International Journal of Information Technology and Language Studies (IJITLS)

Evaluating the effectiveness of using Automated Written Corrective Feedback on EFL University Students' Writing

Bilel Elmotri[1] and Chokri Smaoui[2]

1Department of Languages and Translation, Applied College, Northern Border University, Saudi Arabia

2 Department of English Language, Faculty of Letters and Humanities, University of Sfax, Tunisia

bilelelmotri@gmail.com; c.samaoui@squ.edu.om

Article Information

Article type: Article

Article history:

Received: January 15, 2024
Revised: February 13, 2024
Accepted: April 01, 2024

Keywords:

EFL
AWE
AWCF
Specific feedback
Generic feedback

Abstract

This study investigates how Automated Written Corrective Feedback (AWCF) can facilitate English as a Foreign Language (EFL) writing teaching and evaluation in Arab schools and universities. The study addresses a significant research gap in the Computer-Assisted Language Learning (CALL) literature by examining the impact of Automated Essay Scoring (AES) systems' feedback on EFL University students' writing. A mixed-methods approach was used to conduct a longitudinal controlled experiment at Northern Border University (NBU) in Saudi Arabia. The study involved questionnaires with 120 students and interviews with 27 students and 4 teachers. An exploratory factor analysis was employed to analyze perceptions of AWCF. The data analyzed the impact of ease of use, clarity, usefulness, EFL learners' perceptions, and feedback explicitness on AWCF use. The findings proved that EFL students and teachers found AWE systems easy to use and useful for corrective feedback. Furthermore, students preferred using Grammarly's specific AWCF strategies over Writing Mentor's generic feedback, influencing EFL students' use of AWCF tools. Therefore, developers should consider diverse factors influencing EFL learners' preferences and views of AWCF to create useful, clear, and desirable tools.

1. Introduction

1.1 Background

The effectiveness of second-language (L2) acquisition in linguistic accuracy has been a contentious issue among L2 practitioners for nearly two decades (Goins, 2015). This research offers practical advice for teachers implementing WCF in classrooms, highlighting the need for further research on AWE's role in ESL/EFL writing teaching (Bitchener & Ferris, 2012). CF is also known as negative feedback or negative evidence in SLA. On the other hand, CF, defined as input offered to demonstrate to learners instances of improper language use (Long, 1996), is widely regarded to be advantageous since it may facilitate the process of noticing, a necessary component of L2 acquisition (Gass & Selinker, 1983; Schmidt, 1990, 1992). Earlier studies concentrated on oral corrective feedback, but a growing body of current research demonstrates the potential benefits of WCF to L2 acquisition in text-based computer-mediated communication (Li, Link, & Hegelheimer, 2015).

In comparison, L2 response research focuses on WCF and has yet to reach a consensus on its effects. While some have argued that WCF has no place in L2 writing teaching and should be abandoned (Truscott, 1996, 1999, 2007), others have recognized its value in L2 writing instruction and moved on to study effective WCF delivery methods (Ferris, 2010). Ferris and Roberts (2001) discovered that explicit feedback, with and without error type-specific corrections, was more effective at assisting students in self-editing than providing no feedback. Additionally, offering WCF to L2 writers can reduce teachers' burden since they may need to repeat feedback (Ferris, Liu, Sinha, & Senna 2013).

It might also be challenging because the classification and labeling of student errors by type may be inconsistent, increasing the "chance of instructor error" (Ferris, 2010, p. 199). Effective corrective feedback is crucial in ESL writing schools to improve linguistic correctness and concept communication for L2 writers. Research suggests that AWE corrective feedback may be beneficial for ESL writers, and comparing it to essential characteristics of effective WCF presented by Hartshorn et al. (2010) demonstrates the potential for AWE corrective feedback to be used in ESL writing classrooms and its potential benefits for ESL writers.

1.2. Purpose of the research

This research will try to create the framework for a better understanding of the possible role of AWE/AES systems for ELLs, specifically whether they help to teach writing in an EFL situation where English is rarely used outside the classroom. It will discuss both AES, which is defined as "the provision of automated scores derived from mathematical models based on organizational, syntactic, and mechanical aspects of writing," and automated corrective feedback, or "computer tools for writing assistance rather than writing assessment," because these two applications of AWE/ AES have distinct usage considerations (as cited in Weigle, 2013, p. 36). This study investigates EFL learners' preferences and perceptions of AWE systems and various AWCF strategies, including automated graduated CF. It explores practical learners' perceptions of AWE's ease of use, desirable, clear, and valuable strategies, and the impact of feedback explicitness on students' use. A mixed-method design is employed to understand better EFL learners' use, preferences, and perceptions.

1.3. Research questions

This study attempts to answer the following questions:

1. How do EFL students perceive the ease of use of AWE tools?
2. What are their preferred AWCF strategies regarding desirability, clarity, and usefulness?
3. How does feedback explicitness (Generic or Specific) influence EFL students' use of AWCF tools?

2. Literature review

2.1. Automated Written Corrective Feedback

The increasing interest in English language learning (ELL) skills, mainly writing, necessitates the development of sophisticated software and applications to evaluate these essays globally.NNSs of English need Automated Writing Evaluation (AWE) more than NSs because they represent the largest population and market (Weigle, 2013). Automated Essay Scoring (AES) and AWCF are English language instruction tools that provide automated corrective feedback to students. Developed in 1966, AES has evolved through systems like IntelliMetric 13, Knowledge Analysis Technologies TM,2, and e-rater. AWE tools like MYaccess diagnose, write tests, and support students' feedback. However, their effectiveness in teaching and assessment remains a question (Li et al., 2015; Weigle, 2010, 2013).

Another problem is that AES is still controversial regarding reliability for decision-making at prestigious institutions in University admissions and immigration (Weigle, 2013). On the contrary, employing AWCF tools in classrooms is perceived as more efficient in supporting the teacher's feedback (Warschauer & Grimes, 2008; Weigle, 2013). Private companies create tests for immigration, tertiary education, and admission using internationally recognized tests like IELTS, TOEFL, CET, and the University of California's "Entry Level Writing Requirement," assessing language proficiency for critical decisions (Weigle, 2013). TOEFL tests use human raters and automated systems like MyBestTM, saving time, cost, and effort but having disadvantages in high-stakes tests. While not yet recognized by some institutions, AES is valid and reliable in essay scoring, similar to human scorers (ETS, 2020). Weigle (2013) argues that automated scoring machines cannot accurately assess essay quality, and the algorithms used in scoring are ambiguous, leading to de-professionalization, narrowing of constructs, and

potential consequences. So, the transparency of these systems should be revisited to ensure their reliability (Cheville, 2004; Weigle, 2013).

Classroom writing evaluation involves summative and formative measures, with teachers and learners collaborating. Automated feedback enhances formative assessment, while AES is effective for extensive tests but unsuitable for summative assessment due to teachers' preference for deeper evaluations (Weigle, 2010, 2013). Researchers suggest further study using AWE tools in instruction and assessment, rejecting the direct correlation between human and automated scoring (Li et al., 2015). The effectiveness of AWE tools in assessing communicative and social-linguistic features is under-researched, but they may assist students in revising, editing, and improving their writing behavior, potentially directing learners from content focus to form (CCCC, 2014; Cheville, 2004; Li et al., 2015). CCCC (2014) opposes machine-scored writing assessment, stating that it can distort writing as a complex, context-rich interaction, leading writers to focus more on structure and grammar than content (as cited in CCCC, 2014, para. 6).

