AUTOMATED WRITING EVALUATION SYSTEM FOR FEEDBACK IN THE DIGITAL WORLD: AN ONLINE LEARNING OPPORTUNITY FOR ENGLISH AS A FOREIGN LANGUAGE STUDENTS

Hilal YILDIZ
ORCID: 0000-0001-8706-3546
School of Foreign Languages
Ankara University
Ankara, TURKIYE

Dr. S. Ipek KURU GONEN
ORCID: 0000-0002-7374-8269
Faculty of Education
Anadolu University
Eskisehir, TURKIYE

Received: 01/09/2022 **Accepted:** 06/03/2023

ABSTRACT

It is imperative to use new technologies in a supportive manner to meet the learners' and teachers' demanding needs as educational environments change in the digital age. The continuous expansion of online learning and distance education opportunities responds to the demands of learners and teachers while pioneering the use of technology in education. One advancement in English language teaching and learning in online environments, which assists teachers in reducing their workload and providing students with instant digital feedback, is the automated writing evaluation (AWE) tools. To gain a deeper understanding of the potential and limitations of these digital tools, this study aims to investigate the effectiveness of AWE feedback in error reduction in writing in English and the explore views of students regarding the utility of AWE tools. For this purpose, a total of 38 students at a university in Turkiye participated in the study, and three of their essays were evaluated. Within a concurrent triangulation mixed-method design, the changes in errors of the experimental group (n=18) receiving AWE feedback, and the control group (n=20) receiving teacher feedback were analyzed quantitatively, and the written reflection reports and semi-structured interviews conducted with the students were analyzed qualitatively. The results indicated that teacher feedback and AWE feedback were both effective in reducing errors in 11 categories. AWE feedback appeared to minimize errors in mechanics and usage more efficiently and teacher feedback was required more in content and organization issues. As a result, AWE was found as a complementary and effective tool supporting the improvement of target language writing skills saving time and energy for teachers. Furthermore, students expressed positive views regarding the use of AWE despite minor limitations. The findings of this study in general sheds light on using online digital tools of ubiquitous nature such as AWE to assist language improvement outside the class.

Keywords: Automated Writing Evaluation system for feedback, Online learning in L2 writing, online feedback in L2 writing, L2 writing evaluation in the distance learning environment, Students' views on digital writing evaluation tool.

INTRODUCTION

In English as a Foreign Language (EFL) contexts, writing skills are fundamental to disseminating information, both inside and outside of the classroom. It is generally considered that writing is the most complex of all language skills (Lerner, 1996; Lombana, 2002), owing to the difficulty of mastering writing skills in an EFL context compared to writing in a native language context (Silva, 1993). While writing in another language (L2), learners generally are engaged in the process of going through numerous stages and cycles of writing to improve their writing skills (Deqi, 2005; Jefferson & Radhakrishnan, 2020); however, instructors face a labor-intensive and time-consuming task in providing immediate and personalized feedback. Hence, there has been an increasing interest in identifying the most efficient and satisfying methods of feedback and error correction. In the digitalized world, providing feedback on students' writing is possible via technological tools and may help to overcome challenges stemming from time constraints and teacher workload in typical EFL classrooms. Recent advancements have made it possible to provide online and e-learning opportunities that would take the heavy burden of the teachers to provide individual feedback to L2 writers regardless of the physical boundaries of the classrooms. Thus, the implementation of digital systems can make it possible for learners to access feedback in less time and effort (Jiang, Yu & Wang, 2020; Zeng, 2020).

Aligned with efforts to identify effective ways to provide feedback and improve writing skills, language education methodology evolves toward a hybrid of traditional and digital technology pedagogy with an emphasis on learner autonomy and self-regulated learning (Huang, Kuang & Ling, 2022). However, without any guidance and direct or indirect correction, students may have difficulty focusing on their writing errors and they may find it frustrating and time-consuming (Nordrum, Evans & Gustafsson, 2013). Thus, online learning opportunities guided by the teachers outside the classroom environments may not only assist L2 learners to focus on their errors but also create a space for gaining autonomy and language improvement.

As a result, current research has focused on the use of online learning technologies and web-based applications that may offer feedback to nurture and support students' learning in the improvement of writing skills with an emphasis on self-regulation. In this regard, integrating Automated Writing Evaluation (AWE) technologies into L2 writing is considered a way to mitigate the difficulties faced by teachers and students. An increasing number of foreign or second language teaching programs utilize AWE systems since they offer quick and diagnostic feedback (Grimes, 2008; Grimes & Warschauer, 2010; Zhang; 2020; Zhang & Zhang, 2018). This study aims at investigating the effectiveness of an AWE system compared to traditional teacher feedback in EFL classrooms and to explore how students perceive the use of such systems in receiving feedback for L2 writing improvement.

Online English Language Learning and Writing Skills

Although the term online language learning is interpreted differently, essentially, it is a form of language learning that happens with little to no face-to-face components, in both formal and informal settings (Ho, 2021; Hockly, 2015). Nowadays, e-learning environments, including exclusive online learning environments and blended learning environments, are becoming more prevalent and widely used in English language learning. Additionally, the sudden eruption of the pandemic became a catalyst for making virtual learning comprehensive via the electronic gateway for language skills. Technology-based activities are now so ubiquitous in English language learning that it is hard to find a class that doesn't use technology (Parvin & Salam, 2015) as a greater level of flexibility and accessibility is provided by online learning (Sari & Oktaviani, 2021). Based on how activities are structured, Felix (2003) asserts that online learning environments can be used to engage students individually or in groups in authentic settings to foster a variety of language skills. When it comes to writing skills, online English language learning environments seem to provide a platform in which alternative writing instruction can be conducted, bypassing the constraints imposed by curriculum, syllabi, session times, and classroom interaction (Tai, Pan, & Lee, 2015). Incorporating visual, auditory, and kinaesthetic learning styles, online learning environments provide each user with active control and more authentic experiences (Radianti, Majchrzak, Fromm & Wohlgenannt, 2020). Furthermore, other advantages of online learning environments are increased flexibility to monitor students' progress, providing automated feedback, and more opportunities to practice their writing. Therefore, online learning opportunities such as online automated writing platforms could facilitate teaching writing, receive effective feedback, and encourage students to improve their writing skills (Geta & Olango, 2016).

Feedback is regarded as one of the most influential factors affecting improvement in L2 writing skills (Hattie & Timperley, 2007; Joughin, 2009) since, without sufficient feedback, anything done or performed would be ineffective (Laurillard, 2013). In the EFL writing context, corrective feedback refers to teachers' attempts to minimize learners' errors (Chaudron,1988) and is an indispensable part of L2 writing since it allows learners to notice the difference between their output and target-like output. Furthermore, it stimulates a change in learners' interlanguage (Ruegg, 2010; Van Beuningen, 2010). Recent research focuses on finding out the effectiveness of different types of feedback along with teachers' and learners' preferences (Ferris & Roberts, 2001). In actual L2 writing classrooms, teachers are generally required to give feedback on all the errors or a large number of aspects. Many studies also show that EFL learners prefer teachers to correct all of their errors and inform them about different aspects of their writing performance (e.g., Amrhein & Nassaji, 2010; Ustunbau & Cimen, 2016).

On the other hand, providing personalized feedback on all errors and aspects is argued to be less practical and time-consuming (Long, 2013; Warschauer & Ware, 2006) given the workload and tight schedules. Hence, as a supplementary or additional way of responding to learners' writing, computer-mediated feedback is offered, and recent studies investigate the potential of electronic/automated feedback in EFL classrooms. One of the advantages of using computer-assisted assessment or innovative forms of assessment is that students can receive timely and repetitive feedback without time and place constraints (Brown & Bull, 2013), while some problems on the adequacy and quality of teacher feedback may occur due to time constraints in classroom settings (Gorgogeta & Vlachos, 2019). Another advantage is the potential of such innovative ways to promote self-regulated learning

while allowing teachers to improve writing instruction and enable them to make decisions regarding writing instruction in a short time (Zeng, 2020). As a result, automated systems in writing evaluation and their use in language learning contexts have become the focus of attention recently.

