Konin Language Studies
Faculty of Humanities and Social Sciences, University of Applied Sciences in Konin, Poland
KSJ 12 (2). 2024. 135-156
http://ksj.konin.edu.pl
doi: 10.30438/ksj.2024.12.2.2

Human versus machine feedback: The contributions of Automated Writing Evaluation (AWE) tools to writing quality and self-regulation of language learners

Mohadeseh Jalili ✉
English Department, Imam Reza International University
https://orcid.org/0009-0009-7881-5713
jalilimohadese@gmail.com

Afsaneh Ghanizadeh
English Department, Imam Reza International University
https://orcid.org/0000-0003-1649-1556
afsanehghanizadeh@gmail.com

Abstract

This study explored how feedback from teachers and Grammarly Premium, an Automated Writing Evaluation (AWE) system, impacts intermediate-level English as Foreign Language (EFL) learners. A mixed-methods approach was implemented with 50 students, split into control and experimental groups. Both groups were assessed using the IELTS General Writing Task 1 (2021) and a self-regulation questionnaire. During the study, the control group received teacher feedback, while the experimental group got personalized Grammarly feedback on grammar, vocabulary, and clarity. After the one-month intervention, both groups completed a new writing task similar to the IELTS General Writing Task 1 (2021) as well as self-regulation questionnaire. Results showed significant improvements in writing skills and self-regulation among the participants of the experimental group. In particular, it was found AWE benefited students in 'effort' and 'self-monitoring' dimensions of self-regulation. The results of qualitative analysis demonstrated that AWE is noticeably beneficial for 'Understanding mistakes and correction'. It was also revealed that students' 'desired for precise issue pinpointing' as the most articulated concern, entailing the fact that AWE might lack transparency and details and needs more clarification.

Keywords: automated writing evaluation (AWE) systems, EFL education, Grammarly, self-regulation, teacher feedback, writing proficiency

1. Introduction

Composing an English essay can be a significant challenge for learners and teachers of English as a Foreign Language (EFL) due to the need for a thorough understanding of language and subject matter. Despite teachers' great efforts, it appears EFL students' writing performance does not improve adequately, especially in areas such as content organization, concept development, and grammar structure (Chen, 2022). EFL instructors must assist students not only in improving their language and communication skills but also in providing feedback using appropriate approaches (Alharbi, 2022). Students' writing abilities improve when they receive written feedback along with practical training. It is suggested that instructors' feedback supports students' engagement and revision techniques (Zhang & Hyland, 2022).

Corrective feedback, encompassing both direct and indirect criticism, is widely used in L2 writing education (Bruton, 2009). Direct feedback (specific comments or assistance to address deficiencies or correct errors) is useful for improving students' shortcomings through direct guidance. Indirect feedback (indicating deficiencies or errors without providing direct correction information) helps develop learners' higher-order thinking skills, such as self-regulated learning (Luo & Liu, 2017).

Recent research has explored focused and unfocused feedback responses (Bitchener & Ferris, 2012). Focused feedback (comment on a single linguistic detail in a written work) offers detailed guidance on a limited number of linguistic aspects to aid students in making consistent progress over time (Bruton, 2009). Learners' writing improves when corrective feedback is concentrated on specific language areas, and this improvement persisted over time (Bitchener & Ferris, 2012). However, there is no clear answer as to which type of feedback is more effective or under what circumstances it benefits students the most. In conclusion, various factors, such as personal qualities and instructional situations, often mediate the impact of feedback on learning (Shute, 2008).

In the modern era, scholars have primarily focused on advancements in computer technology, including automated essay scoring and computer-generated feedback. Writing assessment has undergone three waves, with the latest one emphasizing online and computer software evaluation, allowing for assessments to be conducted both in classrooms and on a larger scale (Hamp-Lyons, 2002). One of the emerging trends concerns Automated Writing Evaluation (AWE), which is a computer tool that provides evaluation feedback in the writing

process in various fields, such as the educational system. This technical program can provide immediate qualitative and quantitative feedback by scoring the text, analyzing the structure of the text, and creating an accurate assessment of the written text (Khoii & Doroudian, 2013). Recently, AWE feedback has been applied in various contexts. For instance, research by Wilson et al. (2014) on the use of AWE feedback for L1 students in grades 4-8 showed improved writing quality across revisions, although the gain was modest and diminished over time. Sung et al. (2016) found that AWE comments significantly enhanced grade six students' summary writing abilities over revisions. These studies suggest that regular exposure to AWE feedback can support the gradual development of young L1 students' writing skills (Sung et al., 2016; Wilson et al., 2014).

Similarly, Kellogg et al. (2010) reported that continuous AWE feedback effectively helped college students reduce errors in mechanics, grammar, and usage on a transfer test compared to no feedback or intermittent feedback. However, producing transfer effects on overall writing quality, including grammar, usage, mechanics, style, organization, and development, remains challenging (Kellogg et al., 2010). In a different study, Bond and Pennebaker (2012) noted improvements in pronoun usage among adult writers with AWE feedback. However, achieving transfer effects on overall writing quality remains difficult (Bond & Pennebaker, 2012).