CALL researchers utilized an interactionist approach to enhance students' linguistic capacity and provide immediate feedback for essay editing using software features like word processors and automated writing programs (Chapelle, 2001, 2003, 2010; Chapelle & Erik, 2016). (Garrett, 1987) suggests five types of feedback in CALL: correct answer, error detection, error messages, linguistic problem identification, and intelligent feedback. The first type provides correct answers, the second identifies errors, the third anticipates incorrect answers, and the fourth uses Natural Language Processing. Cotos (2010) suggests that feedback methods, ICTs, and interactionist approaches can improve second language skills and academic performance, while corpus-based theory fails to address linguistic deficiencies in CALL environments. AES programs provide immediate feedback, multiple scoring submissions, and advanced issues like argumentation and voice, reducing anxiety and enabling educators to tackle complex writing tasks (Weigle, 2013). Xi (2010) provides a guide for evaluating automated feedback programs, although it acknowledges that few of these inquiries can be applied to general feedback. Hartshorn et al. (2010) state that daily writing of WCF significantly improved post-test accuracy scores in a treatment group compared to a control group receiving frequent feedback.

2.1.1. Usefulness of Automated Feedback to Students

Many studies are exploring the effectiveness of automated feedback in improving students' composition in their classes. Chen and Cheng (2008) analyzed the use of My Access! in three Taiwanese college classes, where students received both scores and explicit feedback. The teachers were given one hour to prepare and decide how to use the product. The results showed that students preferred automated feedback for formative assessment but found it useless and required more effort. The teachers also reported technical issues and poor usage. The study suggests mechanized feedback would suit learners with lower language skills and essential composing abilities. AWCF systems are increasingly used in L2 classes to provide students with written feedback, reducing instructors' workload and allowing learners to modify and edit their work (Li et al., 2015).

However, it remains unsure how much WCF in any structure boosts second language processing (Bitchener & Ferris, 2012; Ranalli, 2018). Proof from L1 and L2 classroom research indicates that mechanized criticism can improve the nature of L2 understudy composing across drafts of similar content (Stevenson & Phakiti, 2014). Nonetheless, some features distinguish AWCF from the instructor's WCF, which conceivably decreases its utilization in L2 teaching. These incorporate:

(1) Distinctions in supportive data across error types are resolved more by technological tools than educational contemplations.

(2) Inaccuracies and mistakes should be standard in any AWE system's flagging errors, impacting students' eagerness and capacity to utilize the obtained feedback.

(3) The one-size-fits-all nature of AWCF takes practically zero records of individual differences (Ranalli, 2018, p. 653).

2.1.2. Explicitness of AWCF

Research in the WCF literature reveals that explicitness is a critical distinguishing feature between instructors' WCF and AWCF, regardless of direct or indirect feedback. Bitchener and Ferris (2012) suggest that direct feedback identifies

mistakes and provides solutions, while indirect feedback highlights errors and encourages learners to correct them. AWCF systems classify errors but do not offer specific explanations based on error information. Ranalli, Link, and Chukharev-Hudilainen (2017) suggested that Specific feedback reminds a learner of an error, presents a solution, or prescribes specific activities to correct it, unlike generic feedback, which offers broad solutions. Ranalli (2018) claims that no study has researched how explicitness affects students' use of AWCF.

Recent AWE systems lack consideration for learners' second language skills, first language, composing abilities, or educational foundations, leading to a one-size-fits-all approach. WCF studies suggest that user differences affect the effectiveness of feedback for L2 learners. Bitchener and Ferris (2012) called for a WCF study that examines the impact of L2 capability mixed with different factors like error type and explicitness of feedback. Until this time, most investigations in L2 classes on the utilization of AWE instruments (Chen & Cheng, 2008; Dikli & Bleyle, 2014; Lavolette et al., 2015; Lee & Hegelheimer, 2012) have just examined learners at a one-course level. The previously mentioned three factors are:

Feedback explicitness.

The requirement for accuracy assessment.

Individual differences among learners.

They may impact how learners deal with AWCF and deserve a thorough examination. There are three dimensions of AWCF use by L2 learners that these factors might affect.

2.1.3. ETS's WM®

The WM system is a Google Docs Add-On that provides instructional writing support to students, particularly in post-secondary settings. It focuses on four main writing subconstructs: reliability of statements, topic advancement, consistency, and editing. The app is available in Beta version and aims to improve student writing (Burstein et al., 2018; Madnani et al., 2018). The app supports writers internationally, building on past AWE (Attali & Burstein, 2006; Burstein, Chodorow, & Leacock, 2004; Burstein et al., 2018; Shermis & Burstein, 2013). Typical measures for the automated evaluation of large-scale, high-stakes scores (Atttali & Burstein, 2006; Shermis et al., 2015) and online instruction have been supported by AWE systems (Burstein et al., 2004; Burstein et al., 2018). The WM app targets college students and hard-won writers seeking written assistance. It focuses on English conventions but not beyond them. Institutional subscriptions for applications like Criterion® and Turnitin's Revision Assistant 6 restrict accessibility (Burstein et al., 2004).

The WM app, available on Google Docs, provides effective writing assistance to college students and writers, focusing on English conventions and allowing users to choose feedback types and review them for improvement (Burstein et al., 2018; Burstein, Elliot, & Molloy, 2016). It includes a departure perception survey adapted from the System Usability Survey. The app also provides feedback reports, which can be shared with instructors. User event log data is collected for research purposes, including time marks, features, and document revisions (Brooke, 1996). WM offers a 3-question survey for users to assess writing confidence, app usage, and English proficiency, and a departure perception survey adapted from the System Usability Survey.

2.1.4. Grammarly application

Grammarly, founded in 2009 by Maz Lytvyn and Alex Shevchenko, is the world's most precise English grammar checker, offering over 250 grammatical controls and contextual spell check and boasting over four million registered users by 2014, ranking at the fastest-rising Deloitte company index (Cavaleri & Dianati, 2016). Grammarly is an AI software that improves text and document interaction efficiency, corrects grammar errors, and enhances writing comprehension. It reviews sound, ensures reliability, and searches for plagiarism. Grammarly Premium offers over 400 tests, grammar checks, vocabulary enhancement, plagiarism detection, and bug fixes for professionals and students (Grammarly, 2021).

Grammarly was chosen because prior studies suggested it is simple (Li et al., 2015). Grammarly is a free academic writing tool for non-native English writers, offering grammar, pointing, orthography, sentence structure, and style support. It offers a freemium model and a premium subscription for $139.95 per year, licensed for K-12 and universities, with over 600 universities and companies (Brown, 2018). The study utilized Grammarly's freemium version, a free grammar checker,

for EFL students to improve their writing skills. Grammarly is a general-purpose AWE with a freemium model, similar to WM.

3. Methodology

The study explores the impact of Automated Written Corrective Feedback (AWCF) on EFL students' compositions, utilizing a sociocultural and mixed approach, incorporating quantitative and qualitative data to understand students' perceptions, interests, and experiences with AWCF strategies.

3.1 Context and Participants

The study examines the impact of AWE and AWCF strategies on students' EFL writing performance at NBU's preparatory year deanship. The research focuses on engineering students' writing composition to improve their performance. The students were divided into three sections: PAA (34 students), PAB (37 students), and PAC (33 students). The experimental group used Grammarly and WM Applications in computer lab writing classes, while the control group was taught through classical writing instruction. Questionnaire participants were 103 students, 66 from the experimental group and 37 from the control group. Interviews were conducted with 27 students, 8 from the control and 19 from the experimental group. This was done to validate the questionnaire and sample data in investigating the interpretation, interests, and experience of L2 students with automated CF strategies. Teachers' interviews were administered with 4 EFL instructors. 1 taught the control group, 2 taught the experimental group, and 1 another taught the experimental groups in the second semester. Interview data were used to establish the underlying explanations for the expectations and preferences reported by EFL students and teachers. The participants in each research instrument are described in the following table (Table 1):

Cohort	**Students' Questionnaires**	**Students' interviews**	**Teachers' interviews**
Control	37	8	1 (taught the experimental group)
Experimental	66	19	3 (taught the control group)
Total	103	27	4

Table 1. Number of participants in each research instrument.

