra-Rater Reliability of Pretest and Posttest of Writing Quality

		Pre-Rater 2	Post-Rater 2
Pre-Rater 1	Pearson Correlation	.874**	
	Sig. (2-tailed)	.000	
	N	40	
Post-rater 1	Pearson Correlation		.926**
	Sig. (2-tailed)		.000
	N		40

**. Correlation is significant at the 0.01 level (2-tailed).

Based on these results it can be concluded that there were significant agreements between the rater's ratings on the pretest of writing quality, representing a large effect size, r (45) = .874, p = .000), and the posttest of writing quality, representing a large effect size, r (45) = .926, p = .000).

4.3. Exploring the First Research Question

A one-way analysis of covariance (one-way ANCOVA) was run to compare the experimental and control groups' means on the posttest of writing quality after controlling for the effect of their baseline writing ability as measured through the pretest. Besides the assumption of normality, which was discussed above, one-way ANCOVA has three more assumptions (i.e., homogeneity of variances of groups, homogeneity of regression slopes, and the linearity of relationship between the covariate (i.e., pretest of writing quality) and posttest (i.e., dependent variable). The results are shown below.

Table 5
Levene's Test of Homogeneity of Variances; Posttest of Writing Quality by Groups with Pretest

F	df1	df2	Sig.
3.725	1	45	.060

As shown in Table 5, one-way ANCOVA assumes homogeneity of variances of the groups. The non-significant results of Levene's test indicated that the assumption of homogeneity of variances was retained, $F(1, 45) = 3.72$, $p > .05$. The second assumption requires that the linear relationship between the pretest and the posttest of writing quality be roughly equal across the experimental and control groups (i.e., homogeneity of regression slopes). As shown in Table 6, the non-significant interaction between the covariate (i.e., pretest) and the independent variable, representing a weak effect size, indicated that the assumption of homogeneity of regression slopes was retained, $F(1, 43) = .044$, $p > .05$, partial $\eta = .001$.

Table 6

Test Homogeneity of Regression Slopes; Posttest of Writing Quality by Groups with Pretest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	68.308	1	68.308	5.246	.027	.109
Pretest	15319.883	1	15319.883	1176.515	.000	.965
Group * Pretest	.574	1	.574	.044	.835	.001
Error	559.921	43	13.021			
Total	173293.000	47				

One-way ANCOVA assumes that there is a linear relationship between the pretest of writing quality (i.e., covariate) and the posttest (i.e., dependent variable).

Table 7

Test of Linearity of Relationship between Pretest and Posttest of Writing Quality

			Sum of Squares	df	Mean Square	F	Sig.
Posttest * Pretest	Between Groups	(Combined)	18221.38	34	535.923	11.92	.000
		Linearity	17609.33	1	17609.33	391.9	.000
		Deviation from Linearity	612.047	33	18.547	.413	.978
		Within Groups	539.167	12	44.931		
		Total	18760.55	46			
		Eta Squared (η^2)	.971				

According to Table 7, the significant results of the linearity test, representing a large effect size, indicated that the relationship between pretest

and posttest of writing quality was a linear one, $F(1, 46) = 391.9$, $p < .05$, $\eta^2 = .971$.

Table 8 displays the descriptive statistics for the experimental and control groups on the posttest of writing quality after controlling for the effect of their writing quality ability as measured through the pretest.

Table 8
Descriptive Statistics; Posttest of Writing Quality by Groups with Pretest

| Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Experimental	59.670[a]	.623	58.415	60.926
Control	51.849[a]	.960	49.914	53.783

a. Covariates appearing in the model are evaluated at the following values: Pretest = 49.60.

The results indicated that the experimental group ($M = 59.67$, $SE = .623$), after working with the AWE process writing program, significantly outperformed the control group ($M = 51.84$, $SE = .960$) on posttest of writing quality after controlling for the effect of the pretest.

Table 9
Tests of Between-Subjects Effects; posttest of Writing Quality by Groups with Pretest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Pretest	16460.475	1	16460.475	1292.182	.000	.967
Group	590.720	1	590.720	46.373	.000	.513
Error	560.494	44	12.739			
Total	173293.000	47				

Table 9 displays the main results of one-way ANCOVA. The results, representing a large effect size, indicated that the experimental group significantly outperformed the control group on the posttest of writing quality after controlling for the effect of the pretest, $F(1, 44) = 46.37$, $p < .05$, partial $\eta^2 = .513$.

