# Grammarly in Teaching Writing to EFL Learners at Low Levels: How Useful Is It?

Mohammed Nour Abu Guba[1], Asmaa Awad[1], & Abdallah Abu Qub'a[2]

[1] The University of Sharjah, United Arab Emirates

[2] King Faisal University, Saudi Arabia

Correspondence: Abdallah Abu Qub'a, King Faisal University, Saudi Arabia. E-mail: aabuquba@kfu.edu.sa

Received: December 4, 2023 Accepted: January 12, 2024 Online Published: February 6, 2024

doi:10.5430/wjel.v14n3p1 URL: https://doi.org/10.5430/wjel.v14n3p1

**Abstract**

This study aims at identifying the extent to which Grammarly can improve low-level EFL learners' writing and the types of errors that Grammarly can help learners avoid most. Two groups of low-level English as a foreign language learners (30 female students each) taught by the same teacher participated in this study. Errors made by the two groups in a pretest and a posttest were identified and compared. The results show that the experimental group performed better than the control group in the posttest, which suggests that Grammarly can generally help low-level students improve their writing skills. Findings also show that some errors were more amenable to correction and feedback whereas others, namely grammatical errors related to word forms and word usage, were resistant, and no noticeable improvement was attested.

**Keywords:** Grammarly, writing, English as a foreign language, Automatic Writing Evaluation

## 1. Introduction

Grammarly is an Automatic Writing Evaluation (AWE) program that has become very widespread among users of English (e.g., Vernon, 2000; Bigert, 2004; Chen, 2008; Potter & Fuller, 2008; Burston, 2013; Radi, 2014;). This is evident from the large number of Grammarly subscribers reaching 30 million people and 30, 000 teams using it daily to strengthen their writing (Grammarly Inc. 2021). Many learners of English as a foreign/second language (EFL learners) use this tool especially over the past few years when online education has become very common due to the COVID-19 pandemic.

Grammarly, founded in 2009 as an online grammar checker, aims to help writers improve their grammar knowledge and writing skills (Moore, 2018). It has many advantages for learners in that it provides direct feedback by identifying mistakes and suggesting corrections immediately. It also increases writers' confidence as they do not feel shy or afraid of committing errors (Lailika, 2019).

Many studies found that using Grammarly was beneficial to its users, who were at advanced or intermediate levels (Qassemzadeh & Soleimani, 2016; Ghufron & Rosyida, 2018; Russell & O'Nell, 2018; among others); however, we still do not know the extent to which Grammarly can be useful to EFL learners at low levels (beginner and pre-intermediate levels), as their low level might not enable them to understand and benefit from the feedback provided to them. Is using such a tool worth the time, effort, and money spent on it? Note that Grammarly comes in two versions: one is free and offers to check 150 types of errors; the other is the premium subscription, and it checks 400 types of errors. This study deals with the free version as all the participants in this study had access to the free one only.

Arab learners of English face many difficulties in learning English (cf. Abu Guba, 2023; Abu Guba, Mashaqba, Jarbou & Al-Haj Eid, 2023; Abu Guba, Fareh & Yagi, 2023). It is assumed that a software such as Grammarly can be of great help and benefit to learners of English in general. However, it is still not established whether such a tool can be effective for low-level learners, Therefore, this study aims to explore how much and in what ways Grammarly can help low-level EFL learners. More specifically, the current study attempts to answer the following questions:

1. To what extent can Grammarly help improve low-level EFL learners' writing?
2. Which types of errors does Grammarly help learners avoid most?

In the next section, we review relevant literature and show that no studies have focused on the use of Grammarly by EFL learners at low levels. In section 3, we lay out the methods and in section 4 we identify the errors and the degree of their responsiveness to Grammarly feedback. We conclude the study with a discussion of the results and recommendations to EFL learners and teachers.

## 2. Literature Review

The literature on using AWE programs has focused on intermediate and advanced users of English, while little is known about using them among lower-level EFL learners. Past research focused on the user's experiences of Grammarly, and the types of errors detected by Grammarly (e.g., Daniels & Leslie, 2013; Schraudner, 2014; Cavaler & Dianati, 2016; Qassemzadeh & Soleimani, 2016; Jayavalan &

Razali, 2018; Karyuatry et al., 2018; Nova, 2018), and comparisons of the feedback given by Grammarly and writing centers (e.g., Dembsey, 2017; Russell & O'Nell, 2019).

Comparing Grammarly with other spelling and grammar software, Daniels and Leslie (2013) reported that Grammarly was not very useful to learners as it did not suggest corrections of the errors and its explanations were complex. Knowing that such technology-based engines are always developing, it remains to be discovered if it is still the case a decade later. Similarly, Hoang and Kunnan (2016), in their study of the effectiveness of 'MY Access' automated writing program (which is similar to Grammarly), stated that the grammar feedback was designed for English native speaker student writers because non-native speakers may sometimes not understand the suggestions given by the program due to language barriers. This is expected as lack of interaction with the program is problematic especially for low-level students, who need clear instructions, more exercises, and model examples to improve their writing skills.

On the other hand, Qassemzadeh and Soleimani (2016) studied the effect of the feedback given by Grammarly and the feedback given by teachers on learning passive structures by 70 intermediate male and female EFL Iranian learners. Post-test results showed that learning passive structures using Grammarly was more effective than learning passive structures by teachers. Cavaleri and Dianati (2016) investigated students' perceptions of Grammarly in an Australian university. Overall, students believed that Grammarly helped them understand grammar rules and provided them with direct interactivity. The same results were obtained by Yulianti and Reni (2018), who explored the use of Grammarly in teaching writing. They reported that students had positive reactions toward using Grammarly, which was helpful in improving their writing skills. Nova's (2018) study aimed to find the strengths and weaknesses of Grammarly in evaluating the academic writing of three postgraduate Indonesian students. Results of interviewing the students to evaluate their experience with the program showed that the program had many strengths such as useful feedback with explanations and examples, ease of access, and high speed. On the other hand, some weaknesses such as misleading feedback and lack of content and context evaluation were reported.

Research comparing feedback and the usefulness of AWE software with human ones yielded mixed results. Dembsey (2017) compared the feedback provided by Grammarly on three freshmen's essays with the feedback provided by the writing center by ten consultants. Results revealed that Grammarly provided twice more feedback than the consultants because of repeated comments on repeated errors, while the consultants covered more issues in their feedback than Grammarly. However, neither Grammarly nor the consultants provided 100% accurate feedback because both used inaccurate and complex terms which were unfamiliar to most students especially when explaining the sentence-level issues. Dembsey concluded that the writing center consultants' feedback was better because of its being trainable, interactive, and flexible, unlike that of Grammarly, which seemed to be inflexible and driven by algorithms.

