THE JOURNAL OF TEACHING ENGLISH FOR SPECIFIC AND ACADEMIC PURPOSES
Vol. 10, Nº 2, 2022, pp. 301–317

UDC: (811.111+37):316.723 https://doi.org/10.22190/JTESAP2202301S

Review research paper

THE IMPACT OF AUTOMATED WRITTEN CORRECTIVE FEEDBACK ON EFL LEARNERS' ACADEMIC WRITING ACCURACY

Abdulaziz B Sanosi

Prince Sattam bin Abdulaziz University, Saudi Arabia

Abstract. *The positive effect of Corrective Feedback (CF) on students' writing performance has been a topic of dispute for a long time. The dominant belief around this issue is that students benefit from feedback to a certain extent. However, there is no consensus on what type of feedback can achieve that or the effect of CF provision on language learning in general. This study investigates the impact of Automated Written Corrective Feedback (AWCF), namely Grammarly AI-powered writing assistant, on students' academic writing accuracy. After being allotted to control and experimental groups, sixty-four university students participated in the study. The participants underwent a pre-test to validate their homogeneity and levels and a post-test to explore the effect of using Grammarly on the written work of the experimental group. The main finding of the research is that after 14 weeks of using Grammarly, the experimental group members showed a significant improvement in their written accuracy compared to the control members. Moreover, it was found that the progress was represented in a substantial drop in the number of errors pertaining to specific categories while errors of other types remained unaffected. The implications of the findings are discussed, and suggestions for further research are presented.*

Key words*: corrective feedback, Grammarly, AWE, WCF, error treatment*

1. INTRODUCTION

It has been widely acknowledged that "written accuracy is important in the real world" (Ferris 2011, 14). This importance stems mainly from the advent of English as lingua franca and a medium of instruction that requires students to use English for learning disciplinary courses and communicate their knowledge and information while studying at university and post-graduate levels (Coffin et al. 2003). It is also a basic need for academic professionals of different disciplines for scientific publishing, especially with the vast spread of the concept of publish-or-perish that requires them to write, mainly in English (Raitskaya and Tikhonova 2022). Furthermore, In the contexts of English as a Foreign Language (EFL), English for Academic Purposes (EAP) and

Submitted August 4th, 2022, accepted for publication August 13th, 2022
Corresponding author: Abdulaziz B Sanosi, Prince Sattam bin Abdulaziz University, Saudi Arabia
E-mail: a.assanosi@psau.edu.sa

English for Specific Purposes (ESP). It is required to acquire the language and learn the disciplinary content (Wahyuningsih 2020, Utkina 2021)).

Therefore, teachers deprecate writing mistakes and consider them disgraceful. Accordingly, they consider it essential to provide Written Corrective Feedback (WCF) to learners, ask them to revise their work accordingly and repeat the process as often as possible (Burstein, Chodorow and Leacock 2004). Although WCF has been debated for a long time, no consensus has been reached regarding its effectiveness in improving writing performance. Some researchers defended its value for promoting learners' writing and stated that it "significantly improves accuracy" (Bitchener 2008, 15) and makes students "more confident in performing their academic writing task" (Wahyuningsih 2020, 13). Notwithstanding, other researchers, e.g. (J. Truscott 1996), thought contrariwise, stating that it is ineffective and even harmful for learning.

Parallel with the vast advances in computers and related technology; there has recently been considerable interest in Automated Writing Evaluation (AWE) which is a suite of software that automatically assesses students' writing. AWE was introduced in the EFL classroom setting to account for teachers' heavy workload that prohibits them from accurately following their students' writing (Burstein, Chodorow and Leacock 2004). It was recognised that AWE effectively checks students' spelling, grammar, word choice, tone, and plagiarism. However, there is also no decisive proof regarding their impact on improving students' written accuracy, i.e. appropriate use of grammatical conventions, punctuation and other writing mechanics, whether in the long or short term.

Many recent studies investigated AWE; however, most of these studies focused on how this technology can be used in teaching (Barrot 2020, Ghufron and Rosyida 2018, Li, Link and Hegelheimer 2015). Other studies explored how students can benefit from such technology (Jim 2018), yet others evaluated the types of feedback presented by AWE applications (Luo and Liu 2017). Relatively few studies have explored the effect of AWE application in improving students' performance in writing (Qassemzadeh and Soleimani 2016). As this is a significant aspect of research that can generate insightful implications on the use of AWE, the present study aims at investigating the effect of using Grammarly, a well-known AWE application, on improving the academic writing accuracy of Arab EFL students. As a sample, the study targeted university students majoring in English language and literature at Prince Sattam bin Abdulaziz University, Saudi Arabia.

2. Theoretical Background

Error correction in the classroom and assessing students' activities are among the several foci of research in second language acquisition, TESOL and academic writing. Most related theories acknowledge that errors are an inevitable phenomenon in language learning in general (Huang 2009). However, there are different perspectives on dealing with them. For example, according to the behaviourist theory, language learning is a habit formation process achieved through stimulus, response, and reinforcement. Positive feedback is needed to reinforce learners' correct output, while incorrect output needs to be corrected. Consequently, traditional teaching methods based on behaviourist views focus more on accurate language production, and teachers tend to correct students' mistakes to achieve habit formation through conditioning (Demirezen, 1988).

Suggesting that instruction should focus on enabling students to use language successfully in a realistic setting, the Communicative Language Teaching (CLT) approach was presented in the 1970s. The main elements of the approach were inspired by the *comprehensible input* (Krashen 1982). According to the hypothesis, "we acquire by *going for meaning* first, and as a result, we acquire structure" (p. 21). Accordingly, the CLT method contradicts the traditional approach, where the focus was on the structure. Furthermore, it does not encourage continuous error correction. Krashen (1982) considers error correction "a serious mistake" (p. 74) because it makes students defensive, uses an avoidance strategy, and focuses on form more than on meaning.

Another teaching method that merges the two techniques mentioned above is *focus on form*, which suggests targeting accuracy through communicative-based activities (Long 2000). *Focus on form* can be achieved through CF, which targets the meant form to be taught. Applying *focus on form* in class and reacting to students' errors is found to develop students' accuracy (Lightbown & Spada 1990). Another effective method in this strand is to provide CF in written format, a common approach known as Written Corrective Feedback.