Figure 1
Means on Posttest of Writing Quality by Groups with Pretest
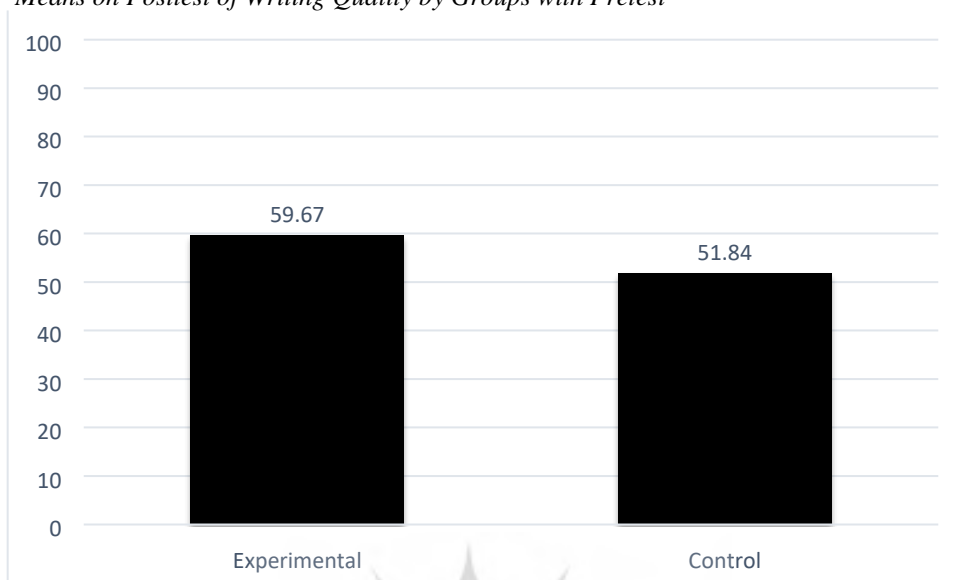
4.5. Exploring the Second Research Question

To answer the second research question (i.e., if the learners find it helpful to be evaluated by AWE or not, and if it helps them in writing new essays), it should be mentioned that learners who took part in the interviews, had different opinions upon the usefulness of this AWE process.

Analysis of the interviews with some students reveals the fact that AWE helped them with drafting their essays. For example, all the students, who used Grammarly, to varying degrees felt that it assisted them in structuring their essays. A less experienced student for example, commented that they felt Grammarly had provided a scaffold, which allowed them to write a better assignment answer. Students commented that Grammarly gave them confidence in their writing. For example, one of them said they were "not really confident about how to structure an essay, and that's where this Grammarly has assisted."

One of them commented that Grammarly gave confidence that they had covered the topic areas required to answer the assignment question and she felt a sense of being secure and accurate while writing the essay. Another participant commented "AWE helped me convey what I mean clearly, and it increased the coherence of my writings." AWE has also helped the participants make fewer grammatical mistakes while writing their essays as an interviewee said, "When I wrote something, Grammarly immediately showed the

feedback, and my mistakes were corrected on the spot, with showing me the reason. Therefore, it helped me be more accurate in writing later essays." A participant also commented that they wrote fewer drafts of their essay

> …. because it's (Grammarly) given me the feedback to be able to get straight to where I need to change, while before I didn't have that so I just relied on other people reading it and thinking I needed to change so it drastically lessened the number of drafts I did.

On the other hand, besides the usefulness of this application, some participants said that the AWE program cannot only replace teacher feedback since the students still need help from the teacher to enhance the content of their writing. For example, an interviewee put "The program is actually restricted to the semantic analysis of the language." Most of the participants made grammatical mistakes and they found it helpful to be corrected immediately by the application rather than themselves. For instance, one of them stated "It's much more helpful to be corrected by Grammarly and learn from your mistakes".

They also said that after using the AWE process, they could identify their weakness in writing through the feedback they received from Grammarly. In the same way, one of the interviewees commented, "Before using this program, I didn't know exactly where my weakness is, but now, I know it and I can start improving it and I think it can really help me." While almost two-thirds of interviewees were positive about Grammarly's feedback, it was observed by some of them that some errors can also be identified by Microsoft Word. Likewise, most of the students found Grammarly's grammar feedback useful and practical, while none considered it pointless. Meanwhile, some of the interviewees uttered their doubts about the accuracy of some grammar feedback, as mentioned below:

> *Excerpt 1*: Longer sentences would be marked as 'grammatical errors.' When this happens, I have to change a sentence into a simpler structure by cutting it shorter.
> *Excerpt 2*: When a main clause and a subordinate clause both have verbs, it's identified as verb error. It is misleading.