The debate continues regarding the use of computer-generated feedback versus teacher feedback, or a combination of both, among researchers and educators. Each type of feedback for L2 writing has its advantages and disadvantages, and using either in isolation can either enhance or hinder EFL learners' writing (Cheng, 2017). Consequently, it is often recommended that AWE feedback be integrated with teacher input, as computer-generated corrective feedback may not be the sole factor influencing the proficiency of EFL learners' writing (Li et al., 2015).

The main purpose of this study is to investigate the effect of feedback taken from an AWE tool, along with teacher feedback on students' self-regulation and their writing quality. Furthermore, the second purpose of this study is to investigate the perceptions of English language teachers and students in writing lessons regarding the feedback taken from an AWE tool.

To attain the above purposes, the following research questions were inspected in this study:

RQ1: Does the AWE tool influence EFL learners' self -regulation?
RQ2: Does the AWE tool influence the quality of EFL learners' writing?
RQ3: How might the AWE tool influence EFL learners' self-regulation?
RQ4: How might the AWE tool influence quality of EFL learners' writing?
RQ5: What are students' perceptions toward AWE?

2. Literature review

Technological improvements have transformed the feedback process in formative writing assessments, allowing for quick and economical assessment of student writing (Cotos, 2014). Automated feedback in second language (L2) writing for English has become more important as a supplement to teacher input, providing advantages including customization and efficiency (Warschauer & Ware, 2006). Initially created for assessing essays in standardized examinations, these systems have been utilized in English as a Foreign Language (EFL) environments to offer instant feedback, particularly beneficial for large classes and for identifying grammatical mistakes (Chen & Cheng, 2008; Wilson et al., 2014).

Automated feedback systems provide consistent evaluations and support learning by identifying errors for correction and providing personalized suggestions (Hoang & Kunnan, 2016). These technologies help reduce human-rater biases and improve learning results by providing personalized feedback adapted to specific student requirements (Leontjev, 2014). Advancements in technology have allowed Automated Writing Evaluation (AWE) systems to offer detailed, interactive feedback beyond just scoring, using automated essay scoring (AES) engines to assess and rate written material based on various factors (Hockly, 2019; Shermis & Burstein, 2003; Deane, 2013).

Automated Writing Evaluation (AWE) systems evaluate texts and offer constructive feedback to enhance writing proficiency (Allen et al., 2016). They employ user-friendly interfaces and resources like visual organizers and online dictionaries to help students (Franzke et al., 2005; Ware, 2014). As AWE platforms become more common in education, other systems have been developed, such as online platforms like Write To Learn and modifications for existing software, providing a range of features for students and teachers (Dikli, 2006; Cahill & Evanini, 2020).

The purpose of developing Automated Writing Evaluation (AWE) was to decrease the need for manual effort in assessment writing and to guarantee unbiased scoring that is not affected by human errors (Stevenson & Phakiti, 2019; Wang et al., 2020). The systems provide benefits such as impartial analysis and prompt feedback, which might enhance student motivation and writing skills (Weigle, 2013; Grimes & Warschauer, 2010).

Educators can utilize AWE to concentrate on advanced writing components and tailor feedback to individual student needs (Cotos, 2014; Jacovina & McNamara, 2017). AWE tools also assist with L2 writing and have been modified for languages other than English to tackle the distinct difficulties encountered by L1 and L2 authors (Warschauer & Ware, 2006; Shermis, 2020).

Although AWE systems have benefits, they are criticized for their accuracy in assessing writing, with some academics doubting their usefulness compared to

human raters (McCurry, 2010; Stevenson & Phakiti, 2014). Research has investigated the influence of AWE feedback on student writing, showing varied outcomes on their effectiveness in enhancing writing abilities and emphasizing the necessity of integrating automated feedback with teacher guidance for the best learning results (Graham et al., 2015; Stevenson, 2016).

Challenges are the danger of bias in scoring, constraints in offering feedback on content and structure, and the possibility of students manipulating the systems to achieve better grades (Deane, 2013; Weigle, 2013; Powers et al., 2002). Although there are challenges, AWE systems are considered a beneficial complement to conventional feedback approaches, enhancing rather than substituting human evaluation (Attali, 2013).

Overall, AWE systems provide notable advantages in efficiency, customization, and the ability to improve learning, but their constraints and the importance of integrating them into educational practices are apparent. Continual development and assessment of these systems are essential for optimizing their efficacy in enhancing writing instruction.

Educational research has lately turned its attention towards metacognition, self-regulation, and self-regulated learning, reassessing these ideas in light of their historical and theoretical foundations. This reassessment seeks to comprehend their interconnectedness and predict future research paths. Despite the vast amount of literature on knowledge, motivation, and education, a continuing problem is the need for clear definitions for these categories. This uncertainty requires comparing contemporary views and their philosophical or psychological roots.

Fox and Riconscente (2008) thoroughly examined the historical factors that impact contemporary perspectives on metacognition and self-regulation. The principles were attributed to the core theories of James, Piaget, and Vygotsky, which offer a complete framework for comprehending metacognition and self-regulation. These theories emphasize the interrelated relationship between these constructs, showing that although they are separate, they are interconnected regarding growth and function. James concentrated on the utilitarian aspect of habit and the reflective approach to self-awareness. Piaget focused on peer-level interaction and the cultivation of metacognitive and self-regulatory skills via interacting with the environment. In contrast, Vygotsky emphasized the importance of formal education and the cultural advancement of higher cognitive functions through language.