Automated Writing Evaluation in L2 Writing

AWE is a computer technology that provides immediate feedback and scores on written compositions (Shermis & Burstein, 2013. Nowadays, such software can be used for writing instruction, formative assessment (the process of evaluating learners' knowledge as they learn), and summative assessment (evaluating learners' knowledge at the end of an instructional period or unit) of low-stakes or high-stakes exams (Hockly, 2019). With the advancement of information, communication, and artificial intelligence technology combined with wide access to computers and the internet, researchers have been able to sophisticate the AWE tools, and more AWE products have been designed to contribute to the assessment process of large-scale and high-stakes exams such as TOEFL and GRE (Dikli, 2006). Recently, AWE has become web-based and available to be used in language learning and teaching context for providing individualized feedback on multiple drafts as well as for scoring purposes.

AWE feedback is also goal-oriented as it provides learners the opportunity to monitor their progress throughout their writing process to determine the gap between their current level and the level of writing proficiency that they desire to acquire (Cotos, 2010). Studies on the effects of formative AWE feedback on students' tendency to revise and improve the quality of writing (Chapelle, Cotos & Lee, 2015; El Ebyary & Windeatt, 2010; Warschauer & Grimes, 2008) strongly suggest that AWE feedback can encourage students to revise and improve their writing. For all these features, AWE is associated with self-regulated learning and autonomous learning. Students who receive indirect automated feedback are more likely to try and succeed through trial and error. In this way, automated feedback promotes self-directed learning, and autonomy (Taskiran, Yazici & Aydin, 2022). Additionally, AWE programs' ability to improve students' writing has been the focus of interest. Studies investigating the effectiveness of AWE in an instructional setting (Attali, 2004; Palermo & Thomson, 2018) found that students' writing performances improved in terms of both form and content as a result of AWE implementation. Furthermore, AWE is reported to improve writing quality in many aspects, such as grammar and mechanics and overall linguistic accuracy (Li, Link & Hegelheimer, 2015; Mohsen & Abdulaziz, 2019). In terms of the AWE tools' ability to decrease errors across different categories, usage, and mechanics errors were identified as less challenging to correct (Long, 2013; Ranalli, Link & Chukharev-Hudilainen, 2017).

The literature on the differences between the effects of utilizing AWE feedback and teacher feedback in formative assessment (Cheng, 2017; Khoii & Dorodian, 2013; Wang, Shang & Briody, 2013) indicates that AWE feedback might be more helpful to improve writing in terms of accuracy when compared to teacher feedback. To illustrate, a study conducted by Bulut (2019) in the Turkish EFL context examined the effect of an AWE tool on writing achievement and the findings revealed that even though both traditional and AWE assessments led to enhanced writing achievement, the students receiving AWE feedback achieved higher grades. Likewise, Gencer (2019) investigated how automated feedback affected writing performance in terms of grammar and mechanics, and the findings revealed that automated feedback had positive effects on students' writing performance in the long term as well as in the short term.

AWE is also found effective in providing scaffolding for better writing performance (Cotos, 2010) and ensures active participation of learners in their learning process. In this regard, learners have an opportunity to become more autonomous and intrinsically motivated to take control of their learning (Attali & Powers, 2008; Cotos, 2014; Weigle, 2013). Moreover, a more recent study conducted by Wilson and Czik (2016) investigated the effects of combining teacher feedback with AWE feedback on students writing motivation and the results indicated that students might benefit more from AWE feedback combined with teacher feedback. Investigating learner perceptions towards the implementation of AWE systems in L2 writing instruction may help to find out the strengths and weaknesses of these systems and how they can be used effectively to improve writing skills.

SIGNIFICANCE AND AIM OF THE STUDY

Reported advantages of AWE, such as providing instantaneous, individualized feedback on both form and content and its ability to predict accurate scores (Attali & Burstein, 2006; Elliot, 2003) have attracted the attention of testing companies, schools, educators, and researchers. However, the use of AWE in classroom settings is relatively new; thus, the use and impact of AWE feedback remain to be open to investigation (Ware, 2011). Additionally, since none of these products were designed to be used in EFL contexts, the research on AWE was limited to psychometric studies, and the studies were either mostly carried out by the software developers in laboratories or they were usually conducted in English as Second language (ESL) contexts (Warschauer & Ware, 2006).

Given the impact of the recent pandemic on educational systems worldwide, the study might inform teachers, students, and administrators in their practices on digital teaching and learning. According to Limpo, Nunes, and Coelho (2020), although technological advancements have been utilized for teaching and learning writing, the significance of these tools is better understood due to the pandemic, thereby making the studies conducted on the use of innovative tools for improving writing even more necessary. Moreover, for the new generation of students who are born into technology and labeled as "digital natives" (Prensky, 2001), the integration of innovative tools into education can promote learning and teaching (Gilakjani, 2017; Hsu, 2016) by taking the attention of learners and engaging them in the learning process. As an innovative tool, AWE might meet the needs of the 21st century if the implementation of AWE is supported by teachers' attempts to improve the writing curriculum (Wilson & Roscoe, 2020), and the findings of this study may contribute to the integration of online learning and distance learning opportunities to improve L2 writing skills. The studies examining the efficacy of AWE and the views of learners may provide a better understanding of the potential of using AWE in EFL writing, yet the integration of AWE into the educational setting has not received much research (Balfour, 2013). Based on these insights, the present study aims at investigating the pedagogical potential of AWE in EFL writing instruction by addressing the following research questions:

1. Is there a significant difference in terms of writing accuracy between a group of students receiving AWE formative feedback with a group of students receiving teacher feedback?
2. How do EFL students view the use of AWE in L2 writing?

METHODOLOGY

A concurrent triangulation mixed-methods design was employed in this study. This design involved a single study including both qualitative and quantitative data collection which were conducted at the same time (Creswell, 2009) to cross-validate, confirm, and substantiate the findings within the study and to add to the depth and scope of the findings. Concerning the quantitative dimension of the study, the comparative research design was adopted to address whether there was a difference between AWE feedback and teacher feedback in terms of writing accuracy. Regarding the qualitative dimension of the study, which investigated the students' views on the use of AWE in L2 writing assessment, data were collected through written reflection reports and semi-structured interviews.

Participants and Settings

The present study was conducted at an English Preparatory Program (EPP) of a foundation university in Turkiye. One year of the EPP includes four modules and each module lasts for eight weeks. Each week students receive a total of 25 hours of instruction in English. Four hours are allocated for writing instruction and practice weekly. In the writing lessons, the process approach to writing instruction is integrated and formative assessment is used for both helping students attain the learning objectives. The participants of this study (n=38) were selected through the convenience sampling method since the participants were accessible and voluntary (Dornyei, 2007). A higher education ethical committee approved the study, and the participants signed a consent form assuring their voluntary participation. They were also informed that they could opt out of the study at any time. The participants' age ranged between 18 to 21 years old (18 years old (n=26), 19 years old (n=10), 20 years old (n=1), 21 years old (n=1) respectively). The study took place during the participants' first term in the EPP. For the purposes of the study, two random writing classes were identified as the experimental groups, including 18 participants, and the other two classes were set as the control groups with 20 participants. All the participants were placed at the intermediate level according to a placement test which was taken prior to any exposure to writing instruction and feedback at the beginning of the term. Even though 'repeat students' (the ones who couldn't get the required score from the final proficiency exam the previous semester/year but continued studying in EPP for the following semester/year) were allowed to use AWE feedback, their data were excluded from the data as their level of proficiency might differ at the beginning of this study and their previous exposure to traditional feedback might affect their views.

Instruments

This study investigated the use of AWE in the formative assessment process, and the views of students. Accordingly, both quantitative (AWE tool *Criterion*™ for the evaluation of student essays) and qualitative (reflection reports and semi-structured interviews) data were employed. A group of experts from English language teaching and information and communication technologies was consulted during the development of all instruments for reliability and validity.

The AWE Tool

The AWE tool used in this study was the *Criterion*™ which is a web-based AWE program that was released in 2002. *Criterion*™ can score essays in terms of organization, style, grammar, mechanics, usage, vocabulary use, and lexical complexity and it provides corrective feedback on a variety of error types such as formation errors, punctuation, agreement errors, wrong word, spelling errors, etc. In the study context, teachers provide feedback via a correction code for errors. The AWE tool, on the other hand, provides explanations of the error and offers suggestions for correction. Figure 1. shows how *Criterion*™ as an AWE tool provides feedback on students' errors in comparison with the teacher's feedback.