Additionally, the interviewees mentioned the kind of feedback they considered helpful and here are some of their opinions upon it. Before that, it's worth mentioning that some feedback and revisions are considered to be good and practical while others tend to be misleading and impractical. In other words, good revisions correctly recognize problems, elucidate ideas, or enhance expressions; neutral revisions neither improve nor worsen the well-

formed or ill-formed original text; 'bad revisions' generate errors or degrade the quality of the original text. Here are some opinions of the interviewees regarding some good revisions they've received and learned from:

> Well, actually, I felt it was useful for me because I myself really love grammatical feedback and I think they tend to be helpful in writing s.th because these kinds of feedback would also teach you s.th beside correcting you. That's why I personally believe that grammatical feedback is perfect. For example, I remember a sentence that I wrote which was "We noticed that the girl was disappeared." Then I received the grammatical feedback by Grammarly which was :>Verb error. Disappear cannot be used in passive voice.>Revision: We noticed that the girl disappeared. So here I could also learn s.th which was so good."

In addition to all the mentioned points, some of the interviewees talked about having stress while they were writing an assignment for their class. They said they would always feel anxious and worried about making mistakes in their essays. Here's one of their opinions:

> To tell the truth, I'm always worried about making grammatical mistakes or even punctuation mistakes in my writing but after using Grammarly, I felt completely confident while writing s.th, because I was corrected immediately, and it gave me confidence and motivation for writing other essays.

All in all, the participants found it totally helpful to receive feedback from Grammarly and being corrected right away, and they prefer to be corrected by this application since it's believed that AWE can gradually lead to remarkable progress in learners' writing particularly in the long run.

5. Discussion

The study aimed at investigating the effects of an AWE tool (i.e., Grammarly) on EFL students' writing development. Another purpose of the study was to examine if the learners found it helpful to be evaluated by AWE and if it could help them in writing new essays. In relation to mentioned objectives, the following research questions were asked to evaluate the effects of an AWE tool (Grammarly) on students' writing development:

1. Does the learners' writing quality change using AWE process writing program?
2. Do the learners find it helpful to be evaluated by AWE software and does it help them in writing new essays?

To address the above research questions, the pre-test and post-test scores were inspected, and some conclusions were drawn. The first conclusion that can be deduced from the study is that applying AWE software has transparently supportive effects on EFL learners writing development. With the aid of a comparison of the holistic scores gained from the pre- and post-tests, it tends to be vivid that receiving the AWE feedback would enhance writing development of university-level EFL students. The fact that students utilizing an AWE tool promoted their writing scores remarkably is steady with a number of research studies which were done on the same topic (Dikli, 2006; Hoon, 2006; Kern & Warschauer, 2000; Li et al, 2015; Warschauer & Ware, 2006; Wang et al., 2013, Al-Inbari & Al-Wasy, 2023).

Another point that can be discussed is that AWE makes learners capable of getting remarkably higher writing scores when compared to traditional pen-and-paper instruction though either method bring about development in writing. As mentioned by Zhang and Hyland (2018), various sources of formative assessment can possess a massive possibility in facilitating student involvement in writing assignments. However, when above-mentioned writing instruction and feedback methods were compared, as Wang et al. (2013) discovered, according to the total impact and the detection of students' thoughts toward their usage of the AWE software, it was seen that students who used AWE display noticeable writing improvement. Hence, it can be mentioned that AWE reveals to be more helpful than traditional pen-and-paper instruction and feedback in terms of university-level EFL students' writing performance by providing constant corrective feedback with vivid explanations.