The assessment highlights that each theorist's method has its advantages and drawbacks. James' introspective approach is informative but restricts the research to a particular channel of consciousness, possibly neglecting the broader range of metacognitive and self-regulatory development. Piaget's retrogressive method from action to comprehension could overlook or misinterpret some facets

of metacognition and self-regulation because of the inferential processes involved. Vygotsky's emphasis on language-based internalization provides useful perspectives on the impact of cultural tools on development. However, it may face obstacles due to cultural diversity and the hierarchical impact of culture on growth.

Dinsmore et al. (2008) examined current empirical research on these ideas, pinpointing meaningful relationships and fundamental conceptual bases that bring together metacognition, self-regulation, and self-regulated learning. They advocate for using accurate language and acknowledging conceptual differences to guarantee clarity and efficiency in research. This alignment emphasizes the importance of monitoring and altering behavior in both dimensions while also warning against seeing them as interchangeable words.

Maggioni and Parkinson (2008) conducted empirical research on epistemic cognition, examining how instructors' views and teaching methods influence students' learning and growth. Their research emphasizes the intricate relationship among instructors' epistemic views, the learning environment, and teaching efficiency, emphasizing the significance of incorporating epistemic cognition into educational methods.

Schunk (2008), Lajoie (2008), and Kaplan (2008) provide significant suggestions for the next research and practical uses in education, emphasizing the promotion of metacognition, self-regulation, and self-regulated learning at various developmental levels. Enhancing these processes in students can improve lifelong learning, especially in technology-integrated environments (Lehmann et al., 2014; Kizilcec et al., 2017).

This literature review examines the intricacies of metacognition, self-regulation, and self-regulated learning from historical, theoretical, and empirical viewpoints. The text emphasizes the significant work of influential theorists, the difficulties posed by unclear concepts, and the necessity of accurate terminology and educational integration for successful learning and research.

3. Method

3.1. Participants

3.1.1. Participants of the Qualitative Phase

Fifty EFL students were included in the present research. The individuals were engaged in the study of English at an intermediate level, dedicating around four hours per week to their educational pursuits. The study took place inside the confines of a semi-private language institute in Mashhad, a city in the northeastern

Iran. The selection of the institution was based on factors related to credibility and practicality. The participants aged 15 to 17 years were allocated into two groups of equal size: an experimental group and a control group. The research included 25 EFL learners in the experimental group, who were provided with AWE feedback over four weeks, covering four distinct themes.

Additionally, a control group of 25 EFL learners got input from their instructor. Prior to the commencement of the study, both control and experimental groups were administered a writing proficiency test and completed a questionnaire as a pretest. This was done to ensure that the participants in both groups were homogenous regarding their language competency and self-regulation level.

3.1.2. Participants of the Qualitative Phase

During the second phase of study, eight volunteer participants from the experimental group were requested take part in a semi-structured interview. This data collection aimed to examine the potential impact of AWE feedback on their writing and self-regulation, identify any drawbacks associated with the feedback, and elicit recommendations for improving its effectiveness and utility.

3.2. Instruments

3.2.1. The test of the International English Language Testing System (IELTS)

The International English Language Testing System (IELTS) is widely acknowledged and used as a standardized assessment tool for evaluating people's language skills in order to facilitate their pursuit of educational, professional, and travel endeavors. The IELTS assessment framework examines applicants' proficiency in four fundamental language skills: Listening, Reading, Writing, and Speaking. This study utilized the writing component of the International English Language Testing System (IELTS) examination and its development in response to two unique feedback approaches

This research focuses mainly on Task 1 of the writing section in the General Training module. In accordance with this, the study employed the genuine writing assignments derived from IELTS General Writing Task 1 for the year 2021. The writing tasks were assessed by two qualified examiners and the man score of the two raters was considered as the obtained score. The correlation between the two sets of scores was deliberated as the magnitude of inter-rater reliability.

3.2.2. Self-Regulation Questionnaire

The self-regulation trait (SRT) questionnaire was designed by O'Neil and Herl (1998). It consists of 32 Likert-scale questions measuring two broad dimensions of metacognition and motivation. Each dimension comprises two sub-scales. Meta-cognition covers the constructs of planning and self-monitoring, and motivation contains effort and self-efficacy. The following table (Table 1) depicts the subscales of the SRT.

Table 1 The Subscales of SRT Along with the Corresponding Descriptions

Factor		Definition	Items
Metacognition	*Planning*	The extent to which one has an assigned or self-directed goal and a plan to achieve the goal.	1-5-9-13-17-21-25-29
	Self-monitoring	The extent to which one needs a self-checking mechanism to monitor goal achievement.	2-6-10-14-18-22-26-30
Motivation	*Effort*	The extent to which one works hard on a task.	3-7-11-15-19-23-27-31
	Self-efficacy	The extent to which one has confidence in being able to accomplish a particular task.	4-8-12-16-20-24-28-32

According to Herl et al (1999), the reliability and validity of the scale have been verified in multiple studies.