Error Type	Sample errors from students' essays	Teacher's feedback	Sample AWE feedback
Ill-formed verbs	They always want *to following* the ….	WF	This verb may be incorrect. Proofread the sentence to make sure you have used the correct form.
Missing or Extra Article	*First* reason why internet is beneficial for people is …	^ or X	You may need to use an article before this word. Consider using the article the
	On the contrary, you can meet *new person*, and …		You may need to use an article before this word. Consider using the article a
Punctuation	Money brings *happiness but* reach people …	P	You may need to place a comma after this word
	There are three reasons to believe *that,* internet is the best invention.		You may need to remove this comma

Figure 1. Sample Feedback Provided by Teachers and the AWE System

When compared to other AWE programs and web-based AWE systems, the *Criterion*™ is considered one of the most accurate AWE tools regarding feedback and scoring (Zupanc & Bosnic, 2015). This program provides formative writing assessment, and the students are allowed to use the system to write multiple drafts outside the classroom or use the system for other lessons or personal study as well. The purpose of the study is not to promote any commercial tool but to investigate the effectiveness of AWE in L2 writing. Thus, other similar programs can also be installed and used in different contexts.

Reflection Reports

Weekly written reflection reports were chosen as the data collection instruments to reveal students' experiences through guided reflection questions immediately after they finish writing their essays using AWE feedback. Students were asked to express their ideas and feelings regarding the use of AWE for feedback, and error correction along with the strengths and weaknesses of the AWE tool. Reflection reports were gathered each week considering the possibility that students' opinions would alter over time. 54 written reflection reports were therefore included in the analysis of qualitative data.

Semi-Structured Interviews

To uncover students' opinions on the use of AWE for formative assessment and evaluation purposes in this study, semi-structured interviews with 13 students from the experimental group who regularly used AWE feedback were conducted during the final week. These interviews allowed for a thorough examination of the participants' viewpoints, their justification, and clarification (Galletta, 2013). Students were asked to state their opinions on the use and effectiveness of AWE for feedback and error correction in detail. Moreover, challenges of the digital writing tool, comparison of teacher and AWE feedback, their feedback

type preference, and potential of using AWE in the future were among the questions directed at the students. The interviews were held at appropriate times for the students and each interview approximately lasted 15 minutes. The interviews were transcribed verbatim.

Data Collection Procedures

The study was carried out in the fall term of the 2019- 2020 academic year at four intermediate-level classes. At the beginning of the study, all the necessary permissions were granted by the Educational Testing Service (ETS), to use *Criterion*™ for this study. In the study context, intermediate-level writing instruction and writing practice focused on fundamental principles of writing in L2 and paragraph writing for the first three weeks. During this period, the students received constant corrective feedback and written comments regarding their errors. This step was identical for both groups as the students from the experimental group also needed to be exposed to the teacher's feedback so that they could unveil their experiences with both feedback types and provide information regarding the similarities and differences. In the fourth week, the requirements, and stylistic features of the essay genre (opinion essay) were introduced to both groups. Additionally, the AWE tool was introduced to the experimental group, tutorials provided by ETS were shared with the students, and students were trained on using AWE features. Furthermore, students were provided with their usernames and passwords. In the fifth, sixth, and seventh weeks, both groups wrote opinion essays. The students in the experimental group (n=18) and the control group (n=20) wrote essays on different topics each week and two drafts were written for each topic. Hence, 228 essays including the second drafts were examined in this study. The topics were the same for both the experimental and the control groups each week, and the length of essays ranged between 250 and 300 words. The first and the second essays were written by hand by both the experiment and the control groups to ensure that using AWE feedback to improve writing was the only independent variable of the comparative analysis since changing the writing environment to digital/computer would be another variable. Hence, the original student essays of the experimental group were transformed into digital files and uploaded to the students' accounts and the students were able to display the AWE feedback to write their second draft.

The control group received regularly used error codes and written comments from the teacher to revise their essays, whereas the experimental group received AWE feedback to write their second drafts. Within the same week, both groups received feedback provided for their second drafts as well. The experimental group wrote reflection reports on the AWE feedback they received. At the end of the eighth week, students' thoughts about their experiences with the AWE tool were obtained through semi-structured interview questions.

Data Analysis

To compare the AWE feedback and teacher feedback in error reduction and find out whether there was a significant difference between these two types of feedback in terms of correcting errors, the first and second drafts of essays were treated, respectively, as pre-feedback and post-feedback conditions. The feedback type was considered as the independent variable and to examine the error reduction across drafts, Ferris's model (2006) which focuses on investigating the effectiveness of a feedback type on common errors was adopted.

To find the common errors, the errors on the first drafts of each writing task were counted. To ensure inter-coder reliability, the errors on 30% of the essays were categorized and counted by another experienced instructor. Miles and Huberman's (1994) formula [(consensus)/ (consensus + disagreement) x100] was used and the inter-rater reliability was found .89 indicating high reliability. As a result, a total of 11 error types were included in the comparative analysis. After the common error types were found, the error counts were normalized by using the formula suggested by Biber, Douglas, Conrad, and Reppen (1998) (error count x essay length/ average word count) since essay length was another variation. The error counts on both drafts of the first, the second, and the third essays were normalized to the average of, respectively, 276, 282, and 295 words. Once the error counts were normalized, Shapiro-Wilk, skewness-kurtosis, and histogram graphs were examined to determine distribution. Since the data did not meet the normality assumptions, the Wilcoxon signed-rank analysis was used because when comparing matched pairs of observations, it is the optimal nonparametric test that can be used (Chamberland, Wang & Roessler, 2008), and this test was conducted

to determine the error reduction from pre-feedback to post-feedback within each group. Moreover, Mann-Whitney U analysis was used to determine whether there was a significant difference between the groups according to the post-tests. In the case of non-normally distributed data, this non-parametric test can determine whether or not the medians of two groups differ significantly from one another (McDonnell, Connell, Hannif & Burgess, 2013).

To analyze the students' views on the use of AWE, written reflection reports and interviews were analyzed. In qualitative data analysis, Constant Comparison Method (CCM) which has roots in grounded theory research (Glaser & Strauss 1967) was used. CCM was utilized to interpret categories emerging from codes grounded in data rather than preconceived ones (Charmaz, 2006). A code in this study refers to the words or phrases that stand in for feelings and thoughts and codes disclose participants' perspectives. To create subcategories, the codes were then contrasted and compared both within and between the sets of data. Finally, new subcategories were combined into main categories. To ensure reliability, another experienced rater analyzed the data as well. For inter-rater reliability, the agreement formula [(the number of agreements)/(the number of agreements + the number of disagreements) x100] suggested by Tawney and Gast (1984) was used, and it was found .91 denoting a high level of reliability.

FINDINGS

RQ1: The Change in EFL Students' Writing Accuracy After AWE Feedback and Teacher Feedback

To find out whether the experimental and control groups were similar in terms of writing accuracy at the beginning of the study, the pre-feedback errors of both groups were analyzed. Mann-Whitney U analyses revealed that although spelling errors showed a significant difference in favor of the experimental group ($U = 109.50$, $p < .05$) in the first writing task and spelling errors ($U = 105.50$, $p < .05$) and capitalization errors ($U = 90.50$, $p < .05$) showed a significant difference in the second writing task, there was no significant difference ($p > .05$) between groups in the final writing task before the provision of AWE feedback. Moreover, when Mann-Whitney U analyses of the first drafts are taken into consideration, it can be inferred that the number of errors made by the students in both groups is mainly similar. In other words, there was no significant difference between the groups in most variables examined in the context of the essays ($p > .05$). This situation reveals that before the students in the experimental group were exposed to AWE feedback, the students in both groups were similar to each other in terms of writing accuracy except for spelling errors and capitalization errors on a single occasion.

To determine the accuracy change depending on the feedback type, the Wilcoxon signed-ranks test was conducted to find out whether there was a significant difference in the pre-feedback and post-feedback error counts of the students in the experimental and control groups. This test was conducted to compare both groups' pre-feedback and post-feedback error counts for their first, second, and third essays to identify the exact change according to the feedback type. The results of the Wilcoxon signed-ranks related to the students' essays are presented in Table 1.