3.2. Data Collection Instruments

3.2.1 Questionnaires

The study used online questionnaires designed on Google Forms to assess the use of AWCF in teaching English writing to students. The questionnaire was sent via WhatsApp and email, ensuring anonymity and considering factors like attitudes and fear of embarrassment. 103 students responded out of 120 in all groups. The research focused on the correlation between AWCF systems in teaching English writing and student perceptions, focusing on linguistic challenges and automated feedback strategies. The questionnaire was designed using Daoud (2000), a well-organized and comprehensive model, and included an optional departure perception survey adapted from the System Usability Survey (SUS). The questionnaire was translated into Arabic and piloted with ten NBU students. The reliability of the questions was tested using Principal Component Analysis (PCA) results. The study also examined the validity of the KMO and Bartlett indices and total variance explained by the table of commonalities. The Bartlett Test showed a KMO of 0.765, indicating the analysis is appropriate (See Table 2).

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,765
Bartlett's Test of Sphericity	Approx. Chi-Square	3476,951
	df	1035
	Sig.	,000

Table 2. KMO index and Bartlett test.

The reliability analysis of our scales was conducted using Cronbach's Alpha, which is considered acceptable if it is greater than or equal to 0.7. All items selected from 5 dimensions showed reliable results, with Cronbach's alpha values of 0.936 for 17 items, 0.832 for 6 items, 0.931 for 7 items, and 0.876 for 10 items (see Table 3).

	Reliability Statistics	
	Cronbach's Alpha	N of Items
Q1	.832	6
Q2	.936	17
Q3	.931	7
Q4	.931	6
Q5	.876	10

Table 3. Cronbach's alpha test.

The researcher administered questionnaires to students during the first semester of 2019-2020 during COVID-19 pandemic which hindered the process of data collection. It was more difficult especially for conducting interviews with the participants during the lockdown, which also delayed the data analysis. 103 out of 104 participants responded to the questionnaires, which were merged into a comprehensive one, coded, and uploaded to the SPSS application. The data collected from the questionnaires provided sample data for the current research and helped establish a relationship between statistics and interpretations.

3.2.2. Interviews

This study utilized semi-structured interviews to assess the usefulness of AWCF in English writing instruction. The interviews were conducted using closed and open questions, with a rating scale used to evaluate teachers' and students' satisfaction with the writing process. Audio recordings were used to capture and code responses, allowing participants to express their ideas in a natural flow of speech. The interviews focused on linguistic issues, familiarity with AWE software, and WCF preferences. Open-ended questions were used to gather ideas and assess learners' overall opinions of WM and Grammarly. The study involved 27 students randomly selected from experimental and control groups and 4 EFL preparatory year teachers who speak English as a second language and taught the sample groups. The interviews were in-person, via email, Facebook, and WhatsApp. The information gathered can be analyzed qualitatively and quantitatively.

3.3. Data Analysis

The study used data collection and analysis tools like SPSS, Excel, and Smart-PLS. Quantitative data was analyzed using questionnaires, while qualitative data was evaluated using categorizations. The data was validated through three phases: data organization, assumption checking, and quantitative and qualitative findings integration. This combination of data analysis ensures accurate results and validates the collected data. This study focuses on the validity of AWCF programs, WM and Grammarly, for the English 101C Writing Assessment, using WM/ Grammarly. The research uses a mixed methods

concurrent transformative technique to analyze both quantitative and qualitative data. The analysis includes two models: the evaluation method of AWE identified by Ranalli et al. (2017), which is built on Kanes' "Argument-Based Validation for AWE as formative assessment," and the Technology Acceptance Model (TAM).

The evaluation method of AWE involves constructing an interpretation/use argument (IUA) and evaluating its comprehensiveness, coherence, and reliability. The Technology Acceptance Model suggests that perceived usefulness and ease of use are two primary factors influencing personal acceptance and application of new technology. This study investigates the acceptance and utilization of AWCF programs WM and Grammarly among college students. The core factors, perceived value and ease of use remain the most critical drivers of technology acceptance. The study uses questionnaires, writing samples, and interview data to examine the students and teachers of the AWCF systems used. Descriptive statistics, frequency statistics, and skewness tests analyzed the data. SPSS and Excel descriptive and nonparametric tests were used to determine the distribution of responses across five scale values. The significance test helps maintain or reject null hypotheses raised by software. All statistical tests yield significant results based on questionnaire data, which are essential for assessing the use of AWCF in writing instruction.

3.3.1. Exploratory factor analysis to validate relations between dimensions

The measurement model was evaluated for convergent validity, the agreement of multiple items measuring the same concepts. The criteria for assessing convergent validity include factor loading (>0.7), composite reliability (>0.7), and average variance extracted (>0.5). The researchers conducted a factor analysis to verify each measurement scale's internal consistency and convergent validity. The reliability of the measurement items was investigated, and items with a quality of representation below 0.7 were deleted. The internal consistency of the measurement scales was verified using Cronbach alpha and Rho Dillon-Goldstein coefficient values. The average variance extracted (AVE) of the latent variables was verified using Smart-PLS software, and all values exceeded the threshold of 0.5. The convergent validity of the research variables was confirmed, except for item q 4.4, which was almost equal to 0.7.

3.3.2. Integrating quantitative and qualitative data

The study combines quantitative data from questionnaires with qualitative research to verify findings and determine their robustness. Qualitative data, such as frequencies, percentages, ratings, complement numbers, and scores, are categorized according to the assessment system. Data from audio recordings of interviews helps explain the symbolic worlds of the people and phenomena being studied. The researchers listened to tapes and transcribed the conversations. Open-ended responses from questionnaires, interviews, and fieldwork observations are used to interpret and validate quantitative data. Triangulation is used to support data reliability and consistency with previous studies. Results are compared to other researchers' data to confirm the study's validity.

4. Results

4.1 EFL students' perception of AWE tools and their preferred AWCF strategies

According to the framework, the two pillars of EFL students' perception of AWE tools and their preferred AWCF strategies are 1/ ease of use and 2/ usefulness and clarity. They will measure features that make up each category by associating, correlating, and comparing variables.

4.1.1 Ease of Use of AWE tools

The study validates the hypothesis that most EFL students and teachers find the AWE systems (Grammarly/WM) easy to use and useful. The framework inference suggests that these systems facilitate autonomous learning and improve writing processes. The questionnaire results showed that most students feel confident navigating WM, with a mean of 4 out of 5 respondents agreeing. The study also found that only 4.1% strongly disagreed with the AWE system.

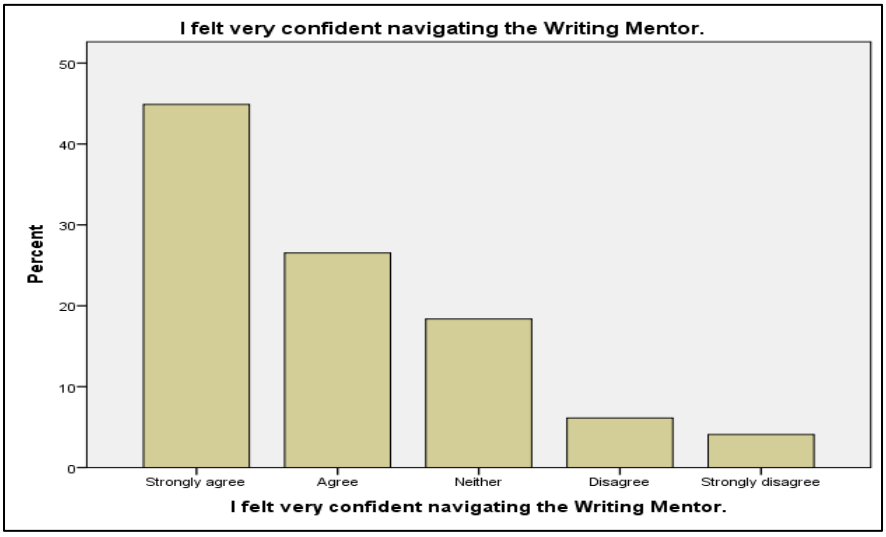

Figure 1. Bar Chart of item 9 from Question 8.

