



Examining the impact of Grammarly on the quality of mobile L2 writing

 Castledown



This work is licensed under a Creative Commons Attribution 4.0 International License.

Gilbert Dizon

gilbert.dizon.jr@gmail.com
Himeji Dokkyo University, JAPAN

John M. Gayed

johnmaurice.gayed@gmail.com
Tokyo Institute of Technology, JAPAN

While the use of automated writing evaluation software has received much attention in CALL literature, as Frankenberg-Garcia (2019) notes, empirical research on predictive text and intelligent writing assistants is lacking. Thus, this study addressed this gap in the literature by examining the impact of Grammarly, an intelligent writing assistant that incorporates predictive text technology, on the mobile writing quality of Japanese L2 English students. Specifically, the study explored if Grammarly had a significant effect on the grammatical accuracy, lexical richness, writing fluency, or syntactic complexity of L2 students' writing when compared to unassisted mobile writing. A total of 31 university EFL students participated in the 8-week study which utilized a counterbalanced design. Participants took part in weekly guided freewriting tasks under each writing condition (Non-Grammarly and Grammarly) over the duration of the study. The descriptive statistics and results from t-tests showed that when students wrote with the assistance of Grammarly, they produced fewer grammatical errors and wrote with more lexical variation. These findings highlight the potential of predictive text and real-time corrective feedback as a way to support L2 writing, particularly among novice writers who may struggle to write effectively in the L2.

Keywords: intelligent computer assisted language learning (iCALL), L2 writing, automated corrective feedback

Introduction

Although native speakers and second language (L2) students generally use similar process patterns when writing, research shows that the L2 writing of adults “is more constrained, more difficult, and less effective” than that of

native writers (Silva, 1993, p. 668). For instance, it has been shown that L2 students tend to focus more on lower-level writing processes, such as grammar and vocabulary use, than higher-level ones (Weigle, 2005), thereby resulting in texts that are of lower-quality (Schoonen, Snellings, Stevenson, & van Gelderen, 2009). Therefore, it is imperative that language students be given the tools and resources necessary to alleviate the enhanced cognitive load that L2 writing places on students (Allen, Crossley, Snow, & McNamara, 2014). The use of web 2.0 technologies such as blogs, wikis, and social-networking services (SNSs) has been seen as a way to bridge this gap between L1 and L2 writing, as these tools promote a more comfortable learning environment, encourage collaboration and interaction, as well as support linguistic improvements in the target language (Wang & Vásquez, 2014). Automated writing evaluation software (AWE) has also been widely-used to support L2 students in their writing development. These systems help foreign language students write more accurately, i.e., reduce the number errors related to grammar, vocabulary, style, organization and mechanics, by giving them automated corrective feedback. Not only can AWE programs provide support for L2 students, but they also can be advantageous for language instructors by saving them time that would be otherwise spent giving feedback on lower-level writing processes, which in turn, affords teachers more time to focus on other aspects of L2 writing (Grimes & Warschauer, 2010). While the use of web 2.0 technologies and AWE programs has been well-studied in L2 research, to the best of the authors' knowledge, no study has looked at the use of predictive text to support L2 writing. Predictive text seems to be a potentially useful tool for L2 students, as the technology offers real time word-choice suggestions to writers based on the context of the words in a given sentence and the first letters typed. This allows novice L2 students to cut down on grammar and vocabulary errors as well as increase the volume of their output. Moreover, advancements in natural language processing and machine learning have led to the development of intelligent writing assistants which offer synchronous feedback on more complex aspects of writing compared to text editors (Frankenberg-Garcia, 2019). Yet, as Frankenberg-Garcia (2019) notes, empirical research on these tools (predictive text and synchronous intelligent writing assistants) is lacking. Therefore, this study addresses this gap in the literature by assessing the effects of Grammarly, an intelligent writing assistant, on four aspects of L2 writing among Japanese EFL students.

Literature review

L2 writing with web 2.0 technologies

Given that the present study examines mobile writing on an online writing platform (Google Classroom), it is important to understand previous research which has investigated the affordances of online writing and mobile-assisted L2 writing. As noted by Reinhardt (2019), when it comes to web 2.0 technologies and L2 learning, “blogs, wikis, and SNSs are the most researched because

they have been around the longest and are the most popular” (p. 2). With that said, the authors have decided to limit the review of web 2.0 technologies to blogs and SNSs as wikis primarily focus on collaborative writing among students, which is in contrast to the online writing platform used in the present study (Google Classroom). While more quantitative and qualitative research needs to be done to explore the impact of web 2.0 on L2 writing, the majority of the studies reviewed in this paper demonstrate that using said technologies can positively influence students’ L2 writing confidence, fluency and accuracy.