Table 1. The comparison of pre-feedback and post-feedback error counts of students' essays

		The First Essay								The Second Essay								The Third Essay							
		Experimental Group				Control Group				Experimental Group				Control Group				Experimental Group				Control Group			
Error Types		N	Mr	Z	p	N	Mr	Z	P	N	Mr	Z	p	N	Mr	Z	P	N	Mr	Z	P	N	Mr	Z	P
Fragment (pre-post)	Nr	8	4,50			12	7,50			12	7,13			12	6,50			13	7,00			13	7,00		
	Pr	0	,00	-2,521	**,012***	1	1,00	-3,111	**,002***	1	5,50	-2,796	,005*	0	,00	-3,064	**,002***	0	,00	-3,181	**,001***	0	,00	-3,181	**,001***
	Ties	10				7				5				8				5				7			
Run-on (pre-post)	Nr	15	8,00			14	8,50			11	6,00			15	8,00			16	8,50			15	8,00		
	Pr	0	,00	-3,408	**,001***	1	1,00	-3,352	**,001***	0	,00	-2,936	,003*	0	,00	-3,411	**,001***	0	,00	-3,516	**,000***	0	,00	-3,408	**,001***
	Ties	3				5				7				5				2				5			
Sentence structure (pre-post)	Nr	12	6,50			11	6,00			6	3,50			7	4,00			11	6,00			9	5,00		
	Pr	0	,00	-3,059	**,002***	0	,00	-2,937	**,003***	0	,00	-2,207	,027*	0	,00	-2,366	**,018***	0	,00	-2,936	**,003***	0	,00	-2,668	**,008***
	Ties	6				9				12				13				7				11			
SVA (pre-post)	Nr	15	8,00	-3,408	**,001***	18	9,50	-3,724	**,000***	13	7,00			15	8,00			12	6,50			16	8,50		
	Pr	0	,00			0	,00			0	,00	-3,180	,001*	0	,00	-3,408	**,001***	0	,00	-3,059	**,002***	0	,00	-3,516	**,000***
	Ties	3				2								5				6				4			
Ill-formed verb (pre-post)	Nr	17	9,00			19	10,00			13	7,00			19	10,00			14	7,50			13	7,00		
	Pr	0	,00	-3,621	**,000***	0	,00	-3,824	**,000***	0	,00	-3,180	,001*	0	,00	-3,823	**,000***	0	,00	-3,296	**,001***	0	,00	-3,180	**,001***
	Ties	1				1				5				1				4				7			
Wrong word (pre-post)	Nr	9	5,00			10	5,50			14	7,50			11	6,36			14	7,50			14	7,50		
	Pr	0	,00	-2,666	**,008***	0	,00	-2,803	**,005***	0	,00	-3,296	,001*	1	8,00	-2,432	**,015***	0	,00	-3,296	**,001***	0	,00	-3,296	**,001***
	Ties	9				10				4				8				4				6			

Table 1. *(Continued)* The comparison of pre-feedback and post-feedback error counts of students' essays

Error Types		The First Essay								The Second Essay								The Third Essay							
		Experimental Group				Control Group				Experimental Group				Control Group				Experimental Group				Control Group			
		N	Mr	Z	p	N	Mr	Z	P	N	Mr	Z	p	N	Mr	Z	P	N	Mr	Z	P	N	Mr	Z	P
Article (pre-post)	Nr	17	9,00			8	4,50			17	9,00			17	9,00			**16**	8,50			10	6,40		,006*
	Pr	0	,00	-3,621	**,000***	0	,00	-2,521	**,012***	0	,00	-3,621	,000*	0	,00	-3,622	**,000***	0	,00	-3,517	**,000***	1	2,00	-2,758	
	Ties	1				12				1				3				2				9			
Preposition (pre-post)	Nr	15	8,00			19	10,00			15	8,00			19	10,00			11	6,00			12	6,50		
	Pr	0	,00	-3,409	**,001***	0	,00	-3,824	**,000***	0	,00	-3,409	,001*	0	,00	-3,824	**,000***	0	,00	-2,934	**,003***	0	,00	-3,059	,002*
	Ties	3				1				3				1				6				8			
Spelling (pre-post)	Nr	13	7,00			18	10,50			13	7,00			18	10,50			15	8,00			16	8,50		
	Pr	0	1,00	-3,180	**,001***	1	1,00	-3,783	**,000***	0	,00	-3,180	,001*	1	1,00	-3,783	**,000***	0	,00	-3,408	**,001***	0	,00	-3,517	**,000***
	Ties	5				1				5				1				3				4			
Capitalization (pre-post)	Nr	11	7,00			16	8,50			11	7,00			16	8,50			10	5,50			9	5,00		
	Pr	1	,00	-2,981	**,003***	0	,00	-3,516	**,000***	1	1,00	-2,981	,003*	0	,00	-3,516	**,000***	0	,00	-2,805	**,005***	0	,00	-2,666	,008*
	Ties	6				4				6				4				8				11			
Punctuation (pre-post)	Nr	18	9,50			20	10,50			18	9,50			20	10,50	-3,921	**,000***	16	8,50			20	10,50		
	Pr	0	,00	-3,724	**,000***	0	,00	-3,921	**,000***	0	,00	-3,724	,000*	0	,00			0	,00	-3,516	**,000***	0	,00	-3,920	**,000***
	Ties	0				0				0				0				2				0			

Based on negative ranks

* Statistically significant ($p < .05$)

Mr: Mean rank

Nr: Negative ranks

Pr: Positive ranks

As observed in Table 1, there was a statistically significant difference in both groups' revisions based on the type of feedback for the first essays ($p < .05$ for all variables). The error counts in both groups were in favor of negative ranks, indicating a reduction in errors. Likewise, in the context of the second student essays, both types of feedback led to a statistically significant reduction in all variables (error types) examined in the experimental and control groups ($p < .05$ for all variables). Furthermore, according to the same analyses conducted for the final writing task, both types of feedback resulted in statistically significant differences in all variables (11 types of error) tested in experimental and control groups ($p < .05$ for all variables). To put it in a nutshell, for both groups, both types of feedback (AWE and teacher feedback) significantly reduced the error counts in each category. In addition to within-groups comparison, between-groups comparison was made using the Mann-Whitney U analysis. Accordingly, whether the type of feedback given to the first drafts of the essays showed a significant difference in the post-feedback scores of the experimental and control groups was examined. The results are presented in Table 2.

Table 2. Post-feedback scores of the experimental and control groups

Writing Tasks (post-feedback)		Experimental (N=18)		Control (N=20)			
		Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks	U	P
The First Essay	Fragment	20,22	364,00	18,85	377,00	167,00	,693
	Run-on	18,00	324,00	20,85	417,00	153,00	,371
	Sentence structure	18,94	341,00	20,00	400,00	170,00	,730
	Subject-verb agr.	17,89	322,00	20,95	419,00	151,00	,210
	Ill-formed verbs	15,89	286,00	22,75	455,00	115,00	,044*
	Wrong word choice	19,61	353,00	19,40	388,00	178,00	,951
	Article	9,78	176,00	28,25	565,00	5,00	,000*
	Preposition	18,75	337,50	20,18	403,50	166,50	,686
	Spelling	18,39	331,00	20,50	410,00	160,00	,465
	Capitalization	18,39	331,00	20,50	410,00	160,00	,357
	Punctuation	13,06	235,00	25,30	506,00	64,00	,000*
The Second Essay	Fragment	19,17	345,00	19,80	396,00	174,00	,849
	Run-on	18,75	337,50	20,18	403,50	166,50	,641
	Sentence structure	18,50	333,00	20,40	408,00	162,00	,534
	SVA	17,67	318,00	21,15	423,00	147,00	,196
	Ill-formed verbs	19,06	343,00	19,90	398,00	172,00	,798
	Wrong word	17,44	314,00	21,35	427,00	143,00	,261
	Article	10,67	192,00	27,45	549,00	21,00	,000*
	Preposition	20,11	362,00	18,95	379,00	169,00	,734
	Spelling	18,22	328,00	20,65	413,00	157,00	,385
	Capitalization	18,94	341,00	20,00	400,00	170,00	,619
	Punctuation	10,72	193,00	27,40	548,00	22,00	,000*
The Third Essay	Fragment	20,42	367,50	18,68	373,50	163,50	,618
	Run-on	17,86	321,50	20,98	419,50	150,50	,319
	Sentence structure	19,22	346,00	19,75	395,00	175,00	,863
	SVA	18,08	325,50	20,78	415,50	154,50	,352
	Ill-formed verbs	17,06	307,00	21,70	434,00	136,00	,152
	Wrong word	19,06	343,00	19,90	398,00	172,00	,809
	Article	12,33	222,00	25,95	519,00	51,00	,000*
	Preposition	17,97	305,50	19,88	397,50	152,50	,578
	Spelling	18,25	328,50	20,63	412,50	157,50	,356
	Capitalization	19,56	352,00	19,45	389,00	179,00	,956
	Punctuation	5,42	277,50	23,18	463,50	106,50	,027*