The second issue to be discussed is that AWE and pen-and-paper method may have similar and different impacts on learners' writing inclination. To initiate, both groups can appear to be similar in terms of planning before writing, admiring being a good writer, enjoying literary analysis papers and research papers, and their tendency to get the highest score on a writing task. However, the students who get traditional feedback might be more extrinsically motivated compared to the students who get AWE feedback. It means that they might have become more dependent on an external effect while writing. Moreover, the control group seems to create more positive attitudes regarding perceived value of writing. This is perhaps since - unlike the students who get AWE feedback- they have to indicate a lot of struggles, preparation, and time on writing because of having restricted number of essay presentation during a restricted period of time and absence of software or peer aid, which might have ended up comprehending writing as a more significant skill which is worth a lot of struggles. Moreover, requiring a teacher to get writing feedback can be enhanced more in the experimental group when

compared to the control group merely since it's thought that there might be some points that the software may fail to recognize and give feedback on.

This finding tends to be consistent with the results of several previous studies, such as Fan (2023) and Lipnevich and Smith (2009), mentioned students are in favor of teacher feedback and comments rather than AWE feedback. However, when the two groups are compared, it can be said that AWE made some positive impacts on learner's intrinsic inclination towards writing. Enjoying writing which can involve creative writing tasks and without attention to being scored, being capable of expressing opinions, detecting writing nice essays and spelling easy, and being motivated to write in their classes were the items that indicated an essential difference between groups. Wilson and Czik (2016) and Liu et al. (2010) stated that online learning environments, for instance AWE, cause students to improve positive attitudes and abilities to share views, and get involved more, which provide greater levels of motivation.

6. Conclusions and Implications

This study was designed and carried out to examine the impact of using an AWE-assisted process writing on Iranian EFL learners' essay writing. Furthermore, this study attempted to investigate the effect of AWE on students' later essay writing. The result of this study was a confirmation of the power of automatic writing evaluation as the learners' key to success in writing more powerfully. The findings of the research indicated that although the students are used to teacher's feedback on their essays, they could feel a considerable improvement in their essays, including spelling, clarity, engagement and accuracy. The second concern of this study was to investigate the impact of AWE on learners' later essay writing. The result indicated a remarkable improvement in their essays and meanwhile they appeared to be more enthusiastic about writing various essays since they attained a high proficiency and grasp of writing points; therefore, some development can also be seen in their subsequent writing tasks.

In relation to English language teaching practice, the study has some implications for language teachers, material developers, and curriculum designers. Firstly, it can be recommended that AWE tends to be a highly helpful technique in teaching writing since it appears to enhance EFL learners' writing development. That's why language teachers could be informed that the significance of online learning areas and combining them to their teaching program for assisting students expedite learning and to boost their writing grades. Since AWE is discovered to be a helpful device to promote learners' writing skills, students would also be inspired to become

associated with such online learning environments and utilize them earnestly and productively. After all, because the recent research discovered the learners, who got feedback from the AWE device, got more prosperous and started to require their teacher more to receive feedback, it tends to be more logical to have both AWE feedback and traditional feedback combined in an effort to conduct the suitable effectiveness of feedback and to guarantee prosperous learning.

Secondly, motivation is highly related with being good language learners and considered as an essential component which has a considerable influence on foreign language learning achievement. Since AWE's individualized feedback meets university-level EFL students' specific needs, increases writing motivation, and encourages learners to take responsibility of their learning, language teachers had better pay attention to this fact and can carry out process-based writing instruction with the help of an AWE tool in order to increase students' writing motivation, autonomy, and self-efficacy. However, one important point that should be taken into consideration is that language teachers may need to supervise especially low students since they may have difficulty understanding the computerized feedback, which may influence them to cause a decrease in their motivation.

Another pedagogical implication for language teachers might be to make use of those rapid developments in technology to get efficiency. To make use of technological advancements, language teachers had better be digitally literate. Besides, teachers' technology literacy is necessary to effectively incorporate it in their learning and teaching processes and facilitate students' learning. Namely, utilizing AWE in EFL writing classes allows teachers to reduce the time spent on huge number of essays and therefore increase the number of writing assignments to provide student ability and self-efficacy.

When it comes to material developers and curriculum designers, they should integrate new teaching environments to writing curriculum. It is a widely accepted fact that technology helps English language learners get involved in the target culture and language more easily and find a voice. Thus, advances in technology can be fully and creatively used and integrated into writing curriculum to help students learn as much as possible. Language learners should also be able to embrace new developments and undertake digital learning activities at any place and time instead of being limited to learning in a traditional classroom in order to get the optimal efficiency in their language learning process.