3.2.3. Grammarly

Grammarly is an innovative AWE technology, functioning as a virtual writing helper. It utilizes sophisticated algorithms to analyze written text for grammatical inaccuracies, syntactical problems, punctuation faults, and other related difficulties. The tool provides immediate recommendations for enhancing sentence structure, language selection, and general coherence, ultimately envisaging to enhance the quality of written writing.

There are two significant iterations of Grammarly: the free version and the premium edition. The complimentary edition offers fundamental grammar and spell-checks functionalities, rendering it a readily available initial resource for persons aiming to enhance their writing precision. On the other hand, the premium edition provides an extensive range of sophisticated functionalities, including context-specific suggestions, word enrichments, style refinements, and plagiarism identification.

One noteworthy characteristic of Grammarly is its capacity to produce PDF outputs. This feature is accessible in both the complimentary and paid editions. In order to generate a PDF output of the written work, it is recommended to either create the content directly or copy and paste it into the Grammarly editor.

In this study, Grammarly played a pivotal role in enhancing the writing skills of the experimental group. The premium version was utilized to provide insightful feedback that went beyond the basics of grammar and spelling. With its sophisticated analysis, Grammarly helped refine sentence structures, enhance vocabulary choices, and ensure coherence. This strategic employment of the premium version was presumed to contribute to develop writing proficiency within the experimental group.

3.3. Procedure

This study employed both quantitative and qualitative methods.

3.3.1. Quantitative Phase

This section presents a comprehensive examination of the methodologies utilized in the research. The main objective of the study was to examine the influence AWE feedback systems on the writing proficiency of young EFL learners at an intermediate level. Furthermore, it aimed to evaluate the impact of AWE on four distinct dimensions of self-regulation. The advanced version of Grammarly, a sophisticated software designed to assist in writing, was employed to enhance the study endeavors.

In order to commence the investigation, a group of 50 students with the intermediate proficiency level in the English language was selected. Having made the required arrangements with the institute's administrative committee, the participants were randomly categorized into two discrete groups: a control group (N=25) and an experimental group (N=25). The experimental group received feedback from the Grammarly software, while the control group was provided with feedback exclusively from their professors. Considerable work was exerted to examine various existing AWE feedback systems, both the localized and the international ones, opting that the accessibility and comprehensibility arising from Grammarly would make it advantageous over other competing programs.

During the experimental phase, the proficiency of both groups in writing and their self-regulation were evaluated using a pretest. The evaluation of participants' writing skills in the study involved the administration of the IELTS General Writing Task 1 for the year 2021. Each participant was given a time limit of 20 minutes to construct a concise paragraph on a specific topic that had been predetermined. The initial writing exercise functioned as a critical baseline by which subsequent development could be assessed. Concurrently, all participants were administered a self-regulation questionnaire to obtain insights into effectively regulating the writing process.

Having assured the homogeneity of the two groups concerning the two dependent variables, the teacher (one of the researchers) engaged the two groups in consistent writing tasks over the whole semester. Having received each writing tasks, the teacher provided the participants of the two groups with the writing prompt. The students in the control group were provided with direct feedback on their writing by the teacher, whereas the experimental group received input from the AWE tool.

The methodology utilized by the experimental group demonstrated a degree of variability. Every student in the experimental group was provided with an individualized PDF document using the Grammarly Premium service. The texts mentioned above offered recommendations for improving grammatical precision, choosing suitable lexicon, structuring sentences effectively, and guaranteeing general coherence in written compositions.

Following the conclusion of the educational session for about one month, participants were given a writing assignment that closely resembled the IELTS General Writing Task 1 for the year 2021. The task required participants to compose a paragraph within a specified time restriction of 20 minutes. The main aim of this study was to evaluate the possible influence of feedback provided to participants, either by instructors or through the use of Grammarly, on the development of writing skills.

Concurrently, the administration of the self-regulation questionnaire took place for both groups as part of a post-intervention evaluation to examine alterations in self-regulatory mechanisms after instructional interventions.

In order to maintain standard criteria and ensure objectivity, a methodology of inter-rater assessment was utilized. The methodology employed in this study entailed the assessment of written compositions by two independent raters. The researchers utilized *Bailey and Brown's Grading* framework for their analysis. It evaluated different facets of writing, encompassing grammatical accuracy, lexical diversity, coherence, and organizational arrangement. The selection of this particular methodology was motivated by reducing potential biases and establishing a consistent approach across the scoring process. The inter-rater reliability was computed, which was quite acceptable (0.91). Table 2 presents an overview of the scale components, along with the corresponding weights allocated to each component.

Table 2 Bailey and Brown's Grading Scale

	Component	Points
A.	Organization	5
B.	Logical development of ideas	5
C.	Structure	3.5
D.	Mechanics	3
E.	Style and quality of expression	3.5

3.3.2. Qualitative Phase

In the qualitative phase, a semi-structured interview – with a set of pre-specified albeit dynamic interview questions – was administered to eight students from the experimental group who were chosen randomly after four weeks. The primary objective of this approach was to elicit and comprehend their perspectives and perceptions regarding the feedback provided by the AWE system. The inquiry encompassed diverse dimensions to understand their experiences and viewpoints comprehensively. The obtained responses were systematically collected and subsequently subjected to rigorous analysis.