SVA: Subject-verb agreement. U: Difference between the two rank totals. P: Significance of the difference between the medians

In Table 2, a complete picture can be obtained by observing the sum of ranks. The sum of ranks above shows that the difference across all error types was in favor of AWE feedback. In other words, thanks to AWE feedback, the participants in the experimental group made less errors in their second drafts. As for the error types which reduced significantly, the Mann-Whitney U analysis showed that the error types that reduced significantly were ill-formed verbs, ($U = 115.00$, $p < .05$), article ($U = 5.00$, $p < .05$), and punctuation errors ($U = 64.00$, $p < .05$) in the first writing task. In the second writing task there was a significant difference only in the article ($U = 21.00$, $p < .05$) and the punctuation ($U = 22.00$, $p < .05$) errors between the groups, and for the final writing task, it was determined that there was only a significant difference in the article ($U = 51.00$, $p < .05$) and punctuation ($U = 106.50$, $p < .05$) errors between the groups after different types of feedback was received by the experimental and the control groups. In other words, post-feedback results yielded that in terms of the ill-formed verbs, article, and punctuation error counts, the average of the error counts of the students in the experimental group decreased significantly while the mean scores of the error counts of these error types did not decrease significantly in the control group.

These results revealed that the feedback given by the AWE system in the experimental group was more effective in correcting the ill-formed verbs, articles, and punctuation errors compared to the feedback given by the teacher in the control group. In other words, regarding the comparison made over the post-feedback error counts of the groups, there was a significant difference between the two groups in terms of the reduction in the ill-formed verbs, articles, and punctuation errors. Sample AWE feedback and teachers' feedback provided for ill-formed verbs, articles, and punctuation errors made by the students in this study are available upon request from the authors.

RQ2: EFL Students' Views on the Use of AWE in L2 Writing

As a result of the qualitative analysis of the weekly reflection reports and semi-structured interviews regarding the implementation of AWE in L2 writing, a total of 922 codes were identified. Constant comparison of the codes resulted in 15 sub-categories and five main categories. The students expressed their views on various aspects of the AWE system. The main categories related to students' views on the use of AWE in L2 writing can be seen in Table 3.

Table 3. The main categories related to students' views on the use of AWE in L2 writing

Main categories	N*
Improvement in writing performance	324
Advantageous features	224
Comparison of AWE feedback and teacher feedback	157
Gains in autonomy	115
Drawbacks	102
Total	922

N*: Number of codes

As visible in Table 3, the main categories including the respective codes within each category regarding the participants' views on AWE were: improvement in writing performance (n=324), advantageous features (n=224), comparison of AWE feedback and teacher feedback (n=157), gains in autonomy (n=115), and the drawbacks (n=102). That is, the students expressed some positive views about the AWE tool as it helped them to improve their L2 writing performance (i.e. reducing errors, promoting learning via feedback, and increasing attention on errors), it has various advantageous features (i.e. technical advantages, advantages in the quantity and quality of feedback), it helped to compare teacher feedback and AWE feedback (i.e. helped to see the differences and the similarities of two feedback types and helped to find out the preferred feedback type), it had certain gains in learner autonomy in L2 writing (i.e. independence from the teacher, increased self-correction, self-awareness of strengths and weaknesses, and increased motivation). What is more, the students mentioned several drawbacks of the AWE tool such as the difficulty in correcting certain errors and some system-related problems. In the subsequent sections, each main category along with the related sub-categories is given in detail with direct quotations from students' responses.

Improvement in Writing Performance

The first main category was related to the effect of AWE feedback on improving writing performance. Students expressed their ideas about how the system helped to reduce errors, increased their attention to errors, and promoted learning the language in general. The subcategories under this main category can be seen in Table 4.

Table 4. The subcategories related to the effect of AWE on the improvement in writing performance

Writing performance	Explanatory statements	N*
Reduction in errors	*I corrected most of my errors*	249
Promotion of learning via feedback	*I learned that nouns are followed by articles*	54
Increased attention on errors	*I realized the frequent errors and tried not to make them*	21
Total		324

N*: Number of codes

The majority of the students pointed out that they were able to correct their errors thanks to the feedback provided by the system. Students were able to notice a reduction in their errors with the help of instant and regular feedback provided by the AWE system. Moreover, AWE feedback was found to promote learning a foreign language. The students' statements revealed that they might have reached beyond completing the given task since the feedback they received also resulted in learning gains in terms of language use. Finally, the students also noted that they realized the frequent errors they made, and they paid attention to avoid making the same errors while writing their second drafts and other essays. The excerpts below display students' ideas regarding the effects of AWE on their writing performance:

> *S4: "Thanks to the feedback I received, now I know what kind of errors will occur in my writing, and I try not to make those errors."*
>
> *S2: "I have probably revised the feedback given by the system about a thousand times, I used to a have problem in constructing sentences, and I was structuring sentences much differently than I do know. I have learned how to construct my sentences properly by revising the feedback there many times."*

Advantageous Features of the AWE System

The second main category emerging from students' statements was the advantageous features of the AWE system (n=224). Table 5. displays the sub-categories related to the advantageous features of the AWE system.

Table 5. The subcategories related to the advantageous features of the AWE system

Advantageous features	Explanatory statements	N*
Technical advantages	*You can use the system regardless of time and place*	146
Quantity of feedback	*It finds all the errors and evaluates every aspect of the writing*	42
Quality of feedback	*It provides detailed and explanatory feedback*	36
Total		224

N*: Number of codes

Concerning the advantages of using an AWE tool, the majority of the students mentioned that the system can be used on different devices such as smartphones and computers, and it can be used at any time and anywhere. Hence, the availability of the system and its ubiquitous nature were regarded as an advantage by the participants. Moreover, participants pointed out that receiving immediate feedback allowed them to revise and resubmit their essays a couple of times outside the school. Therefore, receiving immediate feedback was timesaving and encouraged students to write multiple drafts. The participants also noted

they were pleased with the amount of feedback given by the system, and they expressed that the abundant quantity of feedback provided by the system helped them to improve writing accuracy on their second drafts. Another reported advantageous feature of AWE feedback was the quality of feedback. The students found AWE feedback clear and to the point as the system provided detailed explanations along with direct or indirect hints. The following excerpts exemplify students' views on the advantageous features of the system:

> *S1: "It (the AWE system) shows all the errors, and it tells you exactly what category the error falls into. Since it finds all the errors even if it is a missing comma, you can see all of the errors on the system"*
>
> *S5: "I think the system can't overlook errors because it has an established system. On the other hand, teachers might be exhausted while assessing writing.*

Comparison of AWE Feedback and Teacher Feedback

Students reflected on their experiences of receiving teacher feedback and AWE feedback to improve their writing in L2 some similarities and differences between the two types of feedback were reported. The sub-categories related to the comparison of the two feedback types are displayed in Table 6.