This study was conducted with the participation of 50 university level students who were English language learners. Since it is deficient in generalizability to the population because of small sample size, a further study

can be conducted with a larger sample in order to reach more reliable results. Moreover, a study with participants from different proficiency levels or from various backgrounds can be fulfilled to know if similar results can be reached or not. Additionally, the recent study probed the impacts of AWE on learners' writing development holistically because of the laws of the university where the research was conducted. A further study can be implemented to consider the impacts on the writing field individually. To put it in another way, the way students improve in vocabulary utility, organization, coherence, content, grammar and spelling can be considered separately in order to gain more thorough outcomes and also to compare the post-test results with the software's scores for each domain of writing directly. Furthermore, for removing the researcher impact totally, a further study can be done with the same teacher teaching both groups during the experiment

References

Al-Inbari, F.A.Y., & Al-Wasy, B.Q.M. (2023). The impact of automated writing evaluation (AWE) on EFL learners' peer and self-editing. *Educ Inf Technol* (Dordr), *28*(6), 6645-6665. doi: 10.1007/s10639-022-11458-x.

Attali, Y. (2004). Exploring the feedback and revision features of criterion. *Journal of Second Language Writing*, *14*(3), 191-205.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3). https://ejournals.bc.edu/index.php/jtla/article/view/1650

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

Burstein, J., & Marcu, D. et al., (2003). Developing technology for automated evaluation of discourse structure in student essays. In M. D. Shermis & J. C. Burstein (Eds.), *Automated Essay Scoring: A Cross-disciplinary Perspective* (pp. 209-230). Hillsdale.

Chen, C.E., & Cheng, W.E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology*, *12*(2), 94-112.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7-24. doi:10.1016/j.asw.2012.10.002.

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, *22*, 1–17, doi:10.1016/j.asw.2014.03.006.

Dörnyei, Z. (2007). *Research methods in applied linguistics*: *Quantitative, qualitative, and mixed methodologies.* Oxford University Press.

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

Ferris, D. (2004). The "Grammar Correction" debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing*, *13*(1), 49-62. doi: 10.1016/j.jslw.2004.04.005.

Grimes, j., & Warschauer, M.et al. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. The Journal of Technology, Learning, and Assessment 8(6). https://ejournals.bc.edu/index.php/jtla/article/view/1625/1469

Hoon, T, B. (2006). Reforming ESL writing instruction in tertiary education: The writing center approach. *The English Teacher*, *35*(3), 1-14.

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

Kern, R., & Warschauer, M. (2000). Introduction: Theory and practice of network-based language teaching. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 1-19). Cambridge University Press.

Knoch, U., Rouhshad, A., Oon, S. P. & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, *28*, 39-52. doi:10.1016/j.jslw.2015.02.005.

Li, J., Link, S., & Hegelheimer, V. (2016). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Second Language* Writing, *27,*1-18. doi:10.1016/j.jslw.2014.10.004

Liao, L. (2015). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, *70*(3), 308-319. doi:10.1093/elt/ccv058.

Lipnevich, A., & Smith, J. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology Applied*, *15*(4), 319-33, doi:10.1037/a0017841.

Liu, O. L., Lee, H. S., & Linn, M. C. (2010). Multifaceted assessment of inquiry-based science learning. *Educational Assessment*, *15*(2), 69–86. doi: 10.1080/10627197.2010.491067

Shermis, D., & Burstein, J. (2003). Applications of computers in assessment and analysis of writing. *System*, *2*(8), 74-32.

Shim, S. S., Kiefer, S. M., & Wang, C. (2013). Help-seeking amongst peers: The role of goal structure and peer climate. *The Journal of Educational Research*, *106*(4), 290-300. doi:10.1080/00220671.2012.692733.

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of second language writing*, *18*(2), 103-118. doi: 10.1016/j.jslw.2009.02.003

Tang, J., & *Rich*, C. S. (2017). *Automated writing evaluation in an EFL setting*: Lessons from China. *JALT CALL Journal*, *13*(2), 117-146.

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.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157-180. doi: 10.1191/1362168806lr190

Warschauer, M., & Grimes, D. (2008). Utility in a fallible tool: A multi-site case study of automated writing evaluation. J*ournal of Technology, Learning, and Assessment*, *8*(6). https://www.researchgate.net/publication/41628129_Utility_in_a_Fallible_Tool_A_Multi-Site_Case_Study_of_Automated_Writing_Evaluation

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. doi:10.1016/j.compedu.2016.05.004

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90-102. doi:10.1016/j.asw.2018.02.004.