2.1. Written Corrective Feedback

According to second language learning theories, feedback provision is one way to promote students' motivation and ensure linguistic accuracy (Ellis 2009). Feedback can be positive to reinforce correct production or corrective, defined as "any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner utterance." (Chaudron 1977, 31). Further, corrective feedback can be *oral*, through regular classroom instruction, or written in the forms of notes, symbols, or error corrections (Wirantaka, 2019). The debate around the latter type, i.e., Written Corrective Feedback (WCF), is more contentious. The efficacy and practicality of WCF are always a topic to dispute (Ellis 2009, Beuningen 2010, Maleki and Eslami 2013, Jang 2020). WCF is widely believed as a "means of fostering learner motivation" (Ellis 2009, 3), and there is sufficient evidence that it significantly improves accuracy (Bitchener 2008). However, other views devaluing its usefulness emerged later. The initiative claims in this strand were raised by Truscott (1996), who not only claimed that CF is ineffective for both logical and functional reasons but also believed that it "has harmful effect" (J. Truscott 1996, 327). The significant argument presented by Truscot about the uselessness of WCF is that "grammar correction is a bad idea" (J. Truscott 1999, 111). Furthermore, serious methodological issues are found in the research that yields contrary claims. (J. Truscott 2020).

There are, however, subsequent studies that refute these claims and promote using WCF. For instance, Bitchener (2008) states that providing WCF on specific errors would significantly improve academic writing accuracy. Kang and Han (2015, 1) reported that "corrective feedback can lead to greater grammatical accuracy in second language writing, yet its efficacy is mediated by a host of variables, including learners' proficiency, the setting, and the genre of the writing task.". Their conclusion is based on their review of 21 studies related to CF practice and results. Thus far, there is no conclusive statement about the exact effect of WCF in improving EFL learners' academic writing accuracy. However, it is a standard and indispensable practice in the EFL classroom (Luo & Liu, 2017). Moreover, its role in L2 writing development is "an exciting and dynamic area of investigation and, as such, is likely to continue engaging the energy and insights of established and emerging scholars" (Bitchener and Ferris 2012, 27). These studies yielded many dichotomies regarding types of feedback and errors to be corrected.

2.2. Types of feedback

The focused versus unfocused classification is based on the views about the amount of feedback provided to learners and its distribution to different error categories. In this regard, researchers differentiate between focused feedback, where teachers "focus attention on a few error types rather than try to address all the errors learners make" (Ellis 2009, 6) and unfocused (or comprehensive) feedback, which "involves correction of all errors in a learner's text, irrespective of their error category" (Beuningen 2010, 11). The argument for adopting the first approach is the limitedness of students' cognitive capacity, which is susceptible to overloading with much feedback. This is because "too much feedback at any one time might be de-motivating or too burdensome for cognitive processing." (Bitchener and Ferris 2012, 128). Accordingly, focused feedback is recommended for teachers and researchers. It is perceived to effectively teach the targeted forms and help the researchers design proper test instruments.

Another classification divides feedback into direct and indirect. Direct feedback points to the error and provides the correct linguistic form thereof, and indirect feedback only informs the learner that an error has been committed (Ferris 2011). Other terms to describe these types are explicit versus implicit feedback (Ellis 2009). No consensus as to what type is more effective. Indirect feedback is praised because it engages students in language processing and problem-solving, which foster their learning. However, direct learning is considered better since it provides learners with sufficient information about their errors (Beuningen 2010).

2.3. Types of errors

Researchers distinguish between errors to be corrected by the teachers. Ferris (2011) classified errors into treatable and untreatable. Treatable errors by such a classification can be explained and remedied following specific rules. These errors include article usage, subject-verb agreement, and singular/plural forms. On the other hand, untreatable errors are non-distinctive errors that cannot be corrected according to fixed rubrics such as lexical and some word choice errors. Research findings pertaining to this dichotomy support treatable errors as focusing on them is more likely to lead to a more successful CF (Beuningen 2010). Also, Ferris (2011) suggested that teachers should provide indirect feedback to correct untreatable errors because this type is more responsive to indirect feedback, while direct feedback works well with treatable errors.

Another typology divided errors into local and global. According to Touchie (1986), local errors such as article, preposition, and verb inflection errors do not impede text comprehension. In contrast, global errors such as word order inaccuracy distort the utterance's meaning and disrupt communication. Although it seems intuitive that global errors are more serious and should be given more consideration by teachers (Ellis 2009), it is nonetheless context-dependent (Ferris 2011). Besides, deciding which type of error to be local or global is not always easy for regular teachers.

2.4. Corrective feedback for academic writing

Academic writing is the writing used in universities and scholar publishing to communicate scientific ideas and information to instructors or the wider academic community. This type of writing requires strict standards to distinguish it from other types

of writing. For example, it should demonstrate knowledge of rhetorical conventions, linguistic features, vocabulary, and syntax (Utkina 2021). Therefore, instructors must devote all possible strategies to fostering students' written accuracy. Feedback is one of the essential techniques followed by teachers in this regard (Wirantaka 2019, Wahyuningsih 2020).

In Academic writing classes, especially in higher education settings, two types of CF are usually used: teacher and peer feedback. Both types of feedback are found effective. However, research has proved that teachers' CF of academic writing is a complex process for several reasons. Coffin et al. (2003) quantified factors that cause such a dilemma, the most important of which are: (1) The mismatch between teachers' and students' concepts about academic writing, (2) different and inconsistent comments from teachers on similar pieces of writing, (3) mixed, unclear and confusing comments, and (4) comments are derived by teacher disciplinary background. Nevertheless, it is widely acknowledged that L2 students still prefer teachers' CF over peer feedback which may be more suitable for L1 learners (Hinkel 2004). Therefore, the above barriers need to be tackled to provide more efficient feedback since it is believed that "to improve the quality of students' academic writing, teachers need to improve the quality of their feedback" (Wirantaka 2019).