Table 6. The subcategories related to the comparison of AWE feedback and teachers' feedback

Comparison of AWE and teachers' feedback	Explanatory statements	N*
Differences between AWE feedback and teachers' feedback	*AWE feedback is instant, teacher feedback is delayed*	88
Similarities between AWE feedback and teachers' feedback	*Both types of feedback are indirect and focused on similar error types*	38
The preferred use of feedback in L2 writing	*We can use AWE feedback in general, but teachers can comment on our ideas*	31
Total		157

N*: Number of codes

In general, the students highlighted some differences between these two feedback types. Some of the differences frequently mentioned by students were that they could only receive teacher feedback in the school environment, and they have difficulty receiving feedback for the essays due to the teachers' workload, the number of students in the class, and the limited class hours. Thus, AWE feedback and teacher feedback were perceived differently due to some challenges and constraints related to the school environment. However, according to the students, the AWE feedback and the feedback given by the teachers focused on similar error types, especially the error types under the grammar, usage, and mechanics categories. The students pointed out that both types of feedback guided them to correct their errors on their own instead of correcting the errors directly. Thus, it can be inferred that both teachers' feedback and AWE feedback can be considered indirect feedback types and were effective in writing improvement. Furthermore, the students expressed that the teachers can guide them better than AWE feedback in terms of content, mainly due to the lack of content-related feedback on the AWE system. Hence, it was revealed that the preferred way to use feedback was a mixed use of AWE feedback and teacher feedback. When students were asked about how they would like to benefit from these two types of feedback, the majority of the students expressed their willingness to continue using the AWE tool in the absence of a writing instructor, especially outside the school for self-regulated learning purposes and to be able to practice writing more.

Gains in Autonomy

Another main category that emerged from the students' statements was related to the positive effects of the AWE tool on gaining autonomy. Automated feedback was found to create self-awareness of the strengths and weaknesses in language use, lead to independency, increase self-correction, and promote motivation. The subcategories of this main category on gaining autonomy can be seen in Table 7.

Table 7. The subcategories related to gains in autonomy

Gaining autonomy	Explanatory statements	N*
Independence from teacher	*We don't need to wait for a teacher to give feedback*	68
Increased self-correction opportunities	*I noticed the areas of improvement and studied accordingly*	21
Self-awareness of strengths and weaknesses	I realized that need to improve my grammar	17
Increased motivation	*Receiving feedback quickly and easily makes you want to practice*	9
Total		115

N*: Number of codes

The students pointed out that, using the AWE tool fostered independence from the teacher. As most of the writing practice took place at school guided by the teachers, students were mostly dependent on the feedback provided by the teachers to improve their writing ability. However, students pointed out that thanks to the ubiquitous nature of the AWE tool, they were able to practice outside the school and they started to feel less dependent on their teacher. Furthermore, it was reported that the system allowed for more self-correction opportunities. The participants stated that the AWE feedback urged them to find out how to correct their errors without direct provision; hence, they had more chances to analyze their errors on their own. Additionally, seeing their errors grouped under certain categories helped them to understand their strength in language use and writing skills, and they were also able to identify the areas for further improvement in their writing. Hence, they realized what they needed to focus on to improve their writing skills. The fourth and final subcategory was an increase in self-motivation to write more and receive more feedback. The following excerpts set an example of participants' views on the aforementioned issues:

> *S1: "In addition to writing essays, I would also use the system for my hobbies. For instance, I like to write stories in English, but since it is something I do outside the school, I can't ask my teachers to correct the errors in my stories. However, I will be able to use the system whenever I need to write in English, so I can also use it for other purposes."*
>
> *S11: "Thanks to easy access to feedback, we were not dependent on the time we spend in the classroom. We were also able to work at home, so we were able to do everything on our own."*

Drawbacks of the AWE System

Albeit various positive views related to the use of AWE in L2 writing, students also expressed some drawbacks of the system. Some difficulties in correcting certain errors and system-related challenges were reported. The sub-categories related to the drawbacks of the AWE tool are displayed below in Table 8.

Table 8. The subcategories related to the drawbacks of using the AWE tool

Drawbacks of AWE	Explanatory statement	N*
Difficulties in correcting certain errors	*I couldn't change some of the repeated words*	74
System-related issues	*The system detected a website name as an error*	28
Total		102

N*: Number of codes

Concerning the drawbacks of the AWE feedback, the students expressed that certain errors were rather challenging to correct compared to others. According to the students, they had difficulty correcting the errors related to the stylistic issues, such as the repetition of words and short sentences. Moreover, some students expressed that the system was not able to identify specific names or concepts in the essays and detected them as errors Thus, it can be inferred that receiving AWE feedback in isolation may not be

sufficient for correcting certain errors, or students may need further writing instruction to improve certain aspects of writing. In addition to difficulties in correcting certain errors, it was pointed out by a few students that they experienced some system-related issues such as the complex features of the interface.

All in all, the findings concerning student views on using the AWE system in L2 writing yielded mainly positive results. Both quantitative and qualitative findings indicated the effectiveness of both AWE feedback and teacher feedback and highlighted the use of AWE not as a replacement for teacher feedback but as complementary support. Albeit some reported challenges, the results of the study, in general, revealed numerous advantages of using an automated digital writing feedback tool that can be used outside the class without time and place constraints.

DISCUSSION

The findings of the study point to the effectiveness of both AWE feedback and teacher feedback in improving L2 writing accuracy. That is, both teacher feedback and AWE feedback led to a significant decrease in writing errors. This concurs well with the previous studies in the literature suggesting that using AWE formative feedback results in positive changes in writing performance by reducing errors (Attali, 2004; Li et al., 2015; Long, 2013; Polermo & Thomson, 2018; Ranalli et al., 2017; Warschauer & Grimes, 2008). In the Turkish EFL context, in a study by Gencer (2019), it was found that AWE feedback had positive feedback on students' L2 writing performances similar to the findings of the present study. The results also revealed better scores in the post-feedback assessment of essays on the behalf of the students who received AWE feedback, a finding echoing Bulut's (2019) study in the Turkish context. The students in the study highlighted the practicality of AWE feedback to save time, instant access regardless of the boundaries of the school context, and receive more feedback for multiple drafts without adding a burden to teachers' workload. Therefore, the findings of the current study support the notion that AWE has great potential in not only guiding students toward improving their writing performances but also reducing teachers' burden of correcting errors in each student's essay regularly and facilitating the assessment process.

Besides the general improvement in writing accuracy, it was ascertained that although both teacher feedback and AWE feedback led to a significant decrease in error counts, there was a significant difference between the two groups in three categories, namely, ill-formed verbs, articles, and punctuation. Hence, the AWE feedback was consistently more effective than teacher feedback in correcting article, punctuation errors and correcting ill-formed verb errors while there was no significant difference between the effectiveness of teacher feedback and AWE feedback in correcting the errors in the other eight categories. In this respect, the findings substantiate other previous findings in the literature. To exemplify, both studies conducted by Khoii and Doroudian (2013), and Wang et al. (2013) reported significant findings in favour of AWE feedback as in both studies the experimental groups significantly outperformed the control groups in writing accuracy. Concerning the error types that AWE was more successful in correcting, the findings of this study correlate favorably with previous studies (Long, 2013; Li et al.,2015; Ranalli et al.,2017, Warshauer & Grimes, 2008) as the findings of the current study further support the idea suggested by Ranalli et al. (2017) that AWE feedback can be more useful to improve form-based aspects of writing and the AWE feedback was more effective on word-level errors and most of the improvement was mainly on grammar and mechanics aspects of writing (Warshauer & Grimes, 2008).

Feedback provided by the AWE system and the teachers for error correction differed in significant ways. To illustrate, while teachers used codes such as '^' for 'article' and 'P' for 'punctuation' to address errors, the AWE tool used in this study specified these errors in more detail by providing explanations and suggestions. Students also reported the effectiveness of such detailed feedback compared to the teachers' feedback. Hence, it can be assumed that although error codes are frequently used to give indirect corrective feedback and are preferred by teachers to save time, students may not understand how to correct these errors and they may need further explanations (Lee, 2003). This finding has insights for L2 writing assessment as it shows that AWE feedback is timesaving, and it may also provide guidance and effective feedback. Although this study did not focus on the effectiveness of AWE on discourse-level writing performance, the findings point to the usefulness of AWE feedback on word-level and sentence-level improvement. Thus, the findings may

suggest that AWE feedback can be used to provide formative corrective feedback on errors, and it can be combined with teacher feedback to ease teachers' burden by allowing them to focus only on the aspects of writing which AWE might disregard or does not provide feedback on. As suggested by Ersanli and Yesilel (2023), while assisting teachers with surface-level errors, AWE tools enable them to focus on content and organization instead of repetitive surfaces-level errors.