In this regard, three pivotal questions were posed to the participants:

1) Utility of Grammarly's Feedback: Participants were queried about their perceptions concerning the utility of Grammarly's feedback. Specifically, they were prompted to elucidate the areas of their writing in which they found the feedback beneficial. This inquiry sought to reveal the practical impacts of the AWE system's guidance on their writing process and their metecognition.
2) Challenges and Obstacles: A significant aspect of the interview revolved around participants' insights into the challenges and obstacles they encountered when attempting to comprehend the feedback provided by Grammarly. This line of questioning aimed to uncover potential difficulties and identify any aspects of the feedback that may have been less transparent or easily misunderstood.
3) Suggestions for Enhancement: Lastly, participants were invited to provide constructive suggestions regarding how Grammarly's feedback could be optimized for the benefit of language learners. This open-ended query aimed to extract innovative ideas from the participants, potentially offering insights into how AWE systems could be tailored to more effectively aid language learners in refining their writing skills.

Through this qualitative approach, the study sought to unlace valuable insights into the multifaceted perspectives of participants, shedding light on the nuances of their experiences with AWE feedback. These insights, duly extracted and analyzed, would contribute to a comprehensive evaluation of the potential of AWE systems in enhancing language learners' writing proficiency.

4. Results

The present study examined the role of AWE feedback in EFL students' self-regulation and writing achievement. Feedback on writing is the independent variable

(treatment), and writing proficiency and self-regulation were considered as the dependent variables, being presumed to be influenced by AWE.

4.1. Quantitative Phase

In order to assess the potential disparity in writing proficiency levels between the control and experimental groups, a statistical analysis was conducted using an independent samples *t*-test. There were some degrees of disparity in the mean scores of writing proficiency and self-regulation between the control and experimental groups.

In order to determine the statistical significance of the observed differences, independent samples *t*-tests were conducted. The results verified homogeneity of the two groups in writing proficiency as well as their self-regulation level: writing proficiency (*t*=.54, *p*=.58), level of self-regulation (*t*= -.17, *p*=.86).

In order to assess the potential influence of AWE on writing achievement, independent samples *t*-test was run. The descriptive results of writing level in the two groups are summarized in Table 3.

Table 3 Descriptive Statistics of Language Achievement Across Control and Experimental Groups

	groups	N	Mean	Std. Deviation	Std. Error Mean
Writing score 2	1.00	25	15.55	1.04	.23
	2.00	25	17.26	1.40	.31

According to the data presented in the table, it can be observed that the average scores obtained in the writing test by the experimental group are higher in comparison to those of the control group. In order to determine the statistical significance of the observed difference, an independent samples *t*-test was conducted. The results, displayed in Table 4, indicated a considerable disparity in writing achievement between the two groups, as evidenced by the statistical analysis (*t*= -4.29, *p*=.00).

Table 4 Independent Samples T-Test Showing the Results of Posttest on Writing Achievement

		Levene's Test for Equality of Variances		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
		F	Sig.					
Writing Score 2	Equal variances assumed	1.70	.20	-4.29	48	.00	-1.70	.39
	Equal variances not assumed			-4.32	47.03	.00	-1.70	.39

Another independent samples *t*-test was performed to see if the control and experimental groups differed in self-regulation and its four variables (effort, self-efficacy,

self-monitoring, and planning). Table 5 shows the descriptive statistics of self-regulations and subscales in the two groups. The Table reveals that control and experimental groups have different mean self-regulation and subscale scores, in favor of experimental group.

Table 5 Descriptive Statistics of Self-Regulations and its Subscales Across Control and Experimental Groups

	groups	N	Mean	Std. Deviation	Std. Error Mean
Planning 2	1.00	25	22.21	6.45	1.48
	2.00	25	26.25	3.38	.75
Self-monitoring 2	1.00	25	21.57	5.84	1.34
	2.00	25	26.69	3.83	.85
Effort 2	1.00	25	22.10	5.49	1.26
	2.00	25	26.30	2.65	.59
Self-efficacy 2	1.00	25	21.94	5.81	1.33
	2.00	25	25.90	4.35	.97
Self-regulation 2	1.00	25	87.84	21.68	4.97
	2.00	25	101.90	12.46	2.78

Independent samples *t*-tests were performed to see if the differences are significant. Table 6 shows the two groups differ statistically in total self-regulation (t= -2.46, p=. 02) and all four subscales: Planning, Self-monitoring, Effort, and Self-efficacy. As shown, self-monitoring and effort were more influenced.