Blogs

Focusing on popular web 2.0 blogging tools Lee (2010) conducted a study that showed a strong correlation between blogging and improving students’ grammatical accuracy, lexical richness and creativity. The study involved university students who maintained blogs over a fourteen-week period. The researcher evaluated the writing done by the students throughout the treatment period and found the quality of her students’ writing had shown demonstrable improvements. In a follow-up study, Lee (2017) found that the ease of conducting scaffolded peer feedback helped improve students’ written content and improved their grammatical accuracy. In addition, the study concludes that a blogging platform helps students be more creative with their writing in addition to improving attention to language forms. Both Nakatsukasa (2009) and Fellner and Apple (2006) analyzed English learners’ development while using blogs as a language learning tool. In the two studies, the researchers found that students were able to improve the lexical richness and fluency of their writing and focused more on grammar when they were engaged with the blogging activities introduced in the classroom. Nepomuceno (2011) discusses how blogging can contribute to the development of students’ writing skills. The researcher found that students had a better opinion of online writing versus traditional writing tasks. The students in the study expressed that the ease of access, interactivity and unpressured peer-support as reasons they preferred the blog format.

SNSs

Studies by Dizon (2016) and Wang and Vasquez (2014) show that the use of social networks such as Facebook seems to have a positive effect on L2 writing fluency. In both studies, students who used Facebook outperformed those in control/comparison groups on a writing post-test. Shih (2011) used a blended teaching approach integrating Facebook with first-year university students. Quantitative and qualitative analysis was conducted by the researcher with pre and post-tests, survey analysis and student interviews. The findings showed students were able to improve their writing skills while significantly enhancing their motivation and interest in completing classroom assignments. A study by Yen, Hou, and Chang (2015) used qualitative content analysis to show the effects of role-playing activities via Facebook and Skype in a business college. The

results from their research indicated that students were able to improve their speaking and writing skills. The content analysis also showed improvement in speaking and writing due to the platforms' peer- and self-corrective behaviors.

Mobile-assisted L2 writing

Current research shows that mobile-assisted language learning (MALL) has a mixed effect on L2 learning and that there are specific areas that are in need of further exploration. In an analysis of fifty popular mobile language learning applications, Heil, Wu, Lee and Schmidt (2016) identified significant limitations in the use of these applications for improving second language acquisition. The researchers identified lack of relevant context, an inability to adapt to the individual learner, and limited corrective feedback as key areas that need to be improved with further advances in technology in order for the applications to be more effective for L2 acquisition. Another area of weakness in MALL literature identified by the researchers was the underutilization of writing when compared to listening and reading studies. In a recent study, Lee (2020) examined the difference between smartphone- and handwritten-based writing fluency among first- and second-year Japanese university students. The students (n = 1449) were divided into smartphone-based and paper-based groups. The researcher was able to show through analysis of text samples that the handwritten submissions were longer in length and sophistication. The researcher argues that more research is needed in mobile-assisted L2 writing to better identify the influences on students' writing fluency. Li and Hegelheimer (2013) used a mixed methods approach to examine the effects of using a web-based mobile application on an English as a second language (ESL) writing class. Quantitative data included grammar pre- and post-tests as well as self-editing activity and writing accuracy data taken from two writing assignments, while qualitative data consisted of a survey which was used to understand the learners' views of the mobile app. The researchers found positive gains for students in a post-test of grammar and also found students were able to reduce the number of written errors when writing assignments were analyzed. In addition, students were able to improve their corrective editing self-efficacy after using the mobile app that was used in the study. In a meta-analysis of 37 articles in MALL research from 2007 to 2016, Rustam, Wu-Yuin, and Yueh-Min (2017) found most research used mixed-methods research designs that showed positive trends in language proficiency gains when students used mobile technologies in the classroom. However, they did identify the need for more quantitative analysis using control groups and larger sample sizes to show a stronger correlation with the given treatments and observed outcomes. Additionally, similar to the analysis by Heil *et al.* (2016), the researchers found that writing was an understudied topic among MALL research focusing on a specific aspect of L2 proficiency, with studies centered on listening, speaking, vocabulary, and idioms outnumbering those which examined writing.



Automated feedback in L2 writing

The use of automated writing evaluation (AWE) software has received extensive attention in L2 research, particularly when it comes to the grammatical accuracy of L2 writing. For instance, Bai and Hu (2017) investigated AWE software among L2 English learners in China. They found that grammar feedback from the AWE program was relatively inaccurate and thus was ignored by some of the students. Wang, Shang, and Briody (2013) also looked at the use of AWE software with Taiwanese college EFL students and examined its effect on writing accuracy. The researchers found that not only did the experimental group who used the software significantly reduce the number of grammatical errors in their writing from the pre- to the post-test, but learners in this group also outperformed the control group when it came to grammatical accuracy post-treatment. In another study, Liao (2016) examined how AWE could improve grammatical performance among Taiwanese L2 English writers. The results of the study revealed that the software did help improve grammatical accuracy. More specifically, even though improvements were seen in each revision, significant improvements to new text were not found until the third draft of the essay. Lastly, in a comparative study between online peer feedback and AWE software, Shang (2019) found that the former had a greater positive effect on the grammatical accuracy of Taiwanese L2 English writers. This finding demonstrates that human feedback, whether from a teacher or a peer, can still prove to be useful or perhaps more beneficial than automated feedback.