Research on the CF of academic writing focused on the types of CF provided for students, students' perception of the practice (Leong and Lee 2018), and the effect of different kinds of feedback on students' academic writing. Dewi and Jati (2017) found that teacher-written feedback is the most beneficial method for improving students' writing compared to video-based and peer feedback. Furthermore, Leong and Lee (2018) observed that both teachers and students agreed that feedback should help students become insightful and autonomous learners and that it should be detailed and personalised. However, this does not mean that teachers should act as editors for students' written work but rather instruct them on how to deal with frequent and sometimes persistent errors (Hinkel 2004). Hence, it can be concluded that appropriate teacher WCF is a crucial strategy in academic writing instruction.

2.5. Automated Writing Evaluation

Recently, the WCF provision has witnessed new development. Automated Writing Evaluation (AWE) software is a suite of tools that assess students writing by comparing a written text to built-in databases of writing and specific rubrics to measure students' lexical, syntactic, and grammatical aspects (Elliot et al. 2013, Hockly 2019, Woodworth and Barkaoui 2020, Miranty and Widiati 2021). The theoretical foundation behind using AWE lies in practical reasons. Teachers are usually encumbered by burdens that hinder them from following up with their students or doing so accurately. Implementing new techniques can help them to a far extent as scoring would be faster, cheaper, and more accurate (Rudner and Liang 2002, Miranty and Widiati 2021).

Other AI-powered writing assistants are customarily considered under the umbrella of AWE though they are slightly different from them. An example of these tools is Grammarly which is an AI-powered writing assistant founded by Max Lytvyn, Alex Shevchenko, and Dmytro Lider in 2009 (Grammarly 2022). Its primary aim, as advertised, is to foster online writing and hence enhance communication. Grammarly performs more than its name implies. It is not limited to grammatical correction, but its corrective suggestions include style, clarity, engagement, and delivery. These potentials attracted academic writing researchers' attention since they address the crucial requirements of the style. Several studies supported the claims that it is highly effective in promoting students writing (Barrot 2020, Ghufron and Rosyida 2018, Nova 2018, O'Neill and Russel 2019).

Woodworth & Barkaoui (2020) elaborated on the significant differences between standard AWE tools and these grammar checkers, stating that the latter "cannot be moderated by the teacher, do not evaluate writing quality, and do not include any portfolio and class management tools" (p. 237). Notwithstanding, many recent studies consider Grammarly an AWE tool (Miranty and Widiati 2021), while other studies adopted a more precise term, *Automated Written Corrective Feedback* (AWCF). Such studies appraise AWCF applications stating that they do more than AWE ones since they provide feedback on students' performance and thus help improve students' accuracy (Burstein, Chodorow and Leacock 2004, Guo, Feng and Hua 2021).

As Grammarly gained a gradual reputation as a powerful writing assistant and grammar checker, several studies investigated it concerning different variables. Regarding students' perceptions of the application, Nova (2018) explored three Indonesian students' experiences using Grammarly. The findings indicated that the participants found Grammarly helpful as it provides easy, fast, and free editing services. Some weaknesses of the programs, such as misleading feedback and lack of content and context, were also reported. Other studies compared human feedback to an automated one. An example of such studies is (O'Neill and Russel 2019), which examined the perceptions of one group of students (n=54) who received additional Grammarly feedback and regular feedback from their academic advisors. Further, their perceptions were compared to those of other group members who received feedback only from their teachers. The results indicated that students who implemented Grammarly as a writing assistant were more satisfied with the grammar advice they received. Their satisfaction was related to short-term benefits on their marks and the long-term benefits to their writing.

Comparing Grammarly's potential is not limited to students' perception of it. Ghufron and Rosyida 2018 compared the AWCF provided by Grammarly to feedback provided by teachers. The material was written by two homogenous groups of an Indonesian university, i.e. (control and experimental group), each incorporating twenty participants. Interestingly, the results showed that the students who received AWCF committed fewer errors in the post-test than those whose work was evaluated by the teacher. The improvement of their work was limited to diction, grammar and writing mechanics. However, they showed no development in content and organisation.

Similarly, Para and Calero (2019) used a pre-test/post-test experimental research design to explore the improvement of the writing of 28 university students during a semester of study. The participants used Grammarly and *Grammark*, another AI writing assistant, to get feedback. The study demonstrated a significant improvement in the performance of the participants who used either of the two assistants at the post-test level. Further, it was reported that the improvement was in most aspects of writing accuracy, including grammar, punctuation, mechanics, and style. The researchers traced the improvement to the student's motivation as they learn independently. However, they asserted the role of the teacher in compensating for the limitations of the tools, which they found to be in content development. Another study (Guo, Feng and Hua 2021) investigated the effect of Grammarly on 36 students' second draft of an essay. The result showed that errors have decreased significantly after the revision and that students have responded to most Grammarly suggestions.

Studies on the impact of Grammarly on academic writing accuracy can provide powerful implications to EFL teachers about how, when and to what extent they should integrate Grammarly into their teaching plans. However, studies of this kind are relatively

few compared to the extensive use of Grammarly in different contexts, i.e., general online writing, CMC, social network, and classroom setting.

Research Hypotheses

The paucity of studies of this type, especially in Arab EFL learning settings, is the primary motive of the current study. To fill in such a gap, and because the approach is still modern, this research attempt to test the following hypotheses:

H_0: *There are no significant differences between the writing accuracy levels of students who use Grammarly and those who do not.*

H_a: *Students who use Grammarly will write more accurately than those who do not.*

The following methods and procedures will be adopted to achieve the study aims.

3. METHODS

3.1. Research design

This study adopted a between-group experimental research design where control and experimental groups participated. The experimental group's members used Grammarly to assist them in their writing assignments for fourteen weeks. The effect of such a practice was then assessed to test the research null hypothesis. To measure the effectiveness of the treatment, the researcher used a pre-test/post-test design. First, a pre-test was conducted to check the writing level of the research sample before the treatment. After the treatment, a post-test was administered to fulfil two purposes: first, to measure the effectiveness of using Grammarly by members of the experimental group, and second, to confirm that there are no other factors that affect the post-test results. These will be confirmed by reviewing the control group members' results.