Ghufron and Rosyida (2018) compared the effectiveness of using Grammarly feedback and teachers' corrective feedback in reducing writing errors made by 40 Indonesian university students in terms of diction, grammar, spelling, and punctuation. The results showed that students who used Grammarly had a significant reduction in writing errors than those who were evaluated by teachers.

Russell and O'Nell (2018) examined the perceptions of 54 Australian students who received grammar feedback from Grammarly and 42 students who received feedback from academic advisors. The results indicated that students who received feedback from Grammarly responded more positively to 9 out of 15 survey items such as the amount of time spent, short and long-term benefits in student's writing, the ability to proofread independently with great confidence, speed of response, and range of errors addressed. In addition, students were more satisfied with the advice they received from Grammarly.

More recently, Zhang, Özer, and Bayazeed (2020) examined the perceptions of 42 international students at a USA university using Grammarly compared with face-to-face teaching. Results showed that students believed Grammarly to have many advantages such as being easy to use, giving helpful suggestions, providing instant responses, and being accessible, and quick. These reasons made students prefer using Grammarly to visiting the writing centers even though the students were neutral about the ability of Grammarly to help them to improve their writing skills or their grades.

To sum up, earlier studies generally show that their participants were satisfied with Grammarly and responded more positively to the feedback given by Grammarly. All the previous studies focused on intermediate or advanced students, who have a good command of the English language and who can read and understand the long explanations given by Grammarly. None of the studies has dealt with low-level students' experience with Grammarly. This study aims to fill this gap by analyzing how Grammarly can help improve low-level EFL learners' writing and the extent to which Grammarly can be useful to them.

## 3. Methods

### *3.1 Participants*

Two groups of low-level EFL learners participated in this study. Each group comprised 30 female students (mean age= 18.4) enrolled in an intensive English 3.5-month course at a university in the UAE. The course was a 25-hour per week course, of which 5 hours were devoted to a writing class. Immediately before the start of this course, the students sat for an IELTS exam or the Emirates Standardized Test in English (EMSAT)/or TOEFL ITP and received an overall score of 4 and below in the IELTS (the same score was obtained in the writing section) or below 825 in the EMSAT or below 380 in TOEFL. These scores are equivalent to A2 in the Common European Framework. The two groups were taught by the same teachers for one semester, including the writing teacher.

### *3.2 Procedure*

At the beginning of the course, the two groups were asked to write a short opinion essay on online learning (about 250 words in one hour)

in English as a pretest. The test was delivered on Blackboard with the lockdown browser enabled to monitor students during the test so that they do not receive any kind of help that may affect their true level in writing. The test was analyzed by Grammarly, and the two groups received comparable scores with no significant differences between the two groups as revealed by an ANOVA test (p= 0.655). The two groups were assigned (almost every week), two writing tasks submitted online as part of the assessment in the writing course. The total number of assignments was 20. At the end of the course, the same groups were asked to write an opinion essay on using mobile phones in class as a posttest (under the same conditions applied in the pretest) Note that the two topics in the test were similar to minimize the effect of external variables as much as possible.

One of the groups (the experimental group) was asked to use Grammarly before submitting each of the 20-course assignments. As evidence of using Grammarly, the students were asked to attach the Grammarly report along with the assignment. The other group did not use Grammarly, and this was further verified at the end of the course by asking the students if they had used Grammarly during the course.

The researchers analyzed the two groups' writing using Grammarly at the beginning of the course (pretest) and at the end of the course (posttest). Additionally, all types of errors were identified and categorized into their respective groups to find out which types of writing errors were more responsive to correction. The researchers also compared the overall score received by the two groups. This is a score between 0 and 100 given by Grammarly to every piece of writing. It calculates the accuracy level of a piece of writing based on the number of words and the number and types of writing errors. It is worth mentioning that although the students in the experimental group were trained on how to use Grammarly and understand the errors and feedback once at the beginning of the course, no discussion of errors or feedback given by Grammarly was held afterwards.

Grammarly provides feedback on several types of errors such as 'correctness', 'clarity', and 'engagement'. The current study restricts itself to correctness errors. This is because these errors are easily identified and understood by lower-level students and consequently seem to be more amenable to feedback. Second, these are the types of errors that Grammarly identifies and analyzes for free. Other types are available to premium subscribers only.

The performance of the two groups in the pretest and the posttest was compared using a mixed model ANOVA, with each type of error (on its own) as the dependent factor, the group (2 groups), and time (pretest vs posttest) as the between-subjects factors, controlling for 'participant' to limit the effect of within-subject variance. All statistical tests were run at a .05 alpha level. Also, a mixed model repeated measures ANOVA was conducted to find out the differences in the two groups' performance as a whole.

**4. Results**

First, we give information on the type of errors identified by Grammarly, along with explanations and examples (presented in Table 1). We then compare the performance of each group on each type of error. In general, we notice that some errors were more amenable to correction and feedback whereas others were resistant, and no noticeable improvement was attested.

Table 1. Correctness errors

| Error type | Explanation | Examples from students' essays |
|---|---|---|
| Formatting errors | Errors relating to improper formatting. Most errors here relate to spacing, e.g., no space after a period. | There are two advantages for using mobiles.The speed…. |
| Spelling errors | These relate to spelling mistakes. | - Many students use mobailes.<br>- Online lerning is common. |
| Punctuation and capitalization errors | For example, missing a comma after a dependent clause | - If students send their assignments online they will save time.<br>- students will learn a lot. |
| Grammatical errors | Below are the categories of grammatical mistakes committed by the participants | |
| Faulty subject-verb agreement | For example, using a singular subject instead of plural | - There is many advantages for using mobiles in class. (are)<br>- Online learning give us many chances to learn. |
| Word forms | Using the wrong form/part of speech of a word. For example, using the infinitive with to after 'let', and using incorrect noun number | - Online learning let us to see our weakness in using technology.<br>- There are many type of mobiles. (types) |
| Prepositions | Wrong or missing prepositions, e.g., using 'of' instead of 'in' | - Some students are interested of online learning.<br>- Students become good in learning many things. |
| Determiners and articles | Using determiners/articles (a/an/the/this, etc.) incorrectly. | - Students can use laptop in class. (The indefinite article 'a' should be added)<br>- Students can learn from internet. |
| Word usage | This relates to missing/using words incorrectly, e.g., misusing the quantifier 'many' with uncountable nouns, missing relative pronouns, and confusing words ('there' instead of 'their') | - Some students need many money to buy laptops…..(much)<br>- There are many students use their mobiles in class. (a pronoun is missing here)<br>- Mobiles help students improve there learning skills. |

*4.1 Formatting Errors*

The two groups performed similarly in terms of formatting errors in the pretest, with a mean around 10 of formatting errors for both groups. The number of errors dropped to 2.2 in the posttest for the experimental group but decreased only to 9.2 for the control group (Figure 1).