Writing quality has been another area in AWE research that has been well-studied in L2 literature. Li, Link, and Hegelheimer (2015) looked at the use of AWE among ESL students at an American university and found that use of the system led to a significant decrease in the total number of errors (grammar, usage, mechanics, and style) from the rough to the final draft. In a similar study, Ranalli, Link, and Chukharev-Hudilainen (2017) focused on the accuracy of the feedback given by a popular AWE system as well as the ability of American college ESL students to make use of this feedback to correct errors. The researchers found that while there were issues related to the accuracy of the feedback given (some types of errors did not reach the 70% accuracy threshold set by the researchers), the students were able to correct their mistakes according to the feedback by the program 55–65% of the time. In a 2011 study by Cotos, the participants were able to make significant improvements in the rhetorical quality of their writing, as judged by human raters, from the first to their final drafts by making use of the feedback given by the AWE software. Cheng (2017) examined the use of automated feedback among college EFL students in Hong Kong. Results of the study indicate that the AWE group significantly outperformed the control group, thus demonstrating the positive impact that AWE software can have on overall writing quality.



Grammar checkers

Similar to the use of AWE, there is a strong body of literature examining the capability of grammar checkers (GCs) in detecting specific mistakes in L2 students' writing. For example, Chen (2009) compared the effectiveness of two GCs (Microsoft ESL Assistant and NTNU statistical grammar checker) in identifying common L2 English writing errors. The researcher found that NTNU statistical grammar checker was superior to Microsoft ESL Assistant, given that the former could detect eight out of the ten target errors, while the latter was only able to identify five. In a study detailing the development of a GC based on natural language processing, Chodorow *et al.* (2007) reported that their application could detect preposition errors produced by L2 English students at a 0.8 precision rate with a recall rate (percentage of valid errors detected) of 0.3. Gomon *et al.* (2009) also reported on a GC developed by the researchers called Microsoft Research ESL Assistant, which was specifically created for East Asian L2 English learners. According to their analysis, their GC was able to correctly recall 37% of article errors, 27% of the noun-related errors, and 18% of preposition errors. In a recent study on the use of two GCs (Grammarly and Virtual Writing Assistant), John and Woll (2020) found that neither application provided a high level of coverage, i.e., the number of errors the systems were able to correctly detect versus those overlooked. Therefore, the authors conclude that the examined CGs cannot replace teacher feedback at their current stage of development. Considering these results by John and Woll (2020) and those found in previous investigations of GCs (Chen, 2009; Chodorow *et al.*, 2007; Gomon *et al.*, 2009) it seems as though these tools are not yet able to provide consistently accurate feedback to L2 students at this stage of their development.

In summary, the use of web 2.0 and mobile technologies for L2 writing has been studied extensively in the literature. Similarly, AWE systems and GCs have received ample attention in L2 research due to the affordances they potentially offer to both language teachers and language learners. Moreover, the ubiquity of smartphones has coincided with the rise of MALL-centered research on how mobile devices can be used to enhance the language learning process. Nevertheless, there are gaps that ought to be addressed in these areas of L2 research. First, L2 writing is an area that is in need of further investigation in MALL research, as studies exploring other aspects of L2 learning have been more studied (Heil *et al.*, 2016; Rustam *et al.*, 2017). Furthermore, intelligent AWE software that offers real-time feedback and predictive text suggestions to L2 students is another area in CALL research that is understudied (Frankenberg-Garcia, 2019). Given these gaps in L2 literature, the following question was addressed in the present study:

1. Is there a significant difference in grammatical accuracy, lexical richness, syntactic complexity, or writing fluency between EFL students who used Grammarly and those who did not use the writing aid in L2 writing tasks?