Students found the WM's functions well-integrated, facilitating system use and benefiting from well-organized functions. However, excessive use of IT and linguistic terms can confuse students, leading to time-consuming definitions. They prefer easy access to all functions (see Table 4).

I found the various functions in the WM were well-integrated.					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Strongly agree	30	29.1	30.6	30.6
	Agree	36	35.0	36.7	67.3
	Neither	20	19.4	20.4	87.8
	Disagree	7	6.8	7.1	94.9
	Strongly disagree	5	4.9	5.1	100.0
	Total	98	95.1	100.0	
Missing	System	5	4.9		
Total		103	100.0		

Table 4. Frequencies and percentages of item 5 from question 8.

They are also aware of the importance of the application, which should be used by more users who want to enhance their skills. Most would imagine that most people would quickly learn to use an application like this one (see Figure 2).

I would imagine that most people would learn to use an application like this one very quickly.

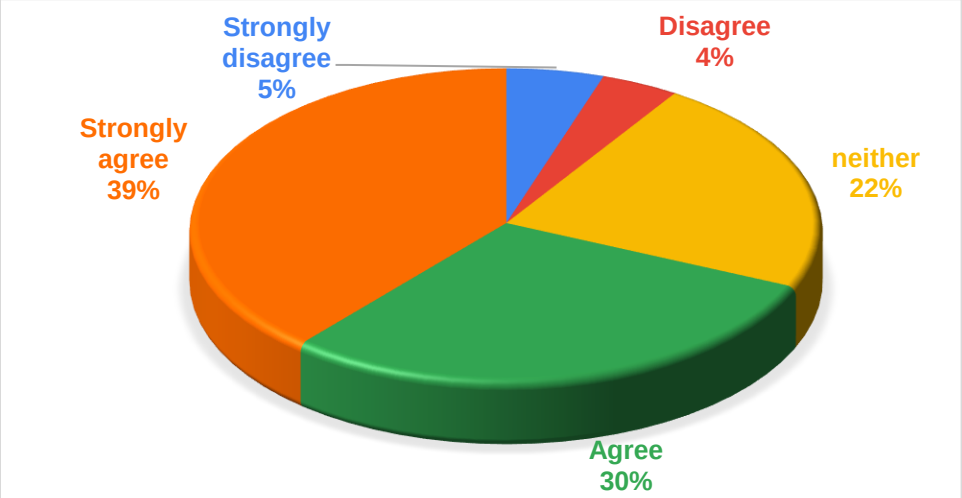

Figure 2. Pie chart of item 7 from question 8.

Results have also shown that 60% of students thought the WM was easy to navigate, and 14% disagreed (see Figure 3). Therefore, students accessed the system easily and used its functions comfortably. Within these questionnaire findings, it is identified that most students found the AWE application easy to use.

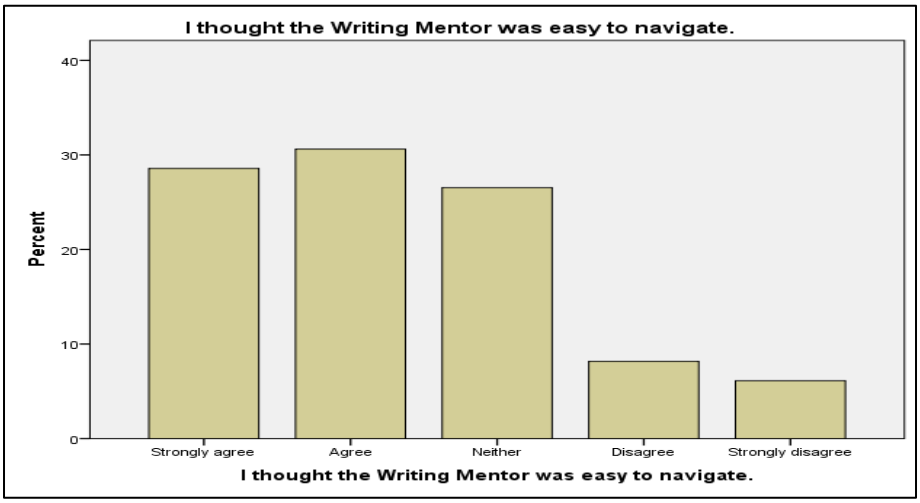

Figure 3. Bar chart of item 3 from question 8.

Regarding the TAM's second critical factor, perceived ease of use, the Grammarly application was also used to compare and confirm this factor. 39 of the 103 students (38%) always and often used Grammarly. 28 participants sometimes used the system. Only 16 students rarely used Grammarly. Twenty students never used Grammarly since they belonged to the Control group and were not introduced to the systems before (see Table 5). The mean of Grammarly's frequency is the highest among other AWE systems (2.98), which is also higher than Microsoft Word spellcheckers (2.47) and WM (2.73) (see Table 6).

Grammarly					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Always	15	14.6	14.6	14.6
	Often	24	23.3	23.3	37.9
	Sometimes	28	27.2	27.2	65.0
	Rarely	16	15.5	15.5	80.6
	Never	20	19.4	19.4	100.0
	Total	103	100.0	100.0	

Table 5. Grammarly's frequency of use among the students.

Descriptive Statistics					
	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Grammarly	103	1	5	2.98	1.328
Microsoft word/ Google docs Spellcheckers	103	1	5	2.74	1.275
WM	103	1	5	2.73	1.308
My writing Lab	103	1	5	2.50	1.378
My Access	103	1	5	2.47	1.440
Turnitin	103	1	5	2.45	1.370
Criterion	103	1	5	2.42	1.280
Valid N (listwise)	103				

Table 6. The mean AWE systems frequency of use among the students.

Students struggled with Grammarly due to a lack of Arabic translations for some features and content. Grammarly's automatic recognition of British or American English text may not be as sensitive, and the user's inability to select text directly affects its usability.

Students' interviews revealed that the AWE systems used in the controlled experiment are user-friendly and accessible, as noted by St 4 from the experimental group: "It's an easy application to use; it gives me more confidence to write English. It helps me to submit a correct paragraph."

4.1.2 Clarity and usefulness of AWCF strategies

The feedback design, including feature detection, wording, and presentation, is crucial for clarity and usefulness. The framework inference, "ramification systematically," assumes that students perceive new systems positively. Question 8 in the questionnaire showed that most students agree with the usefulness of the system items, with means ranging from 3 to 4 out of 5. This ensures that the use of these new systems is beneficial for learning. 70% of students think they would like

to use the WM frequently, with a mean of 4 (see Figure 4). This explains the preference of students to use the application in the future.

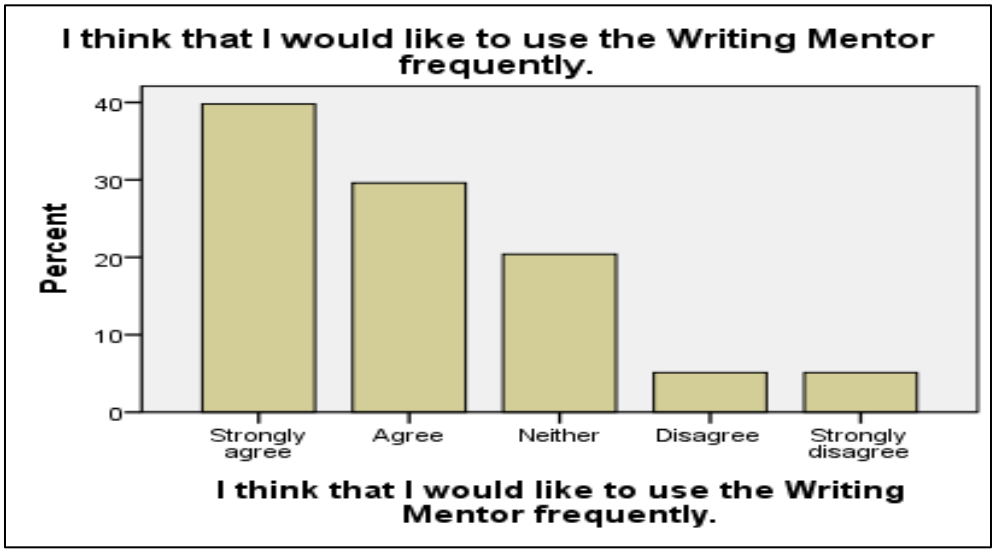

Figure 4. Bar Chart of item 1 from Question 8.