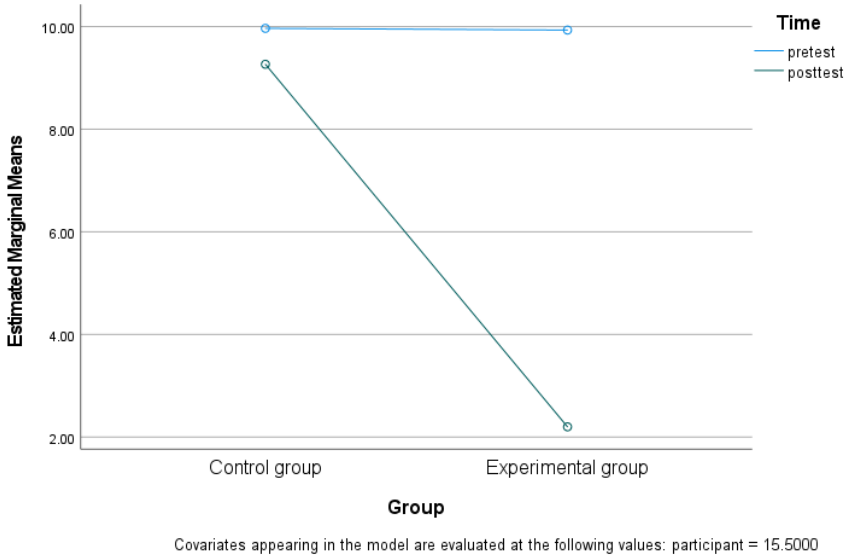


Figure 1. Means of formatting errors

A mixed model ANOVA (Table 2), controlling for 'participant', revealed a significant main effect for the interaction between 'group' (experimental and control) and 'time' (pretest and posttest) ($F(1, 115) = 106.274$, $P=.001$, $\eta p2 =.48$, observed power=1, R-squared= .75). There was also a significant main effect for 'group' ($F(1, 115) =108.299$, $p=001$, $\eta p2 = .48$, observed power = 1) and for 'time' ($F(1, 115) = 152.793$, $p=001$, $\eta p2 = .57$, power=1). No statistically significant differences between participants were attested ($p= .926$).

Table 2. ANOVA results of formatting errors

Dependent variable Formatting errors

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 1282.522[a] | 4 | 320.630 | 91.844 | <.001 | .762 | 367.375 | 1.000 |
| Intercept | 1741.323 | 1 | 1741.323 | 498.798 | <.001 | .813 | 498.798 | 1.000 |
| participant | .030 | 1 | .030 | .009 | .926 | .000 | .009 | .051 |
| Group | 378.075 | 1 | 378.075 | 108.299 | <.001 | .485 | 108.299 | 1.000 |
| Time | 533.408 | 1 | 533.408 | 152.793 | <.001 | .571 | 152.793 | 1.000 |
| Group * Time | 371.008 | 1 | 371.008 | 106.274 | <.001 | .480 | 106.274 | 1.000 |
| Error | 401.470 | 115 | 3.491 | | | | | |
| Total | 9063.000 | 120 | | | | | | |
| Corrected Total | 1683.992 | 119 | | | | | | |

a. R Squared = .762 (Adjusted R Squared = .753)
b. Computed using alpha = .05

*4.2 Spelling Errors*

The experimental group did slightly better than the control group in the pretest (with a mean difference of less than one error (means were 11.8 and 12.7, respectively). In the posttest, the experimental group outperformed the control group with a difference of more than 4 errors less (3.3 and 8.1 errors, respectively) (Figure 2).

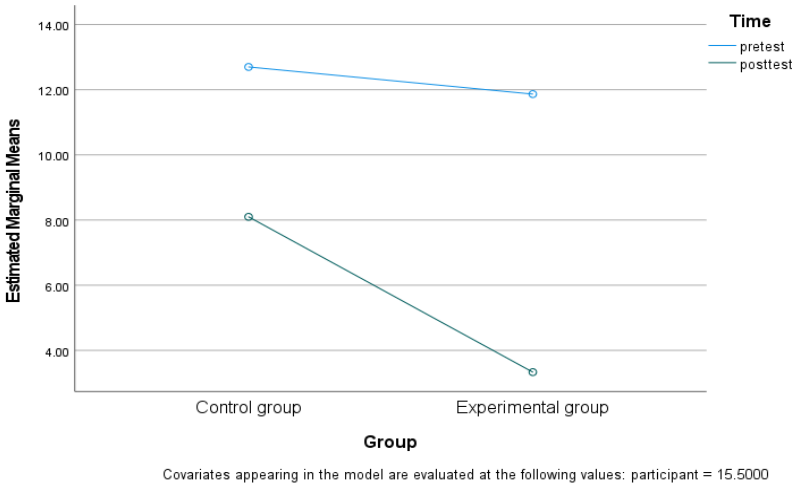


Figure 2. Means of spelling errors

A mixed model ANOVA (Table 3), controlling for 'participant', revealed a significant main effect for the interaction between 'group' and 'time' (F(1, 115) = 25.532, P=.001, ηp2 =.18, observed power=1, R-squared= .75). There was also a significant main effect for 'group' (F (1, 115) =51.753, p=001, ηp2 = .31, observed power = 1) and for 'time' (F (1, 115) = 284.647, p=001, ηp2 = .72, power=1).

Table 3. ANOVA results of spelling errors

Dependent variable: Spelling errors

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 1649.361[a] | 4 | 412.340 | 90.730 | <.001 | .759 | 362.920 | 1.000 |
| Intercept | 2135.983 | 1 | 2135.983 | 469.995 | <.001 | .803 | 469.995 | 1.000 |
| participant | 4.494 | 1 | 4.494 | .989 | .322 | .009 | .989 | .167 |
| Group | 235.200 | 1 | 235.200 | 51.753 | <.001 | .310 | 51.753 | 1.000 |
| Time | 1293.633 | 1 | 1293.633 | 284.647 | <.001 | .712 | 284.647 | 1.000 |
| Group * Time | 116.033 | 1 | 116.033 | 25.532 | <.001 | .182 | 25.532 | .999 |
| Error | 522.639 | 115 | 4.545 | | | | | |
| Total | 11892.000 | 120 | | | | | | |
| Corrected Total | 2172.000 | 119 | | | | | | |

a. R Squared = .759 (Adjusted R Squared = .751)
b. Computed using alpha = .05

*4.3 Punctuation and Capitalization Errors*

Clear differences in the number of punctuation and capitalization errors were obtained in the posttest between the two groups with a mean difference of 4 errors (3.2 for the experimental group and 7.2 for the control group). This is in sharp contrast to the groups' performance in the pretest with a difference of less than one error in favor of the experimental group (Figure 3).