Table 6 Independent Samples T-Test Showing the Results of Posttest Self-Regulation and its Four Subfactors

		Levene's Test for Equality of Variances		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
		F	Sig.					
Planning 2	Equal variances assumed	7.60	.009	-2.46	48	.019	-4.03	1.63
	Equal variances not assumed			-2.42	46.88	.022	-4.03	1.66
Self-monitoring 2	Equal variances assumed	6.21	.040	-3.30	48	.007	-5.12	1.57
	Equal variances not assumed			-3.27	47.84	.010	-5.16	1.59
Effort 2	Equal variances assumed	12.28	.001	-3.68	48	.004	-5.57	1.51
	Equal variances not assumed			-3.63	84.18	.001	-5.57	1.53
Self-efficacy 2	Equal variances assumed	2.09	.156	-2.41	48	.001	-3.95	1.63
	Equal variances not assumed			-2.39	43.31	.022	-3.95	1.65
Self-regulation 2	Equal variances assumed	8.14	.007	-2.49	48	.017	-14.05	5.62
	Equal variances not assumed			-2.46	45.42	.020	-14.05	5.7

4.2. Qualitative Phase

4.2.1. The model of the influence of AWE feedback and students' perceptions

The last three research questions are concerned with how AWE tools influence EFL learners' self-regulation, writing quality, and students' perceptions toward AWE

feedback. Three questions were asked from eight language learners (selected randomly) from the experimental group who received feedback from AWE, and the interview data were collected and coded. All data were carefully studied and analyzed. After an in-depth review and evaluation of the data, the researchers categorized the identified codes into two fundamental categories.: 1) Contribution of AWE for L2 writing, 2) students' perceptions toward AWE feedback, each of them was classified into different subcategories.

The result of the first and second research questions is presented in Figure 1, generated by MAXQDA 2020 software, which exhibits the potential influences of AWE feedback.

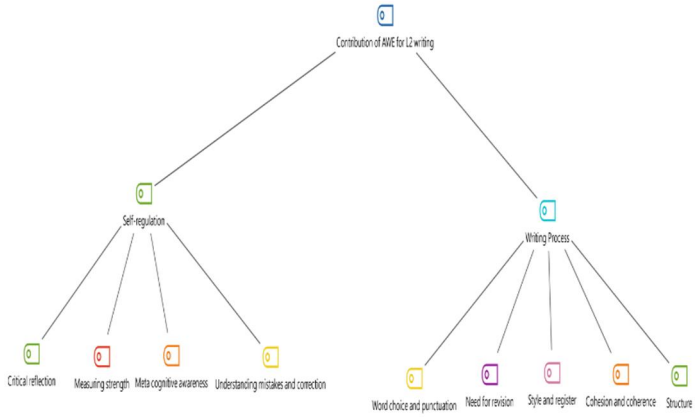

Figure 1 The model of the contribution of AWE for L2 writing

The third and fourth research questions examined the effect of machine feedback on the quality of language learners' writing and their self-regulation process in writing. In the interview that was held with eight language learners from the experimental group, the given answers were analyzed, and two categories of writing process and self-regulation were extracted from the researcher's inquiries about how the machine feedback helped them and what effect it had on their writing skills. Each of these categories includes several subcategories. The self-regulation code includes four subcategories: Understanding mistakes and correction, Metacognitive awareness, Measuring strength, and Critical reflection, and the writing process code includes five subcategories: Style and register, Need for revision, Structure, Word choice and punctuation, Cohesion, and coherence.

The result of fifth research question is presented in Figure 2, generated by MAXQDA 2020 software, and reveals all the concerns that reflect students' perceptions toward AWE feedback.

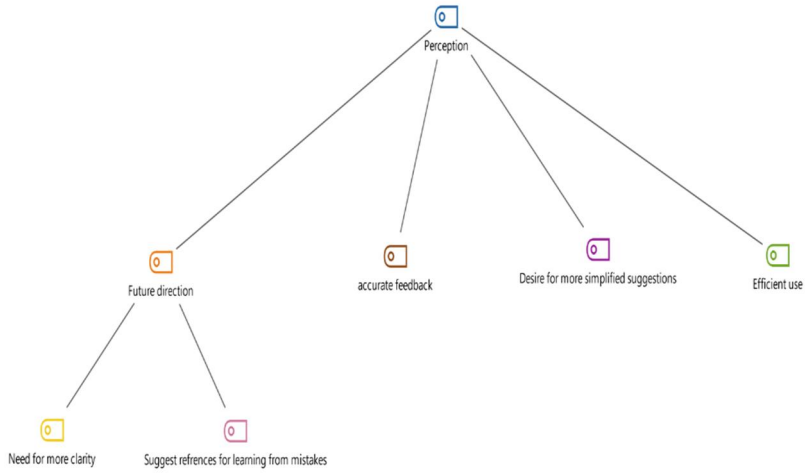

Figure 2 The model of students' perceptions toward AWE feedback

The fifth research question examined students' perceptions toward AWE feedback. In the interview that was held with eight language learners from the experimental group, the given answers were analyzed, and four categories of future direction, accurate feedback, desire for more simplified suggestions and efficient were extracted from the researcher's inquiries.

4.2.2. Frequency of the Codes

Following the examination of the data using the codes shown in Figures 1 and 2, the frequencies of each code were recorded and organized in the subsequent tabulation (Tables 7 & 8).