WM is an instructional model focusing on systematic writing processes, involving planning, goal-setting, writing procedures, reviewing, revising, and editing. It promotes a Model-Practice-Reflect instructional cycle, viewing writing and reading as complementary tasks. Iterative feedback and formative assessment enhance students' skills and promote learning.

Results from the open-ended questionnaire question 9, ("I wish WM had ….") had some similarities with findings from questions 7 and 8. Many students found that the systems were Clear and Useful. For example, student 60 answered: "I wish the University had officially adopted WM."

The study's hypotheses were supported by student interviews, which revealed their preference for a clearer and more useful correction strategy in question 7. Students found some of the research's AWCF of the used systems to be clearer and more useful, as per Student 6: "The applications are highly beneficial as they help me remember information and prevent future errors."

Teachers believe AWCF is useful, except for the control group instructor, T3, who was not introduced to WM and Grammarly. They noticed that AWCF improved students' writing skills and reduced errors. However, it has disadvantages like distraction, lack of creativity, and license issues due to Google Docs access and Grammarly Premium.

4.1.2.1 Feedback design

The feedback design, focusing on clarity and usefulness, will be studied through user, teacher, and researcher perspectives, involving error feature detection and presentation aspects.

Feedback wording:

Most students desire a computer to assist them with grammar errors, correct them, and provide hints when needed.

I eat lunch at a restaurant yesterday.

↓

Computer: You have a grammar error. Do you know what type is it?

You: No.

Computer: Is it a simple past, a present perfect error?

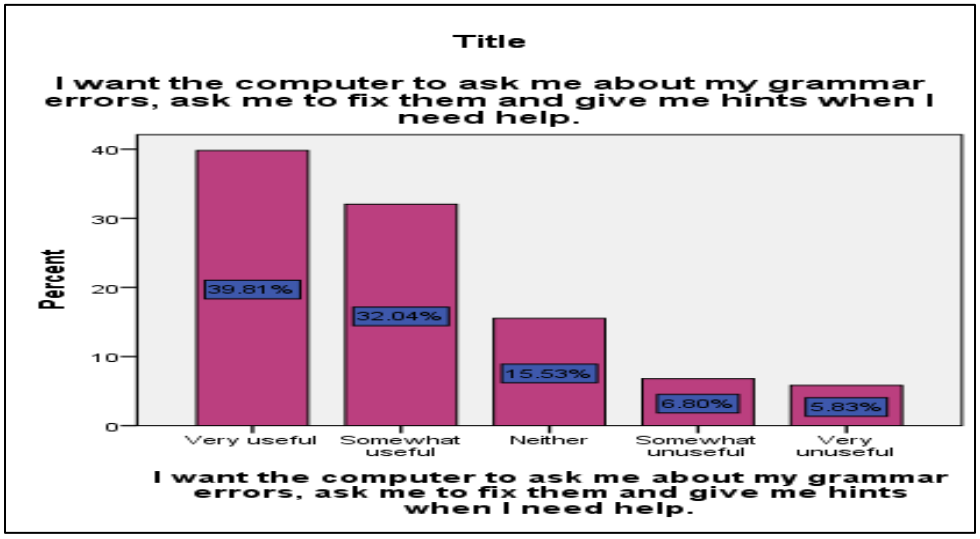

Figure 5. Bar Char of item 6 from Question 7.

62% of students find gradual guidance through application feedback helpful, preferring it over immediate correction, especially in exams or real-world situations requiring mental effort. Using WM in lab sessions resulted in fewer successful corrections on written errors, especially in distinguishing general and specific feedback. Graduated feedback required more mental effort and time. Student interviews showed a preference for strategy 6. Student 18 cites, "The optimal strategy is 6 as it identifies the type of error and provides comprehensive information about it for corrective action."

Highlighting of errors:

The analysis of questionnaire question 7 will focus on highlighting grammar errors. Respondents perceived the feedback strategy as useful by 37%, somewhat useful by 24%, and somewhat and very unuseful by 15%. The mean score was 3.77 (see Table 7), indicating a high percentage but low compared to other strategies.

I want the computer only to highlight my grammar errors.					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very useful	38	36.9	36.9	36.9
	Somewhat useful	25	24.3	24.3	61.2
	Neither	25	24.3	24.3	85.4
	Somewhat unuseful	8	7.8	7.8	93.2
	Very unuseful	7	6.8	6.8	100.0
	Total	103	100.0	100.0	

Table 7. Frequencies and percentages of item 3 from question 7.

Descriptive Statistics					
	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
When I make an error, I want to be corrected immediately.	103	1	5	4.06	1.211
I want the computer to highlight my grammar errors and tell me the correct forms.	103	1	5	3.95	1.199
I want the computer to ask me about my grammar errors, ask me to fix them and give me hints when I need help.	103	1	5	3.93	1.165
I want the computer to highlight my grammar errors and tell me their types.	103	1	5	3.86	1.189
When I make an error, I want to be corrected when I ask for correction.	103	1	5	3.81	1.229
I want the computer to only highlight my grammar errors.	103	1	5	3.77	1.222
Valid N (listwise)	103				

Table 8. **Descriptive statistics of question 7 items in descending order of means.**

Students used WM to write paragraphs, submit, edit, and submit drafts, saving time and effort for the instructor. However, many students struggled to understand the feedback provided by WM, as they often didn't know how to correct mistakes or inappropriate language. The AWE provided no solutions for fixing issues, leading students to seek assistance from the researcher and their teachers. They preferred to be corrected immediately and explicity (see table 8). The teacher's workload decreased when drafts were submitted, but he focused on helping students make sense of the AWE's feedback before submission.

Concerning Item 4 (I want the computer to highlight my grammar errors and tell me the correct forms). For example,

I eat lunch at a restaurant yesterday.

↓

Correct verb: ate.

Most students prefer the AWE system to give the correction immediately, along with highlighting the errors. Indeed, 45% of them answered very useful, and 23% responded by somewhat useful, which makes 68% believe that this corrective feedback strategy is generally useful. However, only 11 % of participants found this strategy unuseful (see Figure 6). According to the ascending order of question 7 items, item 4 is the second-highest preferred strategy by the participants by a mean of 3.95 out of 5 (see table 8).