3.2. Participants

Sixty-four female college students participated in the study. The participants were students at the fifth level (third year) at the English Language and Literature Department, Prince Sattam bin Abdulaziz University in Saudi Arabia. Although there were no available data concerning their exact proficiency levels according to the standard benchmarks, it can be stated that their levels range from intermediate to upper-intermediate. They have studied English as a general course for about ten years at public schools and as a major for four semesters. They have already studied two courses in writing skills, two grammar courses, one vocabulary course, and courses on general linguistics and literature.

When conducting the study, the participants were enrolled in the Applied Linguistics (ENGL 3150) course taught by the researcher. They were divided into two sections by the university's automated registration system, which assigns students to their sections by the precedence of registration. The researcher took advantage of this random allocation and appointed one section as a control group and another as an experimental group. Each of the groups comprised thirty-two students. This procedure was believed to assure the homogeneity of the groups as their levels are comparable based on their grades in the previous courses.

3.3. Instruments

3.3.1. Grammarly

Grammarly is a well-known application that provides feedback on writers' grammatical errors, vocabulary usage and writing mechanics. It provides its service through two plans: free, which can suffice a standard user, and a premium for more advanced tasks. The editing service is presented to users on different platforms. As of July 2022, it has an application for Microsoft Windows, a plugin in Microsoft Word, Outlook and Chrome browsers. It can also be used in portable devices through the virtual keyboard for smartphones or the iPad application.

The researcher presented an instructional session on downloading, installing, and using the application. The experimental group participants were asked to download the application's free version and the plugins for Microsoft Word and Google Chrome. Further, he followed up on the student's progress and answered questions about its use. the participants did not report significant difficulties in using the application.

3.3.2. Writing tasks

The researcher designed two writing tasks to assess the participants' written accuracy at the two stages of the research. Students were asked to write a text of not less than 200 words on one of four topics on both tasks. Each task contains descriptive, narrative, argumentative, and expository options. Table 1 presents the topics for the writing tasks.

Table 1 Writing topics for the pre-test and post-test

Pre-test	Post-test
Yourself	Your Hometown
Your Childhood Memories	Your Experience in Learning the English Language
Advantages and Disadvantages of the Internet	Pros and Cons of Smartphones
A Place You Will Never Forget	An Unforgettable Vacation
How to Cook Your Favourite Dish	How to Shop Wisely

The topics were meant to be straightforward as the primary aim is to measure forms more than content.

3.4. Data collection and assessment

The writing texts were submitted through blackboard LMS, downloaded, and printed for manual review. For the pre-test writing, the task was held asynchronously; students were given three days to present their work through the LMS. However, an online session was held to administer the writing post-test synchronously. Also, the researcher used Respondus LockDown Browser1 to ensure that the participants would use no writing assistants or external software other than Blackboard textbox.

[1] Respondus Lockdown Browser is "a custom browser that locks down the testing environment within a learning management system." (Respondus 2021)

The collected 128 texts were assessed manually by two raters, the researcher and another university professor. The rubrics for evaluating the texts were based on the initial review of the writing material and focused on errors that met two criteria. First, the error should be a treatable grammatical error. This criterion excluded errors in other aspects such as punctuation, style, and spelling. Second, the errors should be marked and responded to by Grammarly as clearly as possible. The researcher used a premium Grammarly account to study the suggestions provided for the errors. The application goals were set to assess the students writing against standard academic writing rubrics, as shown in Figure 1 below.

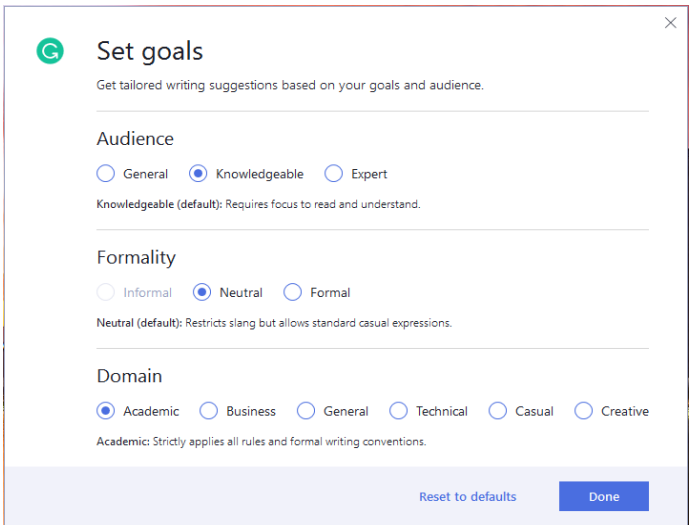

Fig. 1 Goals for tailoring Grammarly feedback to the participants' writing

Accordingly, the top five errors committed by the participants were elected to measure the students' writing accuracy and any potential progress that would occur thereof. The five errors checked for were: (1) articles, (2) subject-verb agreement, (3) verb tense, (4) word choice, and (5) singular/plural form errors. The ceiling score for each writing task was 25 points, and the correction was based on deducing one point for each error of the five mentioned types.

The correction rubric was then explained to the second-rater. The Cohen's Kappa for inter-rater reliability was conducted using SPSS to check the level of agreement between the two raters. The results of the inter-rater agreement were Cohen's *k*: 0.69 for the pre-test and 0.75 for the post-test, which suggest substantial agreement in the result for both tests. Moreover, the researcher checked the Grammarly application frequently to check how it responded to different error types.

3.5. Data Analysis

To generate the results for the study, the researcher conducted an independent samples T-test. The test results were used to interpret the difference between the two groups in terms of mean scores, standard deviation (SD), and the significance level of difference between

the groups. Moreover, the researcher analysed the errors of the two groups to determine which errors were decreased after using Grammarly.

4. Results and Discussion

The overall results of the two tests by the participants of the two groups are presented in Table 2 below.

Table 2 The results of the two tests by members of the two groups

Test	Pre-test				Post-test			
	Total Errors	Mean score	Average error*	SD	Total errors	Mean score	Average error	SD
Control	237	17.59	7.4	3.2	277	17.12	8.7	4.3
Experimental	265	16.71	8.3	4.0	144	20.59	4.4	3.1

*_Note_. The value indicates the average error per text.