Students' views on the AWE system also support the qualitative findings. The present study point to overall satisfaction with the AWE tool and the AWE feedback. The students expressed that they were satisfied with the quantity and the quality of the feedback provided by the system highlighting its various features such as its trustworthiness, availability, and practicality. These views of the students are in parallel with the findings of previous studies regarding students' positive attitudes toward AWE (El Ebyary & Windeatt, 2010, Cheng, 2017; Li et al., 2015; Parra, 2019; Zhang, 2020). One explanation of this can be attributed to students' eagerness to use technology for improving language skills and their need to receive more feedback for language improvement. As suggested by Wang, Chen, and Fang (2011), students' perception of to use of technology in education is parallel with their needs and their familiarity with it.

Another noteworthy finding was that the AWE feedback had positive effects on gaining autonomy. It was expressed by the students that using AWE feedback raised their awareness of their strengths and weaknesses which ultimately increased their learning efforts. Therefore, it can be interpreted that the AWE feedback may have raised students' metalinguistic and meta-strategic awareness which are crucial for a revision stage to be meaningful since detecting errors or flaws in writing and developing strategies improve writing ability (Cotos, 2014). Moreover, regarding gaining autonomy through AWE, the students asserted that using AWE feedback fostered independence from the teacher as they were able to use the AWE system outside the school for many different purposes in addition to the assignments. It is obvious that writing in L2 is one of the most challenging areas for students studying in English-medium programs in the Turkish EFL context (Kamasak, Sahan & Rose, 2021) and providing detailed feedback can be too demanding for English instructors (Gurel, 2010). One implication of the findings might be that the AWE tools can be utilized to promote autonomy, self-regulated learning, and it can be used for various purposes. The AWE tool was also found to increase students' motivation to write more and receive more feedback. The findings regarding the relationship between using AWE and motivation are consistent with previous studies suggesting that AWE tools increase students' motivation to write in L2 and help them revise their writing (Bulut, 2019; Li et al., 2015; Wilson & Czik, 2016; Wilson & Roscoe, 2020) and encourage them to analyze their errors privately by a fast and individualized explanation (Parra & Calero, 2019). Given that intrinsic motivation is believed to be essential to initiate self-regulated learning and promoting learning efforts (Pintrich, 1999; Stone, 2017), the students might have become more motivated as their revisions were more learning-oriented.

Students' reflections on their experiences with the traditional feedback and the AWE feedback yielded that there were some similarities and differences between the traditional feedback and AWE feedback. Some of the similarities reported by the students were the types of error categories and error types for which feedback was provided, the usefulness of both feedback types, and the guidance received by the feedback received from the system and the teacher. Therefore, it can be assumed that AWE feedback is compatible with traditional feedback in leading learners to work on their errors (Calvo & Ellis, 2010). As for the differences, it was mentioned by the students that AWE feedback was more available compared to teacher feedback. However, receiving automated feedback was insufficient in realizing content-related errors. In line with the findings of Zhang and Hyland's (2018) study, these findings indicate that AWE feedback is rather form-focused whereas the teacher's feedback focused more on content and organization. Therefore, when asked what type of feedback they would prefer to receive, the majority of the students expressed their willingness to receive a combination of both feedback types. Similarly, previous studies which investigated the effects of combining teacher feedback with AWE feedback on students' writing motivation (Wilson & Czik, 2016; Mohsen & Abdulaziz, 2019; Palermo & Thomson, 2018) revealed that the students who received a combination of AWE feedback and teacher feedback were more motivated to revise and improve their writing. As a result, this study highlights the effectiveness of using AWE systems along with teacher feedback to improve L2 writing, especially in EFL contexts where students need more feedback and opportunities for correction.

CONCLUSION AND IMPLICATIONS

The findings of the present study highlight the importance of implementing useful and effective online tools that provide effective feedback and error correction opportunities in L2 writing which is of utmost importance in the new era of digitalization that gained momentum due to the current pandemic process. Considering the need for students to get feedback from reliable sources when they are on their own, it is crucial to identify and use effective tools. Therefore, introducing online learning tools such as the AWE is of great importance, especially in EFL writing. As the results of the current study indicate that even though AWE feedback may not replace teacher feedback altogether, it can complement teacher feedback in a way that enables teachers to focus more on global issues of writing such as content and organization. Apart from that, encouraging students to use AWE-like tools outside the classroom as an online learning opportunity can allow learners to become autonomous and have more opportunities to assess themselves and receive feedback on their writing.

Drawn from the aforementioned findings of the current study, some pedagogical implications are that AWE can be used by students during the preparatory year and while continuing their studies at the faculty to check and correct errors and improve linguistic accuracy in written texts, such as assignments and research papers by helping students gain autonomy and engage in self-improvement activities without depending on teachers. One of the implications of this study regarding teachers is that teachers may receive great benefits when assistive technology such as AWE tools are integrated into the summative and formative assessment of L2 writing. By doing so, teachers can reduce their workload significantly, and they can save a lot of time by giving individual or corrective feedback to students on other aspects of writing, such as meaning and organization. In addition, teachers can encourage students to use AWE outside the classroom by assigning writing tasks, as the opportunities for practice and revision in the classroom are limited due to the crowd of the classes and the intensity of the curriculum. Moreover, considering the potential benefits of these systems for teachers and students and the likelihood of popularization of such tools, administrators' attempts to integrate such tools into writing instruction, practice, and assessment will provide great convenience to all users in the future.

Despite favorable findings, the results reported in this study must be interpreted in light of a number of limitations. To illustrate, the time frame of this study was restricted to eight weeks. Hence, a longitudinal study might be designed to explore the issues concerning the implementation of AWE in depth. Moreover, the data were collected from students who were at the intermediate level of English proficiency. Therefore, future studies might need to include other proficiency levels to generalize the findings. Further studies exploring the differences between receiving a combination of AWE and teacher feedback and utilizing either one of these feedback types in isolation might prove important. Moreover, in open and distance learning settings, similar studies can be conducted using AWE. Investigating the effectiveness of a hybrid feedback mode and the views of both students and teachers on combining teacher feedback with AWE feedback might also be explored. Additionally, further studies might explore the relationship between AWE tools and the factors that have great effects on learning such as autonomy and motivation.

Authors' Note: This article was derived from the MA thesis written by the first author under the supervision of the second author.

BIODATA AND CONTACT ADDRESSES OF AUTHORS

Hilal YILDIZ is an English Language Instructor at Ankara University, School of Foreign Languages. She completed her undergraduate education at Hacettepe University, Department of English Language Teaching, and her master's degree at Anadolu University, Department of English Language Teaching. She continues her doctoral studies at Gazi University, Department of English Language Teaching. Her research interests primarily focus on technology-enhanced learning environments, technology-mediated L2 instruction and assessment, computer assisted language learning, and artificial intelligence in language teaching.

Hilal YILDIZ
Department of Foreign Languages, School of Foreign Languages
Address: Ankara University, 06830, Ankara, Turkiye
Phone: +90 0312 485 05 07
E-mail: hllyildiz@ankara.edu.tr

Dr. Safiye Ipek KURU GONEN is an Associate Professor at Anadolu University, Department of English Language Teaching. She has been working as a foreign/second language teacher trainer for over 20 years and is currently teaching various courses related to language teaching methodology and supervising MA and PhD. theses. Her research interests center on pre-service teacher education, reflective practices in teaching, mindfulness practices in foreign language teaching, and technology integration into language teaching. She has published various research papers nationally and internationally.

Safiye Ipek KURU GONEN
Department of English Language Teaching, Faculty of Education
Address: Anadolu University, 26470, Eskisehir, Turkiye
Phone: +90 222 335 05 80/3520
E-mail: ipekkuru@anadolu.edu.tr

REFERENCES

Amrhein, H. R., & Nassaji, H. (2010). Written corrective feedback: What do students and teachers think is right and why?. *Canadian Journal of Applied Linguistics*, *13*(2), 95-127.

Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Journal of Second Language Writing*, 14 (1), 191-205.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, *4*(3), 1-31.

Attali, Y., & Powers, D. (2008). *Effect of immediate feedback and revision on psychometric properties of open-ended GRE Subject Test items: GRE Board Research Report.* Princeton, NJ: ETS.

Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. New York: Routledge.

Bulut, S. E. (2019). *The effects of automated writing evaluation on EFL students' writing achievement and motivation towards writing.* Unpublished Master's Thesis. Istanbul: Istanbul University, Department of English Language Education.

Calvo, R. A., & Ellis, R. A. (2010). Students' conceptions of tutor and automated feedback in professional writing. *Journal of Engineering Education*, 99(4), 427-438.

Chamberland, D. L., Wang, X., & Roessler, B. J. (2008). Photoacoustic tomography of carrageenan-induced arthritis in a rat model. *Journal of biomedical optics*, *13*(1), 011005-011005.

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language testing*, *32*(3), 385-405.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* London: SAGE.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.

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.

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, *34*(1), 18-27.

Cotos, E. (2010). *Automated writing evaluation for non-native speaker English academic writing: The case of IADE and its formative feedback.* Unpublished Doctoral Dissertation. owa: owa State University, Faculty of Applied Linguistics and Technology.

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. New York: Springer.

Deqi, Z. (2005). The process-oriented approach to ESL/EFL writing instruction and research. *Celea Journal*, *28*(5), 66-70.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1)*,* 1–35.

Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford.

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, *10*(2), 121-142.

Elliott, S. (2003). IntellimetricTM: From here to validity. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates.

Ersanli, C. Y., & Yesilel, D. B. A. (2023). The Use of Automated Writing Evaluation Tools to Foster ESL Writing Instruction. In *New Directions in Technology for Writing Instruction* (pp. 193-209). Cham: Springer International Publishing.

Felix, U. (2003). *Language learning online: Towards best practice* (Vol. 3). CRC Press.

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing,* 19(1), 6–23.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be?. *Journal of Second Language Writing*, *10*(3), 161-184.

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. New York. NYU press.

Gencer, o. (2019) *Automated writing evaluation use in an EFL context: From paragraph writing to essay writing.* Unpublished Master's Thesis. Ankara: Ufuk University, Institute of Social Sciences.

Geta, M., & Olango, M. (2016). The Impact of Blended Learning in Developing Students' Writing Skills: Hawassa University in Focus. *African Educational Research Journal*, *4*(2), 49-68.

Gilakjani, A. P. (2017). A review of the literature on the integration of technology into the learning and teaching of English language skills. *International Journal of English Linguistics*, *7*(5), 95-106.

Glaser, B., & Strauss, A., (1967). *The discovery of grounded theory.* Hawthorne, NY: Aldine Publishing Company.

Gorgogeta, M., & Vlachos, K. (2019). A case study to compare the effects of different practices applied when assessing language learners. *Research Papers in Language Teaching and Learning*, *10*(1), 241-264.

Grimes, D. C. (2008). *Middle school use of automated writing evaluation: A multi-site case study.* Unpublished Doctoral dissertation. Irvine: University of California, Department of Information and Computer Science.

Grimes, D. C., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6), 4-43.

Gurel, N. (2010). *An examination of linguistic and sociocultural variables in writing a dissertation among Turkish doctoral students*. Unpublished Doctoral Dissertation. (Publication. New York: State University of New York, Department of Learning and Instruction.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112.

Ho, W. Y. J. (2021). 'I knew that you were there, so I was talking to you': the use of screen-recording videos in online language learning research. *Qualitative Research*, *21*(1), 120-139.

Hockly, N. (2015). Developments in online language learning. *Elt Journal*, *69*(3), 308-313.

Hockly, N. (2019). Automated writing evaluation. *ELT Journal, 73*(1), 82–88.

Hsu, P. S. (2016). Examining current beliefs, practices and barriers about technology integration: A case study. *TechTrends*, *60*(1), 30-40.

Huang, M., Kuang, F., & Ling, Y. (2022). EFL learners' engagement in different activities of blended learning environment. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 1-15.

Jefferson, K., & Radhakrishnan, V. (2020). An Exploration on issues and challenges in teaching writing skills to tertiary-level learners. In S. Kumar and R. Mohanasundram, *Innovations and technologies for soft skill development and learning* (pp. 98-107). Hershey, PA: IGI Global.

Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, *93*, 102302.

Joughin, G. (2009). Assessment, learning and judgement in higher education: A critical review. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1-15). New York, NY: Springer.

Kamasak, R., Sahan, K., & Rose, H. (2021). Academic language-related challenges at an English-medium university. *Journal of English for Academic Purposes*, 49(1), 1-40.

Khoii, R., & Doroudian, A. (2013). Automated scoring of EFL learners' written performance: A torture or a blessing. In *Proceedings of the Edulearn14: 6th International Conference On Education And New Learning Technologies*. Italy: International Education and Training Institution.

Laurillard, D. (2013). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. New York: Routledge.

Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing*, 8(3), 216-237.

Lerner, N. D. (1996). *Teaching and learning in a university writing center*. Unpublished dissertation, Massachusetts: Boston University, Department of Humanities and Social Sciences.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1-18.

Limpo, T., Nunes, A., & Coelho, A. (2020). Introduction to the special issue on technology-based writing instruction: A collection of effective tools. *Journal of Writing Research, 12*(1), 1–7.

Lombana, C. H. (2002). Some issues for the teaching of writing. *Profile: Issues in Teachers´ Professional Development*, *3*(1), 44-51.

Long, R. (2013). A review of ETS's Criterion online writing program for student compositions. *The Language Teacher*, *37*(3), 11-18.

McDonnell, A., Connell, J., Hannif, Z., & Burgess, J. (2013). Voices from higher up the ladder: exploring the team leader role in call centres. *New Technology, Work and Employment*, *28*(2), 145-159.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: SAGE.

Mohsen, M. A., & Abdulaziz, 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.

Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assessment & Evaluation in Higher Education*, *38*(8), 919-940.

Palermo, C., & Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, *54*(1), 255-270.

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.

Parvin, R. H., & Salam, S. F. (2015). The effectiveness of using technology in English language classrooms in government primary schools in Bangladesh. Forum for International Research in Education, 2 (1), 47-59.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, 31(6), 459-470.

Prensky, M. (2001). Digital natives, Digital immigrants. *On the Horizon*, *9*(5), 1-6.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778.

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8-25.

Ruegg, R. (2010). Interlanguage development: The effect of unfocused feedback on L2 writing. *Intercultural communication studies*, *12*(1), 247-254.

Sari, F. M., & Oktaviani, L. (2021). Undergraduate Students' Views on the Use of Online Learning Platform during COVID-19 Pandemic. *Teknosastik*, *19*(1), 41-47.

Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York: Routledge.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, *27*(4), 657-677.

Stone, G. (2017). Implementation of critical literacy for English writing classes in the Thai context. *The New English Teacher*, 11(2), 65-65.

Tai, H. C., Pan, M. Y., & Lee, B. O. (2015). Applying Technological Pedagogical and Content Knowledge (TPACK) model to develop an online English writing course for nursing students. *Nurse education today*, *35*(6), 782-788.

Taskiran, A., Yazici, M., & Erdem Aydin, I. (2022). Contribution of automated feedback to the English writing competence of distance foreign language learners. *E-Learning and Digital Media*, 20427530221139579.

Tawney, J., & Gast, D. (1984). *Single-subject research in special education*. New York: Merrill.

Ustunbas, U., & Cimen, S. (2016). EFL learners' preferences for feedback types for their written products. *The Online Journal of New Horizons in Education*, *6*(4), 68-74.

Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, *10*(2), 1-27.

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.

Ware, P. (2011). Computer-generated feedback on student writing. *TESOL Quarterly*, *45*(4), 769-774.

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22-36.

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

Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. In M.D. Shermis, J.C. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions,* (pp. 36–54). New York: Routledge.

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*(1), 94-109.

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87-125.

Zeng, S. (2020). The potential of online technology for language learning. *English Language Teaching*, *13*(10), 23-37.

Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. Elsevier Inc: Assessing Writing, *43*, 1-14.

Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36(1), 90-102.

Zhang, Z., & Zhang, Y. (2018). Automated writing evaluation system: Tapping its potential for learner engagement. *IEEE Engineering Management Review*, *46*(3), 29-33.

Zupanc, K., & Bosnic, Z. (2015). Advances in the field of automated essay evaluation. *Informatica, 39*(4), 383–395.