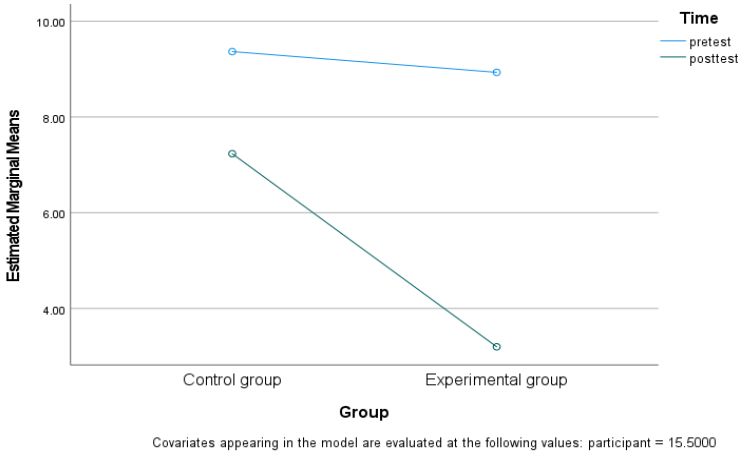


Figure 3. Means of punctuation errors

A mixed model ANOVA (Table 4), controlling for 'participant', revealed a significant main effect for the interaction between 'group' and 'time' (F(1, 115) = 42.380, P=.001, ηp2 =.27, observed power=1, R-squared= .73). The effects of 'group' and 'time' were also significant (F(1, 115) =65.241, p=001, ηp2 = .36, observed power = 1) and (F (1, 115) = 202.365, p=001, ηp2 = .64, power=1), respectively.

Table 4. ANOVA results of punctuation errors

Dependent variable: Punctuation

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 732.209[a] | 4 | 183.052 | 79.812 | <.001 | .735 | 319.248 | 1.000 |
| Intercept | 1179.304 | 1 | 1179.304 | 514.184 | <.001 | .817 | 514.184 | 1.000 |
| participant | 21.242 | 1 | 21.242 | 9.262 | .003 | .075 | 9.262 | .855 |
| Group | 149.633 | 1 | 149.633 | 65.241 | <.001 | .362 | 65.241 | 1.000 |
| Time | 464.133 | 1 | 464.133 | 202.365 | <.001 | .638 | 202.365 | 1.000 |
| Group * Time | 97.200 | 1 | 97.200 | 42.380 | <.001 | .269 | 42.380 | 1.000 |
| Error | 263.758 | 115 | 2.294 | | | | | |
| Total | 7188.000 | 120 | | | | | | |
| Corrected Total | 995.967 | 119 | | | | | | |

a. R Squared = .735 (Adjusted R Squared = .726)

b. Computed using alpha = .05

*4.4 Grammatical Errors*

First, we present each category of grammatical errors separately, and then we report the results of all grammatical errors combined.

4.4.1 Subject-verb Agreement Errors

As Figure 4 shows, although the control group made slightly fewer subject-verb agreement errors than the experimental group in the pretest (6.4 and 6.6, respectively), the experimental group performed much better than the control group in the posttest (2.8 vs. 4.8 errors).

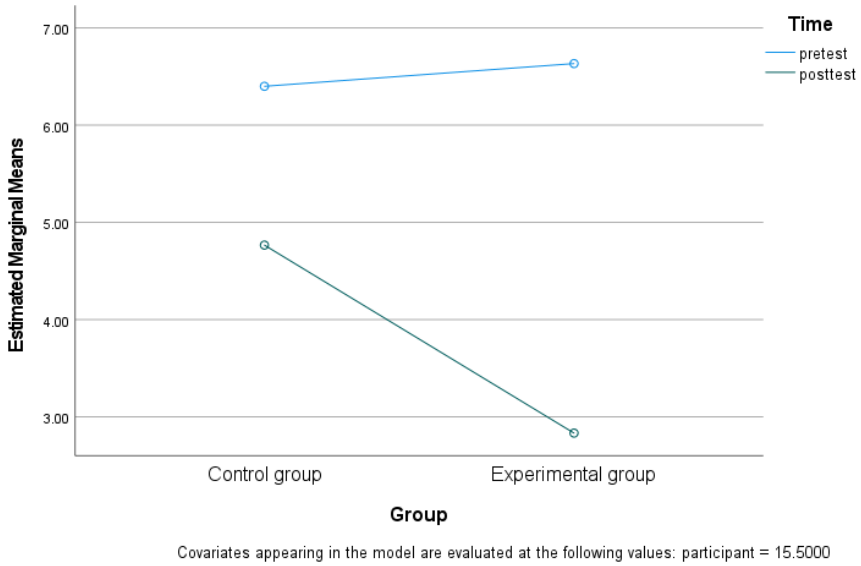


Figure 4. Means of SV agreement errors

Statistically, a mixed model ANOVA (Table 5), controlling for 'participant', yielded a significant main effect, with a weaker effect size than those of other types of errors, for the interaction between 'group' and 'time' (F(1, 115) = 23.122, P=.001, ηp2 =.17, observed power=1, R-squared= .60). The effect of 'group' was also significant (F(1, 115) =14.234, p=001, ηp2 = .11, observed power = .96), and the effect of 'time' was significant (F(1, 115) = 145.401, p=001, ηp2 = .56, power=1).

Table 5. ANOVA results of SV agreement errors

Dependent variable: SV agreement

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 278.876[a] | 4 | 69.719 | 45.785 | <.001 | .614 | 183.141 | 1.000 |
| Intercept | 796.224 | 1 | 796.224 | 522.889 | <.001 | .820 | 522.889 | 1.000 |
| participant | .585 | 1 | .585 | .384 | .537 | .003 | .384 | .094 |
| Group | 21.675 | 1 | 21.675 | 14.234 | <.001 | .110 | 14.234 | .963 |
| Time | 221.408 | 1 | 221.408 | 145.401 | <.001 | .558 | 145.401 | 1.000 |
| Group * Time | 35.208 | 1 | 35.208 | 23.122 | <.001 | .167 | 23.122 | .998 |
| Error | 175.115 | 115 | 1.523 | | | | | |
| Total | 3647.000 | 120 | | | | | | |
| Corrected Total | 453.992 | 119 | | | | | | |

a. R Squared = .614 (Adjusted R Squared = .601)
b. Computed using alpha = .05

**Word form errors**

Figure 5 shows that no substantial differences were obtained in the pretest between the two groups in the number of word-form errors, with a mean around 6.5 errors across the two groups (6.7 for the control group and 6.5 for the experimental group). Likewise, the two groups' performance in the posttest was similar (6.5 for the control group and 6.3 for the experimental group). Improvement in the posttest across the two groups was minimal.