The results of the two independent samples T-test are presented in Table 3 below

Table 3 The results of the independent samples T-test

	t	df	Sig. (2tailed)	Mean difference
Pre-test	954	62	.344	.875
Post-test	-3.66	62	.001	3.46

From the two tables, it can be reported that:

The 32 participants who used Grammarly as a writing assistant for 14 weeks (M = 20.59, SD = 3.1) compared to the 32 participants of the control group (M = 17.12, SD = 4.33) demonstrated significantly better writing accuracy scores, t (62) = -3.66, p = 0.001. These results provided sufficient proof to reject the null hypothesis raised by the study. This finding is justified by the low p-value of (.001), which is much less than the alpha level (a = .05). Therefore, it supports the alternative hypothesis that students who use Grammarly would write more accurately than those who do not.

The improvement in writing accuracy is represented in the overall writing scores of the students who used Grammarly. A considerable reduction of grammatical errors was witnessed in the experimental group members' performance after using Grammarly for fourteen weeks. The mean score of the participants jumped by 3.5 in the post-test. Bearing in mind that the ceiling score of the tests is 25, this can be considered a significant improvement as it represents 14% of the total score.

Two considerations support these results. First, it is noteworthy that participants of the two groups were considered homogeneous before the experiment. This finding is supported by their mean pre-test scores, which were different by only (0.09). This slight difference can be traced to writing topic selection, rating inconsistency, or typical individual differences. Secondly, when comparing the experimental group scores to those of the control group on the post-test, the results showed not only different performance in favour of the experimental group but also a slight drawback in the control group's mean score, i.e.,

control group members' mean score in the pre-test (17.6) was better than their mean scores in the post-test (17.1). This finding rules out any external effect that may contribute to improving experimental group members' writing scores.

These findings align with the results yielded by previous literature on using Grammarly, e.g. (Qassemzadeh and Soleimani 2016, Ghufron and Rosyida 2018, Para and Calero 2019, Guo, Feng and Hua 2021). However, some differences in the methodology and the procedures followed in these studies are noted. For example, Guo, Feng, and Hua (2021) supported the positive effect of Grammarly on students' writing. However, since students wrote two drafts of the same topic (the second of which was assisted by Grammarly), the practice effect could intervene. Also, (Para and Calero 2019) used Grammarly as a teaching complement to give CF to students; hence the development of the scores cannot be traced back only to their use of Grammarly and Grammark applications, although their effects are very likely. Hence, claiming that improving students' writing is because of AWCF only maybe a hasty generalisation.

These considerations are taken into account in this study by taking three precautions: (1) No WCF feedback was provided to the research participants by the teachers during the experiment, (2) the writing post-test was about new topics to eliminate practice effect, and (3) the post-test was administered in a virtual locked environment using Respondus Lockdown Browser to prohibit potential use of writing assistants.

On the other hand, some previous studies adopted methods that are not followed herein and hence might be considered likely limitations of the present study. For example, Qassemzadeh and Soleimani, (2016) applied an experimental method that utilised a grammar pre-test, a post-test, and a delayed post-test administered to the participants in six sessions. A delayed post-test can give insightful implications about the long-term effect of the Grammarly feedback on students' writing accuracy. The finding of the current study cannot answer this question. Likewise, other studies evaluated students' academic writing in many aspects. For example, (Ghufron and Rosyida 2018) assessed students' performance in terms of vocabulary usage, mechanics of writing, and grammar. Thus, unlike the present study, it can provide a broader outlook on the potential of Grammarly on students' academic writing.

This study, however, elaborated on different aspects of grammar by classifying error types and checking which one seemed to be affected by implementing AWCF through Grammarly. To achieve this, the performance of the control group members in the two tests is reviewed. The results are presented in Figure 2 below.

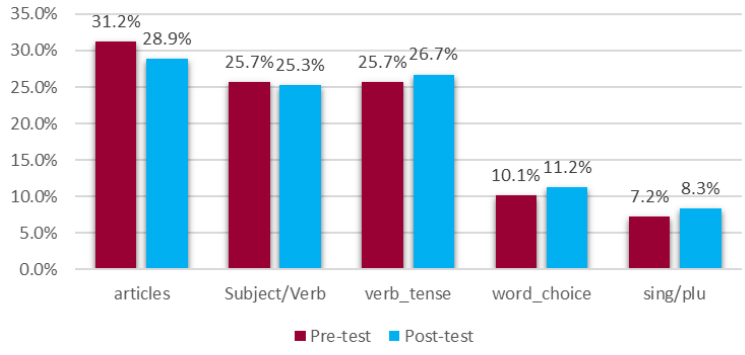

Fig. 2 Percentages of errors by the control group participants

The results show almost a consistent performance in the two tests in terms of the percentage of errors committed in each type. Most committed errors in both tests were those related to article usage, and the fewest ones were those related to word choice and singular/plural forms, respectively. However, errors pertaining to subject-verb agreement and verb tense exchange their ranks. In the post-test more verb tense errors are committed. More importantly, the percentages of each error type in both tests are interestingly similar, with a low difference that reaches only (2.3) in its maximum. These findings represented the basis upon which the experimental group results were analysed to yield the results displayed in Figure 3.