I want the computer to highlight my grammar errors and tell me the correct forms

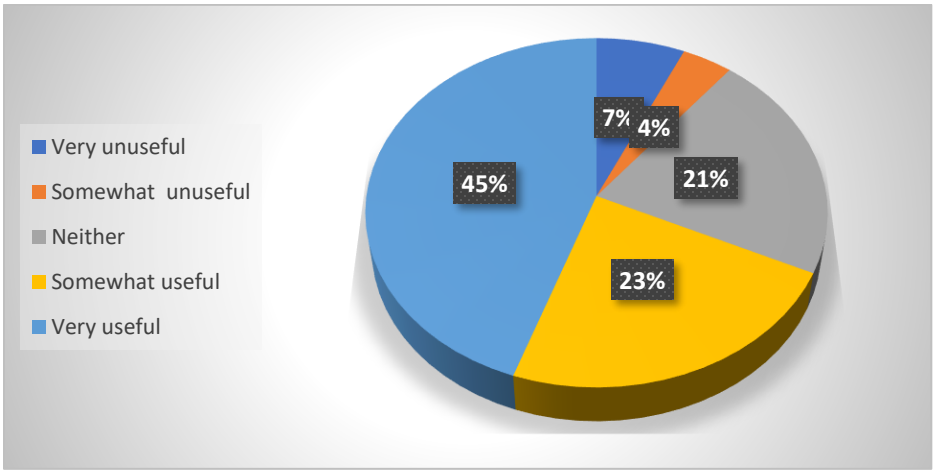

Figure 6. Pie chart of item 4 from question 7.

During the controlled experiment, students found Grammarly less complicated and easier to use than WM. They asked fewer questions about feedback and found Grammarly to be more user-friendly. The Grammarly writing interface is less crowded than WM, with a layout that includes a text writing box, error box, color, shape-coded error marks, and a roll-over box system. Grammarly offers a large writing area on the left side of the screen, with everything marked as an error highlighted and corrections provided in the margin to the right. Students were occasionally confused by the design of WM but not Grammarly. Findings from students' interviews validated the preference for highlighting strategies. Some students contended that having the right and direct correction, like in Strategy 4, is important. For example, student 24 compares strategies 4 and 6: "Strategy 4 is very useful because it highlights my error and gives me the correct answer."

In addition to direct correction, some students opt for immediate feedback like strategy 1 (1. When I make an error, I want to be corrected immediately).

Student 27 observes, "Strategies 1 and 4 are the best as they allow for immediate error correction and faster information sharing."

The last highlighting technique as an AWCF strategy is in item 5, which is like the following:

I want the computer to highlight my grammar errors and tell me their types. For example,

I eat lunch at a restaurant yesterday.

↓

This verb does not match the timeframe of the sentence. You should use a verb in the Simple past.

This strategy was seen as very useful by 38%, less than item 4, and somewhat useful by 30% more than the percentage found in item 4. However, the number of respondents who perceived it as useful is the same as for item 4 (68%). Nevertheless, 13% of participants considered it unuseful (see Figure 7). This proves that students agree that knowing the type of error is important and useful but less useful than providing immediate correction and interaction with students, like in strategy 6. Item 5 received the lowest mean (3.86 out of 5) among the highlighting strategies like items 3 and 4 (see Table 8 above).

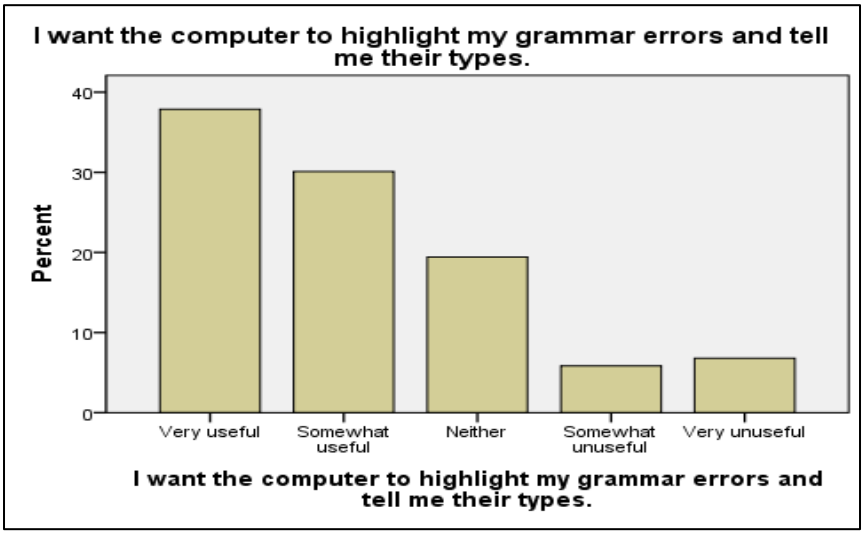

Figure 7. Bar chart of item 5 from question 7.

The students' questionnaires revealed that highlighting grammar errors and their types is less effective than other highlighting strategies, with most interviewees not focusing on this strategy, and only one student mentioned its importance. Student 21 compares strategy 5 to strategy 6. He declared, "The best strategy is 6 because it detects my errors and suggests answers."

Other aspects of feedback preference (color and timing):

Students' responses to question 7 from the questionnaire will focus on timing preference, with item 1 indicating immediate error correction. After making an error, immediate correction is considered very useful by 52%, somewhat useful by 22 %, and 13% of participants (see Figure 8). This item is ranked first in the descending table of descriptive statistics of question 7 items in descending order of means by a mean of 4.06 and a standard deviation of 1.211 (see table 9 above). This explains the high desirability of immediate correction of errors by students. The timing of the corrective feedback is essential for students because they need to know their errors before they continue writing until the end of the assignment.

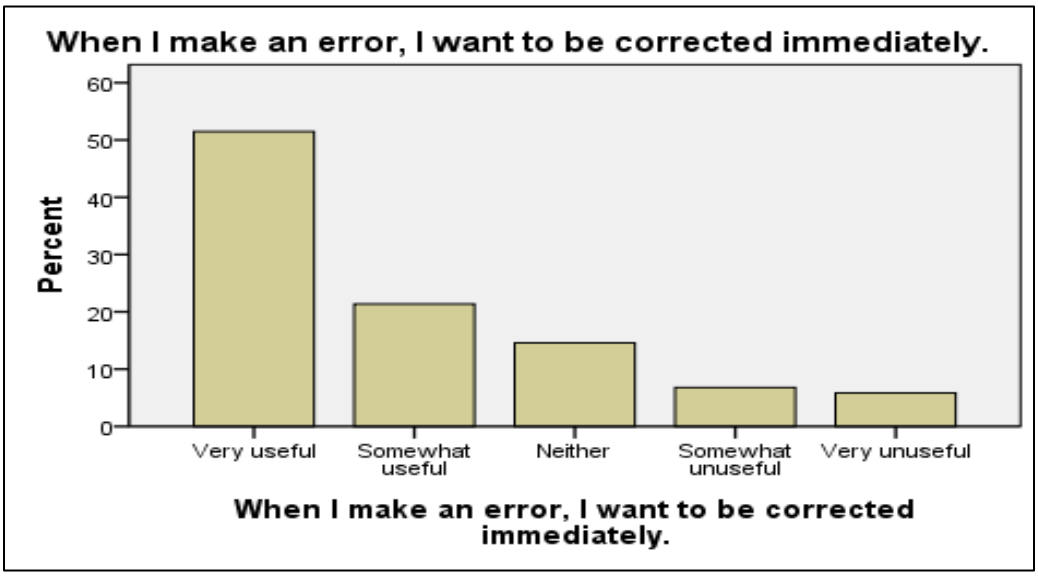

Figure 8. Bar chart of item 1 from question 7.

Grammarly provides immediate correction for underlined errors using red color, with suggested corrections highlighted in green. The paid version offers underlined and highlighted corrections and categorizes errors into correctness,

clarity, and engagement. It also provides immediate evaluation and performance scores. The free version offers some options underlined in red, with the correct answer highlighted in green.

The study revealed that 38% of respondents viewed Item 2 as very useful, 27% considered it somewhat useful, 20% neither, and 15% unuseful. This suggests that some students prefer immediate correction at the end of a writing assignment, as it may distract them from focusing on the flow of ideas and arguments. The percentages of usefulness indicate that students want to be corrected when they ask for correction, which is essential but not more important than immediate correction. Item 2 is ranked before the last item in the descriptive statistics of question 7 items, with a mean of 3.81 (see table 9 above).