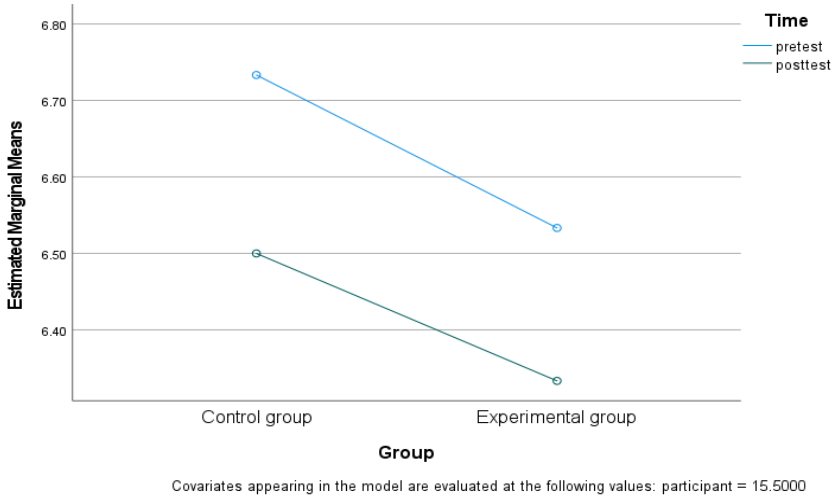


Figure 5. Means of word form errors

Although the two groups performed slightly better in the posttest, none of the differences was statistically significant, as revealed by mixed model ANOVA test (Table 6), group (F(1, 115) =1.585, p= .211, ηp2 = .014, observed power = .24, R-squared=.003), time (F(1, 115) =2.214, p=.139, ηp2 = .019, observed power = .31), and group x time (F(1, 115) =.013, p=.909, ηp2 = .000, observed power = .05).

Table 6. ANOVA results of word form errors

Dependent variable: Word forms

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 2.780[a] | 4 | .695 | 1.093 | .364 | .037 | 4.371 | .335 |
| Intercept | 1250.985 | 1 | 1250.985 | 1966.825 | <.001 | .945 | 1966.825 | 1.000 |
| participant | .355 | 1 | .355 | .558 | .456 | .005 | .558 | .115 |
| Group | 1.008 | 1 | 1.008 | 1.585 | .211 | .014 | 1.585 | .239 |
| Time | 1.408 | 1 | 1.408 | 2.214 | .139 | .019 | 2.214 | .314 |
| Group * Time | .008 | 1 | .008 | .013 | .909 | .000 | .013 | .051 |
| Error | 73.145 | 115 | .636 | | | | | |
| Total | 5185.000 | 120 | | | | | | |
| Corrected Total | 75.925 | 119 | | | | | | |

a. R Squared = .037 (Adjusted R Squared = .003)
b. Computed using alpha = .05

4.4.2 Preposition Use Errors

The performance of the two groups in the pretest was exactly the same (Mean= 6.7 errors). By contrast, clear differences were attested in the posttest (Figure 6), with the experimental group outperforming the control group, with a mean difference of 1.5 errors (5.1 for the control group and 3.4 for the experimental group).

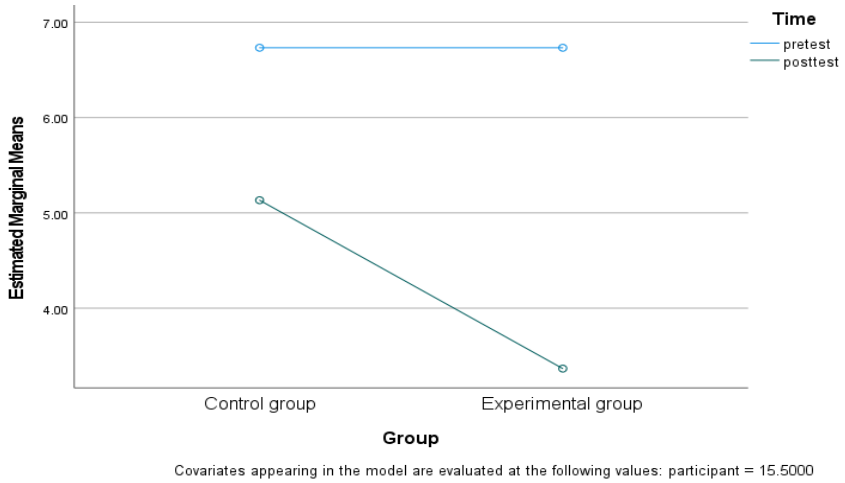


Figure 6. Means of preposition use errors

ANOVA results (Table 7), controlling for 'participant', revealed a significant main effect for the interaction between 'group' and 'time' (F(1, 115) = 34.127, P=.001, ηp2 =.23, observed power=1, R-squared= .74). The effects of 'group' and 'time' were also significant (F(1, 115) =34.127, p=001, ηp2 = .23, observed power = 1) and (F(1, 115) = 269.721, p=001, ηp2 = .70, power=1), respectively.

Table 7. ANOVA results of preposition use errors

Dependent variable: Prepositions

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 233.110[a] | 4 | 58.278 | 84.962 | <.001 | .747 | 339.849 | 1.000 |
| Intercept | 803.168 | 1 | 803.168 | 1170.930 | <.001 | .911 | 1170.930 | 1.000 |
| participant | 1.285 | 1 | 1.285 | 1.874 | .174 | .016 | 1.874 | .274 |
| Group | 23.408 | 1 | 23.408 | 34.127 | <.001 | .229 | 34.127 | 1.000 |
| Time | 185.008 | 1 | 185.008 | 269.721 | <.001 | .701 | 269.721 | 1.000 |
| Group * Time | 23.408 | 1 | 23.408 | 34.127 | <.001 | .229 | 34.127 | 1.000 |
| Error | 78.881 | 115 | .686 | | | | | |
| Total | 3931.000 | 120 | | | | | | |
| Corrected Total | 311.992 | 119 | | | | | | |

a. R Squared = .747 (Adjusted R Squared = .738)
b. Computed using alpha = .05

4.4.3 Determiner and Article Errors

As seen in Figure 7, the experimental group did slightly better than the control group in terms of using determiners and articles correctly in the pretest (7.7 vs. 8.1 errors, respectively). In the posttest, the performance of the control group was almost the same, with a mean of 6.5 mistakes, whereas the performance of the experimental group was much better, with about 3 errors less in the posttest (M=3.7 mistakes).