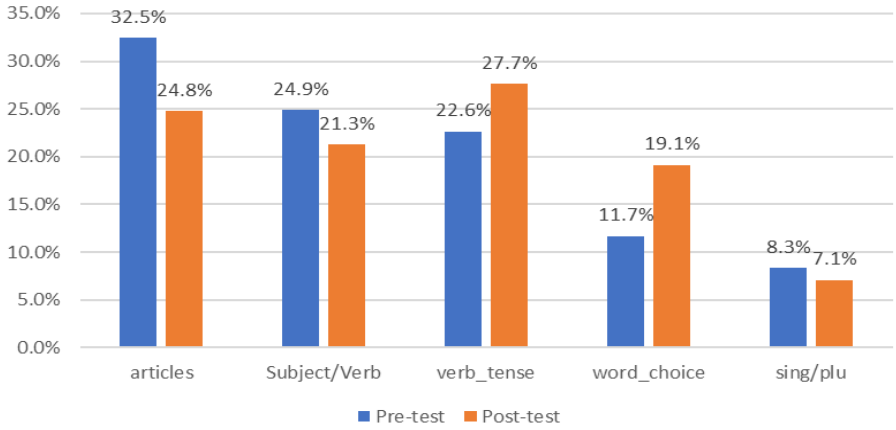

Fig. 3 Percentages of errors by the experimental group participants

Contrary to the control group results, a noticeable drop in the percentage of the most committed errors is witnessed. Errors in article usage decreased by 7.7%, and errors in the subject-verb agreement dropped by 3.6%, which can be considered a substantial difference. On the other hand, errors related to other types had not decreased; instead, percentages of verb tense and word choice errors had increased in the post-test. Concerning the last type, singular plural forms, no significant difference was witnessed as the difference was only 1.2 in favour of the post-test.

From these findings, it can be inferred that Grammarly feedback effectively teaches errors related to article usage, subject-verb agreement, and, to a lesser extent, singular plural form. Errors related to verb tense and word choice are not likely to be remedied using Grammarly's AWCF. These findings contradict what was found by Ferris (2006), who claimed that students made progress in reducing errors relating to verb tense and committed more article errors after being exposed to CF regarding such error types.

It is not probable that the improvement in specific error types is because of the nature of the feedback provided by Grammarly for each error category. Figures 4 and 5 below display examples of Grammarly corrective suggestions for the five error types taken from the participants writing in the pre and post-tests.

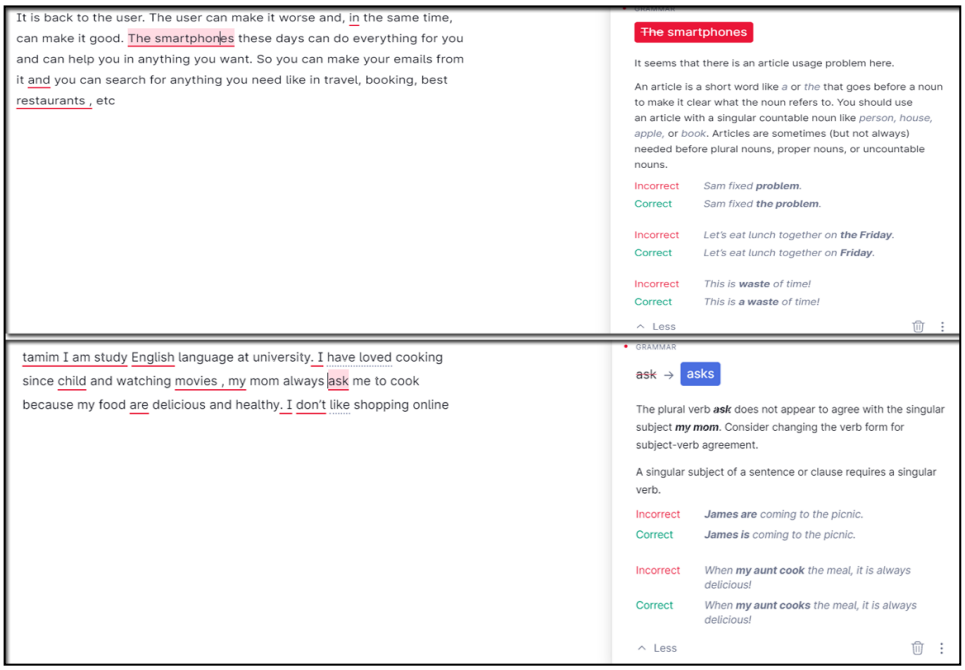

Fig. 4 Grammarly suggestions for correcting article and subject-verb agreement errors

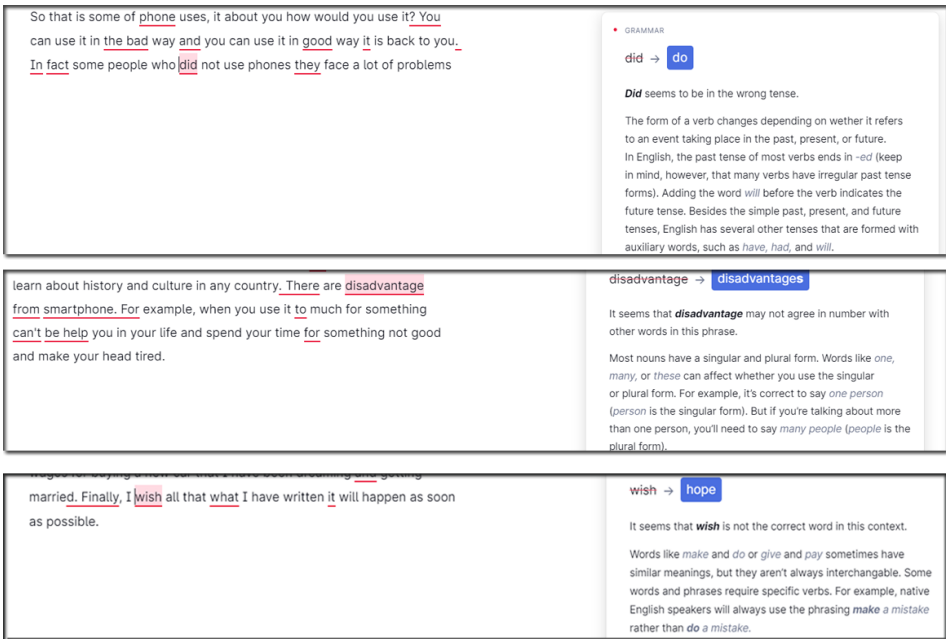

Fig. 5 Suggestions for correcting tense, singular/plural form, and word choice errors

The examples show that the Grammarly feedbacks were almost consistent. They started by displaying the mistake crossed out in red, suggesting correction in blue, giving a general metalanguage-based description of the mistake, explaining the rule to follow, and ending with a few examples. Articles subject-verb agreement errors, for instance, have no more elaborated or clear feedback. This fact entails other reasons why the participants improved in a specific category and regressed or showed no progress in others.