Table 7 Frequency of the Identified Codes of the Contribution of AWE for L2 Writing

Parent code	Code	Cod. seg. (all documents)	% Cod. seg. (all documents)
Contribution of AWE for L2 writing	Self-regulation	7	5.26
Contribution of AWE for L2 writing	Writing process	9	6.77
Self-regulation	Understanding mistakes and correction	10	7.52
Self-regulation	Metacognitive awareness	8	6.02
Self-regulation	Measuring strength	1	0.75
Self-regulation	Critical reflection	5	3.76
Writing process	Style and register	7	5.26
Writing process	Need for revision	1	0.75
Writing process	Structure	6	4.51
Writing process	Word choice and punctuation	7	5.26
Writing process	Cohesion and coherence	2	1.50

Due to the abundance of codes, the use of AWE feedback has had the most significant effect on 'understanding mistakes and correction', and it has also led to 'self-regulation in writing skills' via its influence on various attributes, as designated in Table 7.

Concerning students' perceptions toward AWE, the most frequent codes in language learners' perceptions were 'future direction', and the 'need for clarity and efficient use' (Table 8).

Table 8 Frequency of the Identified Codes of the Students' Perceptions Toward AWE Feedback

Parent code	Code	Cod. seg. (all documents)	% Cod. seg. (all documents)
Perception	Future direction	9	6.77
Perception	Accurate feedback	6	4.51
Perception	Desire for more simplified suggestions	6	4.51
Perception	Efficient use	7	5.26
Future direction	Need for more clarity	7	5.26
Future direction	Suggest references for learning from mistakes	4	3.01

5. Discussion and conclusions

This study examined how AWE tools like Grammarly and instructor feedback affect EFL students' self-regulation and writing skills. The research showed effectiveness of the intervention. EFL students' writing proficiency improved statistically after the intervention. Regarding self-regulation, the experimental benefitted from AWE, particularly in self-monitoring and effort components of self-regulation.

Recent research highlights the effectiveness of Automated Writing Evaluation (AWE) systems in improving EFL students' writing skills. This article summarizes the results of a recent study that showed a considerable improvement in writing skills among EFL learners who used Automated Writing Evaluation (AWE) in comparison to a wide range of existing literature (Cotos, 2014; Wilson et al., 2014; Shermis & Burstein, 2003). The significant enhancement in grammatical correctness and mechanical features of writing, demonstrated by the experimental group's better performance, supports the effectiveness of AWE as a valuable addition to traditional teaching methods (Nunes et al., 2022; Han & Sari, 2022).

However, incorporating AWE into EFL courses comes with obstacles. Feedback precision critiques, particularly about content organization, require continuous improvements in AWE systems (Zhang & Cai, 2019). Additionally, focusing on fixing faults at the sentence level, while helpful for correcting grammatical mistakes, may unintentionally neglect the development of more advanced writing skills like coherence and argumentation (Chen & Cheng, 2008).

To enhance the educational effectiveness of AWE, it is crucial to utilize strategic teaching methods. Utilizing a variety of writing prompts and combining AWE feedback with teacher input may provide a well-rounded learning environment, enhancing student engagement and writing skills (Parra G & Calero S, 2019; Stevenson, 2016). An integrated paradigm combines automated feedback for immediacy and efficiency with human review for depth and contextuality, enhancing learners' writing experiences.

In conclusion, the results of the latest study support the growing agreement on AWE's ability to improve EFL writing results significantly. To fully realize this potential, it is necessary to solve existing constraints through technical improvements and sophisticated teaching methods, ensuring the balanced development of basic and advanced writing abilities.

This discussion summarizes research findings on improving self-regulation in EFL students using Automated Writing Evaluation (AWE) within a theoretical framework covering classic and modern views on metacognition and self-regulated learning. The observed data, showing significant enhancements in self-regulation and its related aspects (such as planning, self-monitoring, effort, and self-efficacy) in individuals exposed to AWE, supports the theories proposed by Fox and Riconscente (2008), Lehmann et al. (2014), and Kizilcec et al. (2017). The enhancements confirm the critical significance of AWE in promoting self-regulation, highlighting the connection between cognitive activities enabled by technology interventions and the metacognitive frameworks outlined by James, Piaget, and Vygotsky.

The qualitative findings emphasize how AWE influences the ability to identify and fix errors, underscoring the crucial role of adaptive feedback in promoting metacognitive awareness and regulatory techniques. This supports Dinsmore et al.'s (2008) focus on the detailed definition of self-regulation in educational research and highlights the capacity of Automated Writing Evaluation (AWE) to enhance learners' self-regulatory skills.

Furthermore, the significant impacts on self-monitoring and effort support the critical discussion about improving learner independence through metacognitive and self-regulatory support in technology-enhanced learning settings. This study confirms the theoretical foundations of self-regulation presented by past and present experts. It also advances the discussion by providing empirical evidence of the effectiveness of AWE in enhancing self-regulatory skills in EFL learners.

This study enhances the educational discussion on combining technological capabilities with teaching methods to develop advanced self-regulation and metacognitive skills. It suggests further research on improving feedback systems in Automated Writing Evaluation (AWE) tools to support independent, continuous learning paths.

In brief, the qualitative results of this study support previous research on AWE tools and self-regulation, writing talent, and student perspectives. The resources

provide valuable perspectives on how AWE technologies affect learners, emphasizing their value in EFL writing teaching.