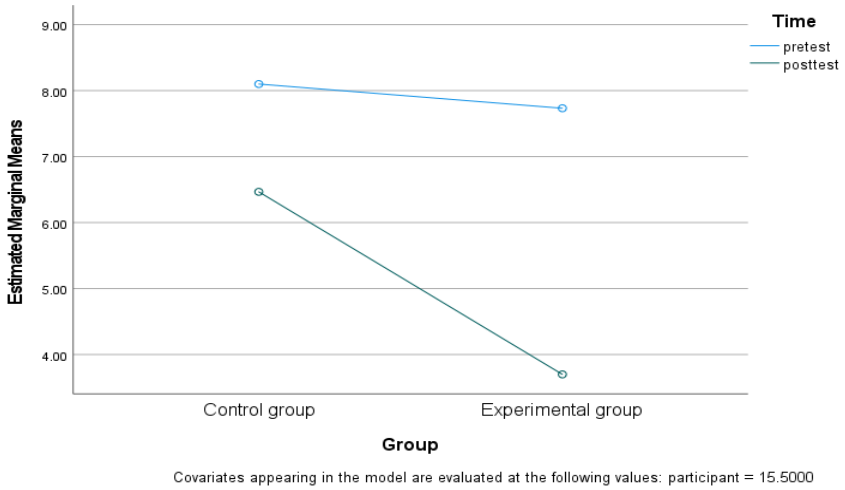


Figure 7. Means of determiner and article errors

All the differences between the two groups were statistically significant (Table 8). The interaction between 'group' and 'time' (F(1, 115) = 42.534, P=.001, ηp2 =.27, observed power=1, R-squared= .75) was significant. Also, the effects of 'group' and 'time' were significant (F(1, 115) =72.497, p=001, ηp2 = .39, observed power = 1) and (F(1, 115) = 237.118, p=001, ηp2 = .67, power=1), respectively.

Table 8. ANOVA results of determiner and article errors

Dependent variable: Determiners

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 375.198[a] | 4 | 93.800 | 92.353 | <.001 | .763 | 369.411 | 1.000 |
| Intercept | 1472.348 | 1 | 1472.348 | 1449.636 | <.001 | .927 | 1449.636 | 1.000 |
| participant | 17.532 | 1 | 17.532 | 17.261 | <.001 | .131 | 17.261 | .985 |
| Group | 73.633 | 1 | 73.633 | 72.497 | <.001 | .387 | 72.497 | 1.000 |
| Time | 240.833 | 1 | 240.833 | 237.118 | <.001 | .673 | 237.118 | 1.000 |
| Group * Time | 43.200 | 1 | 43.200 | 42.534 | <.001 | .270 | 42.534 | 1.000 |
| Error | 116.802 | 115 | 1.016 | | | | | |
| Total | 5562.000 | 120 | | | | | | |
| Corrected Total | 492.000 | 119 | | | | | | |

a. R Squared = .763 (Adjusted R Squared = .754)
b. Computed using alpha = .05

4.4.4 Word Usage Errors

As can be seen from Figure 8, the performance of the two groups was alike in the pretest, with a .03 difference. The same trend held for the posttest; the experimental group made 5.8 word-usage mistakes, while the control group made 5.83 mistakes.

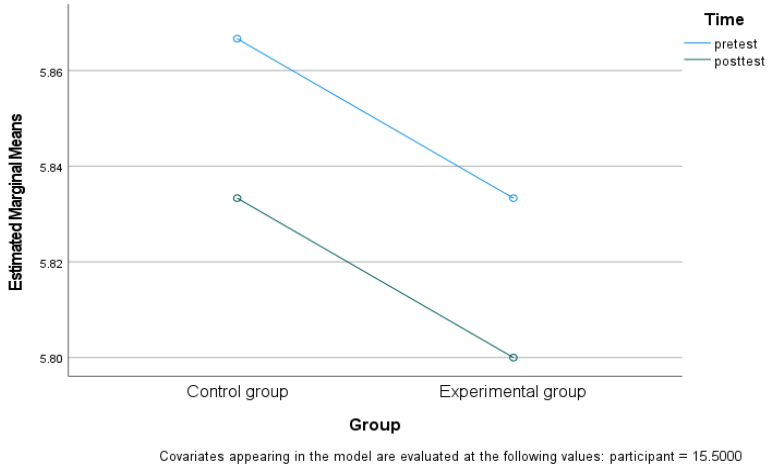


Figure 8. Means of word usage errors

As can be seen from Table 9, no statistically significant differences were attested (group x time (F(1, 115) = .000, P= 1, ηp2 =.000, observed power=.05, R -= .017), group (F(1, 115) = .071, P=.790, ηp2 =.001, observed power=.06), and time (F(1, 115) = .071, P=.790, ηp2 =.001, observed power=.06)).

Table 9. ANOVA results of usage errors

Dependent variable: Word usage

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 2.808[a] | 4 | .702 | 1.499 | .207 | .050 | 5.997 | .452 |
| Intercept | 1062.800 | 1 | 1062.800 | 2269.331 | <.001 | .952 | 2269.331 | 1.000 |
| participant | 2.742 | 1 | 2.742 | 5.854 | .017 | .048 | 5.854 | .670 |
| Group | .033 | 1 | .033 | .071 | .790 | .001 | .071 | .058 |
| Time | .033 | 1 | .033 | .071 | .790 | .001 | .071 | .058 |
| Group * Time | .000 | 1 | .000 | .000 | 1.000 | .000 | .000 | .050 |
| Error | 53.858 | 115 | .468 | | | | | |
| Total | 4140.000 | 120 | | | | | | |
| Corrected Total | 56.667 | 119 | | | | | | |

a. R Squared = .050 (Adjusted R Squared = .017)
b. Computed using alpha = .05

4.4.5 All Grammatical Errors

The fact that two subtypes of grammatical errors were not statistically different between the two groups necessitated comparing all the grammatical groups together to find out whether the effect of the two non-significant differences cancels out the effect of the three significant differences. Mixed-model repeated measures ANOVA tests (given in Table 10), controlling for 'participant', revealed a statistically significant main effect for the interaction of group and time (F(1, 115) = 46.262, P=.001, ηp2 =.29, observed power= 1), meaning that Grammarly did affect the performance of the experimental group positively. Also, the main effect of group was statistically significant (F(1, 115) = 59.337, P=.001, ηp2 =.34, observed power= 1), and that of time was significant (F(1, 115) = 453.961, P=.001, ηp2 =.80, observed power= 1).