When I make an error, I want to be corrected when I ask for correction.					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	Very useful	39	37.9	37.9	37.9
	Somewhat useful	28	27.2	27.2	65.0
	Neither	20	19.4	19.4	84.5
	Somewhat unuseful	9	8.7	8.7	93.2
	Very unuseful	7	6.8	6.8	100.0
	Total	103	100.0	100.0	

Table 9. Frequencies and percentages of item 2 from question 7.

WM is a method of corrective feedback that students prefer over immediate correction. It corrects errors by highlighting the entire word with different colors based on the error type. The system also labels the types of errors on the screen, such as basic feedback (well-developed, coherent, well-edited grammatical errors, punctuation, capitalization, and misspelled words) and advanced feedback (convincing). Students receive scores, final evaluations, and badges for each type of error, allowing them to understand their errors and improve their writing skills.

Teachers' and students' interviews confirmed the study's hypothesis, revealing that students prefer immediate correction after finishing their paragraph. Student 10 said that:

"There is a problem in the WM application; it detects the error but does not correct my mistakes immediately.."

However, some students preferred to be corrected after finishing writing the paragraph. For example, student 15 explains:

"After writing a paragraph, I correct my mistakes, evaluate my level, give myself a grade, and strive to improve next time.."

Students prefer feedback at the end of assignments to concentrate on content and avoid immediate corrections, as they believe it distracts them and may hinder timely completion. Student 19 states that:

"I prefer to correct my errors after I finish the assignment because it might take time for immediate correction."

Concerning highlighting colors, interviewed students gave their opinions based on their experience with the AWCF used in the experimental study. Student 20 argues: "WM is preferred for correct answers, outperforming Grammarly, Word spellcheckers, and Google Translate due to its user-friendly interface, color-coded error highlights, and clear word usage guidance."

He suggests that displaying different types of errors in different colors is the most effective design strategy, as it visually and systematically highlights errors for students, allowing automatic correction to become more ingrained in their minds, as seen in the WM application photos.

5. Discussion

These research results resemble findings in similar studies of AWE systems. Some of the most recent research has emphasized the user-centric perspective (Alsallami, 2017) (see also Brown, 2018; Burstein et al., 2018; Ranalli, 2018; Ranalli et al., 2017). Brown (2018) found that learners struggled with AWE's "Write and Improve" feedback system, requiring assistance to correct errors. The researchers chose Grammarly, a general-purpose AWE designed to assist non-native English authors due to its simpler structure and less crowded writing interface. Grammarly offers ample writing space, highlighted errors, and suggested remedies, making it easier to use than Write and Improve layouts (Brown, 2018, p.133). Li et al. (2015) demonstrated that Grammarly is easier for students and that a free edition is offered to users. Students found Criterion's feedback beneficial in mechanics, sentence-level vocabulary, and grammar but not in organization and substance.

Findings similar to the current study in other EFL contexts hint at a shared perception of the design of AWE tools. The study highlights the importance of explicitness in designing AWCF, with a clarity ranking higher than helpfulness in evaluations. Ranalli (2018) suggested that recognizing errors doesn't always mean understanding how to correct them. AWE tool makers should aim to provide as much specific feedback as possible, such as identifying the location of a run-on sentence or a fragment. While some tools like Grammarly focus on providing specific feedback, it is essential for AWE tools to provide both generic and specific feedback based on user needs, capabilities, and the instructional focus of the writing assignment, as assigned by instructors or students (Ranalli, 2018, pp. 653-654).

Based on more recent research implications and recommendations, the present research proved the importance of AWE tool design in increasing their usefulness. Ranalli (2021) concluded that AWE development is crucial for improving its usefulness and ease of use. Despite significant investment in commercial products like Grammarly and Criterion, evaluation work has primarily focused on error detection and correction, neglecting feedback design's impact on learning and trust. AWE developers should optimize devices for successful user engagement, especially for L2 students, and provide information on the dependability of different types of feedback. Tools should offer explicitness levels and highlight errors for L2 learners' grammatical growth (Ranalli, 2021, p. 14).

These research results are in line with findings in other EFL contexts. Chapelle et al. (2015) proposed the concept of usefulness in their argument-based validation framework, stating that Criterion feedback benefits students when making revision decisions. However, additional classroom-based research on AWE has clarified the precise utility of AWE program feedback. Indeed, the usefulness of the feedback can be determined by the linguistic features addressed, clarity, specificity, and relevance (Ranalli et al., 2017). These studies examined the attitudes and opinions of teachers and students toward AWE feedback. Chen and Cheng (2008) found that "My Access!" is ineffective for students, with 45% deeming it ineffective. The study suggests that language instructors should think critically before implementing AWE programs and create authentic teaching materials that align with the objectives of the writing curriculum. This analysis should consider students' specificities, writing aims, teachers' perceptions, and technological skills.

According to studies on revision and feedback, numerous components of AWE programs were underutilized. Attali (2004) examined the effectiveness of Criterion's feedback and revision features for L1 and L2 students in US high schools from 2002-2003. Only 30% of 33,171 articles were changed or resubmitted, while 70% were submitted without modification. This suggests that students did not fully benefit from Criterion's feedback due to difficulties understanding and correcting errors and reluctance or motivation to use the program.

However, other studies examined the effectiveness of AWE in promoting tertiary-level ESL writing. Specifically, Chapelle et al. (2015) identified six distinct types of user interaction with Criterion feedback (no change, add, remove, delete, transpose, and change). The researchers discovered that students made no revisions to 50% of errors, attributed to the program's frequent inaccuracy of feedback. According to these findings, Ranalli et al. (2017) argue that learners often avoid using support features in learning technologies like dictionaries, annotations, glossaries, and feedback texts due to perceived distraction, which impacts their mental effort and cognitive load. The use of help choices in writing can be challenging due to the cognitive burden, exacerbated by the simultaneous focus on the writing task and handling assistance recommendations, requiring significant mental effort to manage the writing process.

Similar to the present study, Cavaleri and Dianati (2016) reported that over 80% of students found Grammarly helpful and easy to use, with 94.4% rating it 4 or 5, with 5 being "very easy." However, two students found it difficult due to American grammar, and a student expressed dissatisfaction with the site's difficulty browsing. Jones et al. (2013) found that some students face a high cognitive barrier in comprehending grammatical principles, requiring advisers to engage with them to deconstruct Grammarly feedback propositions, which may be necessary for their writing. Graham et al. (2016) suggested that the US Institute of Education Sciences (IES) collaborates with expert panels to create practice guides based on evidence models, like Teaching Secondary Students to Write Effectively. These guides encourage students to engage in focused writing processes and teach appropriate strategies using a Model-Practice-Reflect instructional cycle. Burstein et al. (2018) demonstrated that WM follows Graham's guidelines, promoting systematic writing processes, a Model-Practice-Reflect instructional cycle, and complementary tasks like rereading and targeted feedback, implementing formative assessment for learning. Bennett (2011) emphasizes the need for new development in formative assessment to focus on specific approaches centered on exact subject domains, emphasizing the modeling of the writing process and its ability to deliver instruction.

In short, the advantages of programs such as the ones used in this study cannot be denied, although their full potential has not yet been used. There are also shortcomings, some inherent to the system itself, which researchers are looking for ways to minimize. Still, others are indeed due to students not yet coming to grips with all the subtleties of these automated systems, which calls for more efforts on the part of teachers and these system promoters to make them more easily accessible.