From the researcher's viewpoint, the findings can be justified by considering the nature of the verb tense and the word choice errors. Errors of these types are more related to content and semantics than linguistic structure. Therefore, correcting them requires a comprehensive understanding of context, more profound language knowledge, and a richer vocabulary. On the other hand, verb-subject agreement and article errors can be mastered by following fewer and less complicated rules.

Another justification for this finding is that the relative frequency of errors (and the CF thereof) contributed to the student's understanding of feedback and corrective suggestions. This relative frequency might cause a type of *focused* CF. In this situation, students are continuously exposed to a specific type of corrective suggestion, which can facilitate the process of learning and thus create a practice factor. Focused feedback promotes students' performance and is considered more effective than unfocused or comprehensive CF (Ellis 2009, Beuningen 2010, Ferris 2011). In the context of the present study, this hypothesis is supported by the fact that the students' improvement was witnessed in the two most committed error categories. Moreover, the slightest improvement was noticed in singular/plural errors which were the fewest committed errors and thus received minor corrective suggestions.

As far as the researcher knows, this finding is unprecedented. Moreover, it contradicts one of the pioneer studies in the field (Ferris 2006). Therefore, more studies are needed to support the claim accompanying focused versus comprehensive feedback, treatable versus untreatable errors, and global versus local error dichotomies. Such studies should also take account of the limitations of the present study. They should be conducted in more extended periods, utilise larger samples, and adopt delayed post-tests to assess the long-term effect of Grammarly on students' written accuracy on specific error categories.

The current study's findings imply the efficacy of using AWCF in teaching academic writing to EFL learners. Teachers can use Grammarly as a writing assistant inside and outside the classroom to enable ongoing CF for students. However, as there are still some reservations regarding the clarity and accuracy of AWCF, the researcher sides with the suggestions of the previous research that AWCF should be an assistant teaching tool rather than essential. Therefore, teacher interference should always be present to clarify ambiguity and correct inaccurate suggestions. As far as the students are concerned, enabling Grammarly and utilising its different potentials would be a valuable suggestion. However, using the program should be associated with students' understanding of its limitations and their own repeated mistakes to focus on them and gain the most of the applications.

5. CONCLUSION

Many previous studies and theories found WCF beneficial for students' academic writing. However, there is no consensus on how WCF can foster writing, which aspects it develops, or whether the perceived improvement will remain in the long term. The present study attempted to contribute to filling such a gap by raising a null hypothesis

that Grammarly would not affect the written accuracy of students who used it for fourteen weeks. Adopting a pre and post-test, the researcher found that the performance of the experimental participants was significantly better than that of the control group members. Moreover, the researcher found a considerable drop in the number of errors in specific categories, namely article usage, subject-verb agreement, and, to a lesser extent, singular plural forms. Notwithstanding, no improvement regarding other types of error, verb tense and word choice, was found.

These results imply the importance of integrating Grammarly in teaching plans as an instructive complement and accustom EFL learners to use it in a broader range. The study suggests conducting longitudinal research to investigate Grammarly's impact in the long term. Also, it will be more advantageous to explore its effect in specific categories of grammatical errors and other aspects of writing, i.e., diction and mechanics.

REFERENCES

Barrot, Jessie. 2020. "Integrating Technology into ESL/EFL Writing through Grammarly." *RELC* 1-5. https://doi.org/10.1177/0033688220966632.

Beuningen, Catherine V. 2010. "Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions." *International journal of English studies* 10 (2): 1-27. http://dx.doi.org/10.6018/ijes.10.2.119171.

Bitchener, John. 2008. "Evidence in Support of written corrective feedback." *Journal of second language writing* 17: 102-118. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, John, and Dona R Ferris. 2012. *Written corrective feedback in second language acquisition and writing.* Routledge.

Burstein, Jill, Martin Chodorow, and Claudia Leacock. 2004. "Automated essay evaluation: The Criterion online writing Service." *AI Magazine* 25 (3): 27-36. https://doi.org/10.1609/aimag.v25i3.1774.

Chaudron, Craig. 1977. "A descriptive model of discourse in the corrective treatment of learners' errors." *Language Learning* 27: 29-46. https://doi.org/10.1111/j.1467-1770.1977.tb00290.x.

Coffin, Caroline, Mary Jane Curry, Sharon Goodman, Ann Hewings, Theresa M Lillis, and Joan Swan. 2003. *Teaching Academic Writing.* New York: Routledge.

Demirezen, Mehmet. 1988. "Behaviorist theory and language learning." *Journal of Hacettepe University Faculty of Education* (3): 135-140.

Dewi, Finita , and Anuncius Gumawang Jati. 2017. "The Effect of Three Different Types of Corrective Feedback on Students' Academic Writing in Higher Education." *Beyond Words* 5 (2): 106-114. https://doi.org/10.33508/bw.v5i2.1118.

Elliot, Norbert, Anne Ruggles Gere, Gail Gibson, Christie Toth, Carl Whithaus, and Amanda Presswood. 2013. "Uses and limitations of Automated Writing Evaluation software." *WPA-CompPile Research Bibliographies* 23: 1-7. https://comppile.org/wpa/bibliographies/Bib23/AutoWritingEvaluation.pdf.

Ellis, Rod. 2009. "Corrective feedback and teacher Development." *L2 Journal* 1 (1): 3-17. https://doi.org/10.5070/l2.v1i1.9054.

Ferris, Dona. 2006. "Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction." In *feedback in second language writing: Contexts and issues*, edited by K Hyland and F Hyland, 81-104. Cambridge University Press.

—. 2011. *Treatment of error in second language student writing.* 2nd. The University of Michigan.

Ghufron, M Ali, and Fathia Rosyida. 2018. "The role of Grammarly in assessing English as a foreign language EFL writing." *Lingua Cultura* 12 (4): 395-403. https://doi.org/10.21512/lc.v12i4.4582.

Grammarly. 2022. *About.* https://www.grammarly.com/about.

Guo, Quian, Ruiling Feng, and Yuanfang Hua. 2021. "How effectively can EFL students use automated written corrective feedback (AWCF) in research writing?" *Computer Assisted Language Learning* 1-20. https://doi.org/10.1080/09588221.2021.1879161.