Table 10. ANOVA results of all grammatical errors

Measure: MEASURE_1
Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[a] |
|---|---|---|---|---|---|---|---|---|
| Intercept | 5172.661 | 1 | 5172.661 | 5210.476 | <.001 | .978 | 5210.476 | 1.000 |
| participant | 4.728 | 1 | 4.728 | 4.763 | .031 | .040 | 4.763 | .581 |
| Group | 58.907 | 1 | 58.907 | 59.337 | <.001 | .340 | 59.337 | 1.000 |
| Time | 450.667 | 1 | 450.667 | 453.961 | <.001 | .798 | 453.961 | 1.000 |
| Group * Time | 45.927 | 1 | 45.927 | 46.262 | <.001 | .287 | 46.262 | 1.000 |
| Error | 114.165 | 115 | .993 | | | | | |

a. Computed using alpha = .05

4.4.6 All Types of Errors

Taking all types of errors together, statistically significant differences were found between the two groups. A mixed-model repeated measures ANOVA (presented in Table 11), controlling for 'participant', revealed a significant main effect for the interaction between 'group' and 'time' (F(1, 115) = 171.914, P=.001, ηp2 =.60, observed power=1). Also, significant main effects were attested for group (F(1, 115) =233.662, p=001, ηp2 = .67, observed power = 1), and 'time' F(1, 115) =930.662, p=001, ηp2 = .89, power=1). No main effect was found for the differences between participants (F(1, 115) = .236, P=.628, ηp2 =.002, observed power=.077).

Table 11. ANOVA results of all types of errors

Measure: MEASURE_2
Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[a] |
|---|---|---|---|---|---|---|---|---|
| Intercept | 10018.738 | 1 | 10018.738 | 4547.377 | <.001 | .975 | 4547.377 | 1.000 |
| participant | .521 | 1 | .521 | .236 | .628 | .002 | .236 | .077 |
| Group | 514.801 | 1 | 514.801 | 233.662 | <.001 | .670 | 233.662 | 1.000 |
| Time | 2050.426 | 1 | 2050.426 | 930.662 | <.001 | .890 | 930.662 | 1.000 |
| Group * Time | 378.759 | 1 | 378.759 | 171.914 | <.001 | .599 | 171.914 | 1.000 |
| Error | 253.367 | 115 | 2.203 | | | | | |

a. Computed using alpha = .05

These results show that the experimental group benefited from using Grammarly for all types of errors except for two subtypes of grammatical errors, namely errors related to word forms and word usage.

4.4.7 Overall Score

Both groups had similar pretest overall scores (23.6 for the experimental group and 22.8 for the control group). By contrast, the differences in the posttest increased to just under 5, with the experimental group scoring as high as 38.2. Although this does not seem to be a high score in Grammarly, it is a significant improvement (Figure 9), as also demonstrated by the results of the mixed-model ANOVA test as can be seen in Table 12.

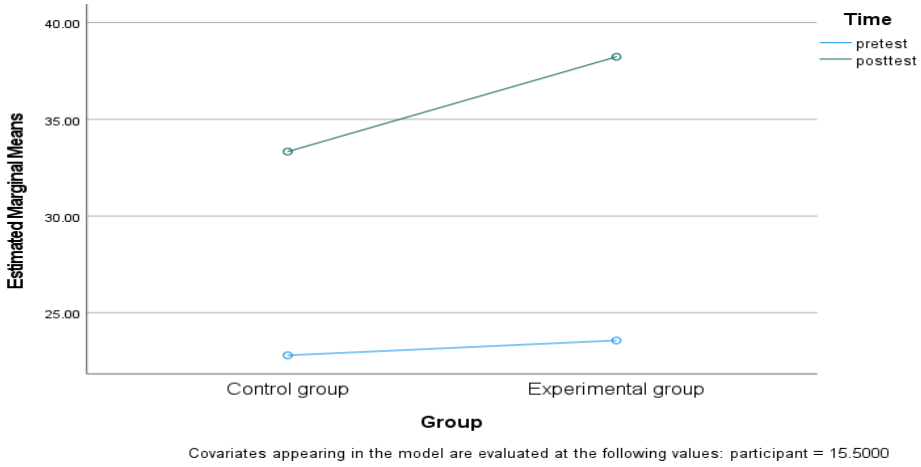


Figure 9. Means of overall score

Statistically, the main effect of the interaction between 'group' and 'time' was significant (F(1, 115) = 6.297, P=.013, ηp2 =.052, observed power= .92, R-squared= .69), with a very weak effect size. The effect of 'group' was also significant (F(1, 115) =11.835, p=001, ηp2 = .093, observed power = .96) and the effect of 'time' was significant (F(1, 115) = 234.062, p=001, ηp2 = .67, power=1).

Table 12. ANOVA results of overall score

Dependent variable: Overall score

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power[b] |
|---|---|---|---|---|---|---|---|---|
| Corrected Model | 5421.892[a] | 4 | 1355.473 | 66.613 | <.001 | .699 | 266.452 | 1.000 |
| Intercept | 29700.144 | 1 | 29700.144 | 1459.576 | <.001 | .927 | 1459.576 | 1.000 |
| participant | 290.125 | 1 | 290.125 | 14.258 | <.001 | .110 | 14.258 | .963 |
| Group | 240.833 | 1 | 240.833 | 11.835 | <.001 | .093 | 11.835 | .927 |
| Time | 4762.800 | 1 | 4762.800 | 234.062 | <.001 | .671 | 234.062 | 1.000 |
| Group * Time | 128.133 | 1 | 128.133 | 6.297 | .013 | .052 | 6.297 | .701 |
| Error | 2340.075 | 115 | 20.348 | | | | | |
| Total | 112074.000 | 120 | | | | | | |
| Corrected Total | 7761.967 | 119 | | | | | | |

a. R Squared = .699 (Adjusted R Squared = .688)
b. Computed using alpha = .05

## 5. Discussion and Conclusion

### *5.1 General Discussion*

It has been shown that the experimental group benefited from using Grammarly for all types of errors except for two subtypes of grammatical errors, namely word forms and word usage errors. The types of errors that improved most were formatting errors, spelling errors, punctuation and capitalization errors, and grammatical errors, respectively. This improvement was also confirmed by the overall score obtained by the experimental group in the posttest. The statistically significant differences in the performance of the two groups in the posttest confirm that Grammarly can generally be an effective tool that can improve the writing of even low-level students. These findings agree with those of Nova (2018) and Russell and O'Nell (2019).