6. Conclusion

The study investigates EFL learners' perceptions and preferences towards AWCF (Asynchronous Writing Feedback) systems and their strategies. It assesses the ease of use, clarity, and usefulness of AWE systems (Writing Mentor and Grammarly) and their preferred CF strategies. The research used a mixed-methods approach, collecting, analyzing, and interpreting data from questionnaires and semi-structured interviews. An exploratory factor analysis was conducted to understand the underlying elements influencing respondents' perceptions and preferences for AWCF. The findings revealed that students and teachers found AWCF systems user-friendly and useful, with Grammarly being more beneficial than WM implicit feedback. Direct feedback is more precise and more useful, but students still need guidance to use indirect and graded feedback systems effectively. The study emphasizes the importance of AWCF in EFL classrooms and emphasizes the need for developers to balance different characteristics to accommodate diverse concerns. Regular and thorough AWCF is preferred due to its ability to produce error-free work and raise understanding of grammar issues. In contrast, direct correction is preferred due to time and accuracy concerns. Design-oriented research is needed to address feedback's textual and non-textual qualitative qualities and students' concerns about accuracy.

This current study significantly contributes to the existing knowledge base in the field of EFL writing education in several keyways. It provided valuable insights into the perceptions and preferences of EFL learners regarding Automated Writing Correction and Feedback (AWCF) systems. It assesses the ease of use, clarity, and usefulness of popular AWE systems like Writing Mentor and Grammarly, offering a deeper understanding of their use in educational settings. The study also compares explicit/direct and generic/indirect AWCF strategies, highlighting their effectiveness in EFL writing instruction. The mixed-methods approach, combining questionnaires, semi-structured interviews, and exploratory factor analysis, enhances the validity and reliability of the findings. The study emphasizes the importance of developers balancing various characteristics of AWCF systems to address diverse learner concerns effectively. This guidance is invaluable for developers seeking to design and refine AWE systems that meet the needs and preferences of EFL learners. The study also highlights the pedagogical implications of AWCF in EFL classrooms, advocating for metalinguistic engagement and graded AWCF among learners. The study calls for design-oriented research to address the qualitative aspects of feedback and students' concerns about accuracy in AWCF systems. This call underscores the need for continued innovation and refinement in developing AWE systems to maximize their effectiveness in supporting EFL writing instruction in the digital age.

AWCF is a vital tool in EFL writing education, providing specific feedback. However, its effectiveness relies on students' feedback management skills. Future research should explore the relationship between human and automated scoring, evaluating human scores' accuracy and the theoretical relevance of automated language scores (Zribi & Smaoui, 2021). Automated scoring engines, such as the Intelligent Academic Discourse Evaluator (IADE), can assess student writing outcomes across different populations. Despite calls to reduce AES systems, they are still essential for scoring and feedback.

Teachers should educate students about their current position, encourage independent paragraph construction, discourage private tutoring, and emphasize the employment benefits of EFL. Automated scoring systems can help identify students' writing difficulties, allowing educators to focus on content and meaning. Educational policymakers should consider how technology fits into the Personalized Learning process, enabling unprecedented levels of personalization. Inclusive course progression policies, faculty hiring, accountability policies, and teacher resources can help.

References

Alsallami, N. (2017). *L2 Learners' Perceptions and Preferences of Automated Corrective Feedback.* Graduate Theses and Dissertations. 15244. https://lib.dr.iastate.edu/etd/15244,

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

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in education: principles, policy, 18*(1), 5-25. Retrieved from https://doi.org/10.1080/0969594X.2010.513678

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

Brooke, J. (1996). Sus: a "quick and dirty'usability. *Usability evaluation in industry, 189*.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *AI Magazine, 25*(3), 27-27. doi:https://doi.org/10.1609/aimag.v25i3.1774

Burstein, J., Elliot, N., Klebanov, B. B., Madnani, N., Napolitano, D., Schwartz, M., . . . Molloy, H. (2018). Writing mentor: Writing progress using self-regulated writing support. *Journal of Writing Analytics, 2*, 285-313.

Burstein, J., Elliot, N., & Molloy, H. (2016). Informing Automated Writing Evaluation Using the Lens of Genre: Two Studies. *CALICO Journal, 33*(1), 117-141. doi:10.1558/cj.v33i1.26374

Cavaleri, M. R., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language Learning, 10*(1), A223-A236. Retrieved from https://journal.aall.org.au/index.php/jall/article/view/393

CCCC. (2014). Writing Assessment: A Position Statement. Retrieved from https://ncte.org/statement/writingassessment/

Chapelle, C. (2001). Computer-assisted language learning. In *Computer Applications in Second Language Acquisition* (pp. 44-94). Cambridge:: Cambridge University Press.

Chapelle, C. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology* (Vol. 7): John Benjamins Publishing.

Chapelle, C. (2010). The spread of computer-assisted language learning. *Language Teaching, 43*(1), 66-74.

Chapelle, C., & Erik, V. (2016). 20 years of technology and language assessment in language learning & technology. *Language LearningTechnology, 20*(2), 116-128.

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.

Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *The English Journal, 93*(4), 47-52.

Cotos, E. (2010). *Automated Writing Evaluation for non-native speaker English academic writing: The case of IADE and its formative feedback.* (PhD). Iowa State University,

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

ETS. (2020). About the TOEFL iBT Test. Retrieved from https://www.ets.org/toefl/test-takers/ibt/about

Ferris, Liu, H., Sinha , A., & Senna , M. (2013). Written corrective feedback for individual L2 writers. *Journal of second language writing, 22*(3), 307-329. doi:https://doi.org/10.1016/j.jslw.2012.09.009

Ferris, & 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. doi:https://doi.org/10.1016/S1060-3743(01)00039-X

Garrett, N. (1987). Modern Media in Foreign Language Education: Theory and Implementation, chapter A psycholinguistic perspective on grammar and CALL. *National Textbook Company*, 19.

Gass, S., & Selinker, L. (1983). *Language Transfer in Language Learning. Issues in Second Language Research*: ERIC.

Goins, M. (2015). *Written corrective feedback: strategies for l2 writing instructors. SLW News: The Newsletter of the Second Language Writing Interest Section, TESOL.org . http://newsmanager.commpartners.com/tesolslwis/issues/2015-10-07/3.html*.

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., . . . Regional Assistance , I. o. E. S., US Department of Education. Retrieved from the NCEE website: http://whatworks. ed. gov. (2016). Teaching secondary students to write effectively (NCEE 2017-4002). *Washington, DC: National Center for Education Evaluation Regional Assistance , Institute of Education Sciences, US Department of Education. Retrieved from the NCEE website: http://whatworks. ed. gov*.

Grammarly. (2021). Retrieved from https://support.grammarly.com/hc/en-us/categories/115000018611-About-Grammarly

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL quarterly, 44*(1), 84-109.

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language, LearningTechnology, 19*(2).

Lee, J., & Hegelheimer, V. (2012). *A hybrid use of Criterion® and teacher feedback in process writing.* Paper presented at the EUROCALL conference, Gothenburg, Sweden.

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. doi:10.1016/j.jslw.2014.10.004

Madnani, N., Burstein, J., Elliot, N., Klebanov, B. B., Napolitano, D., Andreyev, S., & Schwartz, M. (2018). *Writing mentor: Self-regulated writing feedback for struggling writers.* Paper presented at the Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations.

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653-674. doi:10.1080/09588221.2018.1428994

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, 100816.

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.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning1. *Applied Linguistics, 11*(2), 129-158. doi:10.1093/applin/11.2.129 %J Applied Linguistics

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 357-385.

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

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

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327-369.

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*(2), 111-122.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*(4), 255-272.

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

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*(3), 335-353. doi:https://doi.org/10.1177/0265532210364406

Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. *Handbook of automated essay evaluation: Current applications new directions*, 36-54. doi:https://www.routledgehandbooks.com/doi/10.4324/9780203122761.ch3

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? In: Sage Publications Sage UK: London, England.

Zribi, R., & Smaoui, C. (2021). Automated versus Human Essay Scoring: A Comparative Study. International Journal of Information Technology and Language Studies, 5(1).

Zribi, R., & Smaoui, C. (2021). Rater-task interaction effects on testing EFL learners' written performances. Language Testing, 4, 23-36.