Hinkel, Eli. 2004. *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar.* Mahwah, Lawrence Erlbaum Associates.

Hockly, Nicky. 2019. "Automated writing evaluation." *ELT journal* 73 (1): 82-88. https://doi.org/10.1093/elt/ccy044.

Huang, Qian. 2009. "Accuracy & fluency: Inspiration from error-correction of interlanguage theory." *Asian social sciences* 5 (2): 84-86.

Jang, Sung-Soo. 2020. "The efficacy of different types of metalinguistic information in L2 written corrective feedback." *English teaching* 75 (4): 33-56. https://doi.org/10.15858/engtea.75.4.202012.33.

Jim, Ranalli. 2018. "Automated written corrective feedback: How well can students make use of it?" *Computer Assisted Language Learning* 3 (7): 653-674. https://doi.org/10.1080/09588221.2018.1428994.

Kang, EunYoung, and Zhaohong Han. 2015. "The efficacy of written corrective feedback in improving written L2 accuracy: A meta-analysis." *The modern language journal* 99 (1): 1-18. https://doi.org/10.1111/modl.12189.

Krashen, Stephen. 1982. *Principles and practice in second language acquisition*. Pergamon Press Inc

Leong, Alvin Ping, and Hwee Hoon Lee. 2018. "From Both Sides of the Classroom: Perspectives on Teacher Feedback on Academic Writing and Feedback Practice." *The Journal of Teaching English for Specific and Academic Purposes* 6 (1): 151-164. https://doi.org/10.22190/JTESAP1801151L.

Li, Jinrong, Stephanie Link, and V Hegelheimer. 2015. "Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction." *Journal of second language writing* 27: 1-18. http://dx.doi.org/10.1016/j.jslw.2014.10.004.

Lightbown, Patsy M, and Nina Spada. 1990. "Focus-on-Form and Corrective Feedback in Communicative Language Teaching: Effects on Second Language Learning." *Studies in Second Language Acquisition* 12: 429-448. http://dx.doi.org/10.1017/S0272263100009517.

Long, M, H. 2000. "Focus in form in task-based language teaching." In *Language policy and pedagogy. Essays in honor of A. Ronald Walton,* edited by R Lambert and E Shohamy, 179-192. John Benjamins.

Luo, Yang, and Yan Liu. 2017. "Comparison between peer feedback and automated feedback in college English writing: A case study." *Open Journal of Modern Linguistics,* 7: 197-215. http://dx.doi.org/10.4236/ojml.2017.74015.

Maleki, A Ataollah, and Elham Eslami. 2013. "The effects of written corrective feedback techniques on EFL students' control over grammatical construction of their written English." *Theory and practice in language studies* 3 (7): 1250-1257. http://dx.doi.org/10.4304/tpls.3.7.1250-1257.

Miranty, Delsa, and Utami Widiati. 2021. "Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts." *Pegemjournal of education and instruction* 11 (4). https://doi.org/10.47750/pegegog.11.04.12.

Nova, Muhamad. 2018. "Utilising Grammarly in evaluating academic writing: A narrative research on EFL students' experience." *Journal of English Education and Applied Linguistics* 7 (1): 80-96.

O'Neill, Ruth, and Alex Russel. 2019. "Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly." *Australian Journal of Educational Technology* 35 (1): 42-56. https://doi.org/10.14742/ajet.3795.

Para , G Lorena, and Ximena S Calero. 2019. "Automated Writing Evaluation Tools in the Improvement of the Writing Skill." *International Journal of Instruction* 12 (2): 209-226. https://doi.org/10.29333/iji.2019.12214a.

Qassemzadeh, Abolfazl, and Hassan Soleimani. 2016. "The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners." *Theory and practice in language studies* 6 (9): 1884-1894. http://dx.doi.org/10.17507/tpls.0609.23.

Raitskaya, Lilia , and Elena Tikhonova. 2022. "An In-Depth Glimpse into Research on Academic Writing." Journal of Language and Education 8 (2): 5-18. doi: https://doi.org/10.17323/jle.2022.14586.

Respondus. 2021. *Lockdown Browser*. https://web.respondus.com/he/lockdownbrowser/.

Rudner, Lawrence, and Tahung Liang. 2002. "Automated essay scoring using Bayes' theorem." *The Journal of technology, learning, and assessment* 1: 1- 21. https://ejournals.bc.edu/index.php/jtla/article/view/1668/1512

Touchie, Hanna Y. 1986. "Second language learning errors: Their types, causes, and treatment." *JALT journal* 8 (1): 75-80. https://jalt-publications.org/sites/default/files/pdf-article/art5_8.pdf

Truscott, John. 1996. "The case against grammar correction in L2 writing classes." *Language learning* 46: 327-369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x.

—. 1999. "The Case for "The case against grammar correction in L2 writing classes": A Response to Feris." *Journal of second language learning* 8 (2): 111-122.

—. 2020. "The Efficacy of written corrective feedback: a critique of a meta-analysis." *Researchgate*.https://www.researchgate.net/publication/341453398_The_Efficacy_of_Written_Corrective_Feedback_A_Critique_of_a_Meta-analysis.

Utkina, Tatiana. 2021. "Teaching academic writing in English to students of economics through conceptual metaphors." *The Journal of Teaching English for Specific and Academic Purposes* 9 (4): 587-599. https://doi.org/10.22190/JTESAP2104587U.

Wahyuningsih, Sri. 2020. "The role of corrective feedback on academic writing performance: EFL students' perceptions." *Edulingua: Jurnal Linguistiks Terapan dan Pendidikan Bahasa Inggris* 7 (1): 13-21.

Wirantaka, Andi. 2019. "Investigating Written Feedback on Students' Academic Writing." *Advances in Social Science, Education and Humanities Research* 353: 1-7.

Woodworth, Johnathan, and Khaled Barkaoui. 2020. "Perspectives on Using Automated Writing Evaluation Systems to Provide Written Corrective Feedback in the ESL Classroom." *TESL Canada journal* 37 (2): 234-247. https://doi.org/10.18806/tesl.v37i2.1340.