Furthermore, the fact that there was no noticeable improvement among the experimental group participants with respect to using word forms and word usage indicates that Grammarly cannot help low-level learners improve their writing with respect to all types of errors. This is in partial agreement with Demsey (2017) and Hoang and Kunnan (2016), who stated that Grammarly is suitable for English native speaker writers, who have the metalanguage that enables them to understand the long explanations given by Grammarly. Lack of improvement on

these two types of grammatical errors may relate to the difficulty of these errors (as also reflected by the performance of the control group) and the difficulty of understanding the feedback given by Grammarly on such errors on the part of learners. Moreover, errors relating to usage are especially problematic due to the discrepancy between Arabic usage and English usage (Abu Guba & Abu Qub'a, 2020). This means that writing instructors need to give more attention to these problematic aspects. This is in line with Van Beuningen, De Jong, and Kuken's (2012) observation that students with limited knowledge of English language cannot understand the explanations and feedback given by Grammarly.

One might argue that the improved results might not be caused by using Grammarly; rather they could reflect the accumulated effect of teaching as the two groups completed an intensive course in English, However, the fact that the experimental group outperformed the control group in the posttest, although the two groups' performance was alike in the pretest, shows that using Grammarly played a role in improving students' writing.

To conclude, using Grammarly to improve writing skills is beneficial not only to high-level students, as established by previous research (cf. Section 2), but also to low-level students, as established in this study. However, the contribution of Grammarly to improve complex grammatical aspects such as using words and their forms properly is limited.

## 6. Limitations and Recommendations

One limitation of this study is the sample size, which is relatively small. A future study with a larger sample would lend more support to our findings. Also, this study dealt with only the free version of Grammarly; studies using the premium subscription, which checks 400 types of errors is recommended.

In light of this study results, we suggest that students should use Grammarly on a regular basis to improve their writing quality and skills. This would help learners become not only more independent but also self-confident writers. Instructors should encourage students to use Grammarly and guide them to understand and benefit from the feedback provided. Moreover, instructors should focus on the other types of errors that were difficult to rectify as AWE technology cannot totally replace the direct feedback given by teachers.

## Acknowledgments

'This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. 5561]'

## References

Abu Guba, M. N. (2023). Prosody trumps orthography in second language phonology: The case of consonant gemination. *Journal of Psycholinguistic Research*, *52*, 997-1015. https://doi.org/10.1007/s10936 023-09933-8.

Abu Guba, M. N., & A. Abu Qub'a (2020). Perceptions of clichés by Arab English bilinguals: Implications to academic writing. *Linguistics Journal, 14*(2).

Abu Guba, M. N., B. Mashaqba, S. J., & O. Al-Haj Eid. (2023). Production of vowel reduction by Jordanian-Arabic speakers of English: An acoustic study. *Poznan Studies in Contemporary Linguistics*, *59*(1), 1-25. https://doi.org/10.1515/psicl-2022-2011

Abu Guba, M. N., S. Fareh & S. Yagi. (2023). Speech rhythms in English and Arabic: A contrastive review. *International Journal of Arabic-English Studies, 23*(1), 183-202. https://doi.org/10.33806/ijaes2000.23.1.10

Beuningen, C. G., Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62***,** 1-41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

Bigert, J., Kann, V., Knutsson, O., & Sjöbergh, J. (2004). Grammar checking for Swedish second language learners. In CALL for the Nordic Languages: Tools and Methods for Computer Assisted Language Learning (pp. 33-47). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-12497

Burston, J. (2013). Review of BonPatron: An online spelling, grammar, and expression checker. *Calico Journal*, *25*(2), 337-347. https://doi.org/10.1558/cj.v25i2.337-347

Cavaleri, M., & 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 and Learning*, *10*, A223-A236.

Chen, C. F., & Cheng, W. Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology, 12*(2), 94-112. Retrieved from https://www.lltjournal.org/item/2631

Daniels, P., & Leslie, D. (2013). Grammar software ready for EFL writers. *OnCue Journal*, *9*(4), 391-401.

Dembsey, J. M. (2017). Closing the Grammarly® gaps: A study of claims and feedback from an online grammar program." *Writing Center Journal*, 36(1), 63-96, 98-100. https://doi.org/10.7771/2832-9414.1815

Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Lingua Cultura, 12*. 395. https://doi.org/10.21512/lc.v12i4.4582

Grammarly Inc,. (2021). Retrieved 30, Oct, 2021from https://www.grammarly.com/about

Hoang, G. T. L., & Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of MY Access. *Language Assessment Quarterly*, *13*(4), 359-376. https://doi.org/10.1080/15434303.2016.1230121

Jayavalan, K., & Razali, M. (2018). Effectiveness of online grammar checker to improve secondary students' English narrative essay writing. *International Research Journal of Education and Sciences, 2*(1), 1-6.

Karyuatry, L., & Rizqan, M. (2018). Grammarly as a tool to improve students' writing quality: free online-proofreader across the boundaries. JSSH (*Jurnal Sains Sosial dan Humaniora*), *2*, 83. https://doi.org/10.30595/jssh.v2i1.2297

Lailika, H. I. (2019). *Students' Peceptions of The Use of Grammarly as an Online Grammar Checker in Thesis Writing* (Doctoral dissertation, UIN Sunan Ampel Surabaya).

Moore, B. (2018). "Grammarly". PCMAG. Retrieved 21/9/2021.

Nova, M. (2018). Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience. *Premise: Journal of English Education, 7*(1), 80-97. https://doi.org/10.24127/pj.v7i1.1332

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42-56. https://doi.org/10.14742/ajet.3795

Potter, R., & Fuller, D. (2008). My new teaching partner? Using the grammar checker in writing instruction. *English Journal, 98,* 36-41.

Qassemzadeh, A., & Soleimani, H. (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,* 1884-1894. https://doi.org/10.17507/tpls.0609.23

Radi, O. (2014). *Teachers' perceptions of the individual case studies' literacy performance and their use of computer tools.* International Association for Development of the Information Society.

Schraudner, M. (2014). The online teacher's assistant: Using automated correction programs to supplement learning and lesson planning. *CELE Journal*, *22*, 128-140.

Vernon, A. (2000). Computerized grammar checkers 2000: Capabilities, limitations, and pedagogical possibilities. *Computers and Composition*, *17,* 329-349. https://doi.org/10.1016/S8755-4615(00)00038-4

Yulianti, E., & Reni. (2018). Utilizing Grammarly in teaching writing recount text through genre-based approach. *International Journal of Science, Technology and Society, 6*(1), 1-5. https://doi.org/10.11648/j.ijsts.20180601.11

Zhang, J., Ozer, H. Z., & Bayazeed, R. (2020). Grammarly vs. face-to-face tutoring at the writing center: ESL student writers' perceptions. *Praxis*, *17*(2), 33-47.