The findings of the present study offered valuable insights regarding the influence of feedback provided by AWE technologies, such as Grammarly, on the self-regulation (in particular, self-monitoring and effort). The qualitative data provided additional support for the quantitative findings. The impact of AWE feedback on students' word choice, punctuation, and self-regulation in writing skills was noteworthy, as stated by the students. Nonetheless, there was an inclination towards more meticulously identifying issues in the input provided for AWE. This in turn entails more studies on the effectiveness of AWE and the ways it can be modified to present more comprehensible corrective feedback with greater clarification of the nature and quality of the feedback. Whether sentence-level rather than text-level feedback would enhance AWE's capability in supporting learners improve their written tasks could be studied in future research. All in all, as there are so many factors which might shape feedback appropriateness and learners' preferences in receiving them, it is probable to be too abrupt to make certain verdicts about whether human feedback or computer feedback is more effective.

References

Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing, 52*, 100610.

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 316-329). The Guilford Press.

Attali, Y. (2013). Validity and reliability of automated essay scoring. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation* (pp. 181-198). Routledge.

Beishuizen, J., & Steffens, K. (2011). A conceptual framework for research on self-regulated learning. In G. Dettori & D. Persico (Eds.), *Self-regulated learning in technology enhanced learning environments* (pp. 1-19). Brill.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Bruton, A. (2009). Designing research into the effects of grammar correction in L2 writing: Not so straightforward. *Journal of Second Language Writing, 18*(2), 136-140.

Cahill, A., & Evanini, K. (2020). Natural language processing for writing and speaking. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring* (pp. 69-92). Chapman and Hall/CRC.

Chen, A.-H. (2022). The effects of writing strategy instruction on EFL learners' writing development. *English Language Teaching, 15*(3), 29-37.

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*(3), 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*, 18-27.

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

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

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment, 5*(1), 1-15.

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*, 391-409.

Ferris, D., & 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.

Fox, E., & Riconscente, M. (2008). Metacognition and self-regulation in James, Piaget, and Vygotsky. *Educational Psychology Review, 20*, 373-389.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary Street®: Computer support for comprehension and writing. *Journal of Educational Computing Research, 33*(1), 53-80.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523-547.

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment, 8*(6), 1-43.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing, 8*(1), 5-16.

Han, T., & Sari, E. (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2025327

Hibert, A. I. (2019). Metacognitive processes and self-regulation in the use of automated writing evaluation programs. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming learning with meaningful technologies: 14th European Conference on Technology Enhanced Learning* (pp. 655-658). Springer International Publishing.

Hoang, G. T. L., & Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of MY Access. *Language Assessment Quarterly, 13*(4), 359-376.

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

Jacovina, M. E., & McNamara, D. S. (2017). *Intelligent tutoring systems for literacy: Existing technologies and continuing challenges* [Technical report]. Arizona State University.

Kaplan, A. (2008). Clarifying metacognition, self-regulation, and self-regulated learning: What's the purpose? *Educational Psychology Review, 20*, 477-484.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173-196.

Khoii, R., & Doroudian, A. (2013). Automated scoring of EFL learners' written performance: A torture or a blessing. In *Conference proceedings: ICT for language learning* (p. 367). Libreriauniversitaria.it Edizioni.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education, 104*, 18-33.

Lajoie, S. P. (2008). Metacognition, self regulation, and self-regulated learning: A rose by any other name? *Educational Psychology Review, 20*(4), 469-475.

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior, 32*, 313-323.

Leontjev, D. (2014). The effect of automated adaptive corrective feedback: L2 English questions. *APPLES: Journal of Applied Language Studies, 8*(2), 43-66.

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.

Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics, 7*(4), 197-215.

Maggioni, L., & Parkinson, M. M. (2008). The role of teacher epistemic cognition, epistemic beliefs, and calibration in instruction. *Educational Psychology Review, 20*, 445-461.

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing, 15*(2), 118-129.

Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning, 38*(2), 599-620.

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.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103-134.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20*, 463-467.

Sherafati, N., & Mahmoudi Largani, F. (2023). The potentiality of computer-based feedback in fostering EFL learners' writing performance, self-regulation ability, and self-efficacy beliefs. *Journal of Computers in Education, 10*(1), 27-55.

Shermis, M. D. (2020). International applications of automated scoring. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring* (pp. 113-132). Chapman and Hall/CRC.

Shermis, M. D., & Burstein, J. C. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition, 42*, 1-16.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51-65.

Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (2nd ed., pp. 125-142). Cambridge University Press.

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education, 95*, 1-18.

Tang, J., & Wu, Y. A. (2017). Impact of AWE rubrics and automated assessment on EFL writing instruction. *International Journal of Computer-Assisted Language Learning and Teaching, 7*(2), 58-74.

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

Ware, P. (2014). Feedback for adolescent writers in the English classroom. *Writing & Pedagogy, 6*(2), 223-249.

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. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 36-54). Routledge.

Wilson, J., Olinghouse, N. G., & Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities: A Contemporary Journal, 12*(1), 93-118.

Zhang, W., & Cai, W. (2019). Research on English writing feedback based on online automatic evaluation and reform. *Education Research, 3*, 102-103.

Zhang, Z. V., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51*, 100586.

Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review, 2*, 173-201.