Methodology

Research design

A counterbalanced research design was utilized in the 8-week study. During the first four weekly writing sessions, approximately half of the students wrote on a set of predetermined topics under the Grammarly condition while the other half wrote without the use of any aids. After these four weeks were completed, the students switched writing conditions and wrote about a second set of topics. In total, 248 pieces of writing were produced by the participants. The students' writing was then analyzed to check for significant differences in lexical sophistication, syntactic complexity, writing fluency, and grammatical accuracy between the two writing conditions (Grammarly and non-Grammarly). In order to focus on a specific set of grammatical errors, only treatable errors were analyzed. According to Ferris and Roberts (2001), treatable errors are mistakes that L2 students have a higher chance of correcting on their own in future writing. To be specific, these errors included verb errors (verb tense, subject-verb agreement, verb omissions), noun ending errors (singular/plural noun errors, possessive noun errors) and article errors (incorrect articles, article omissions, unnecessary articles). Table 1 below shows sample errors of each type from the participants' writing. In terms of lexical richness, Laufer and Nation's (1995) Lexical Frequency Profile (LFP) was used. The LFP examines the proportion of different word frequencies in the writing of L2 students, specifically, the number of words written beyond the 2000 word-frequency level divided by the total number of words. Writing fluency refers to the number of words the students were able to write within an allotted time (Wolfe-Quintero, Inagaki, & Kim, 1998). Finally, syntactic complexity refers to the amount of subordination (clause per t-unit) used by the students in their writing. As noted by Lu and Ai (2015), subordination is thought of as one of the variables that best indicates syntactic complexity among novice and intermediate writers, which also mirrors the ability levels of the participants in the present study.

Table 1. Examples of errors.

verb tense	When I was a kid I <i>chase</i> dragonflies in Kochi.
subject-verb agreement	We <i>loves</i> Korea.
verb omission	I learn many countries culture in there. And I (<i>had</i>) free time in the night.
singular/plural noun	But this university force <i>student</i> to study something we don't want to study, or don't need.
possessive noun	I can go to back <i>parents</i> home during winter vacation.
incorrect article	<i>The</i> festival is held in my hometown. It is very interesting.
article omission	I have (<i>a</i>) part-time job in LAWSON.
unnecessary article	But <i>the</i> winter vacation ends shortly.

Note: Errors are italicized. Omissions are in parentheses.



Participants

Thirty-one college EFL students participated in the study, which consisted of both male and female students as well as first- and second-year students. The participants were enrolled in a communicative language course at a private Japanese university that met three times a week over the course of the fall 2019 semester. Four classes took part in the study, with the researchers responsible for teaching two classes each. Placement in the classes was based on the participants' scores on the institution-based assessment version of the Eiken exam, a widely-used standardized English-language assessment in Japan that evaluates test-takers' reading, writing, listening, and speaking skills. According to the participants' Eiken test scores, their approximate L2 ability levels were A1–A2 on the Common European Framework of Reference for Languages (CEFR) scale. Table 2 below lists the participants' Eiken test levels along with their CEFR equivalent.

Table 2. Participants' Eiken test levels with CEFR equivalent

Class	# of students	Year	Eiken level	CEFR level
Class A	5	1st	4 & 5	A1
Class B	8	1st	3	A1
Class C	10	2nd	Pre-2	A2
Class D	8	2nd	Pre-2	A2

Grammarly

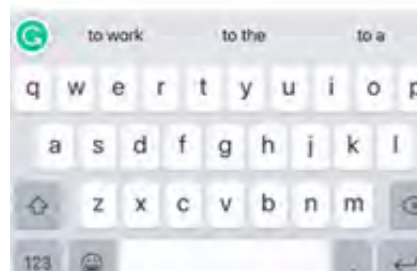
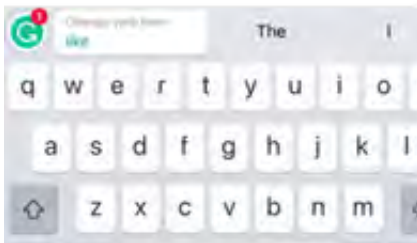
According to the Grammarly website, the intelligent writing assistant's "sophisticated AI system [is] built to analyze English sentences. Unlike other tools (most spell checkers, for instance) that rely on a rigid set of rules, Grammarly takes context into account when making corrections or suggestions" (Grammarly, 2020). The mobile Grammarly keyboard has two main features, auto correction, which flags grammar errors and offers a correct alternative (see Figure 1) and predictive text, which suggests words or phrases depending on the previous text written (see Figure 2). Lastly, while other intelligent writing assistants are available (e.g., Microsoft Editor, ProWritingAid), none of them offer the same level of versatility as Grammarly since it is available as a mobile keyboard, browser extension, web application, and Office 365 plug-in. Consequently, the writing assistant can be used freely on nearly any device or in conjunction with most applications, thereby making it a convenient tool for novice English writers.



Figure 1. Example of Grammarly mobile keyboard auto-correction feature.



Figure 2. Example of Grammarly mobile keyboard predictive text feature.



Treatment

All the participants in the study took part in guided freewriting tasks under the two conditions: Grammarly (experimental) and non-Grammarly (control). In the control condition, students were not allowed to use any aid and the default predictive text feature was turned off on their mobile devices. It is important to note that students' academic performance was not adversely affected even if they produced more errors in the control condition. Writing in both conditions was done through Google Classroom, which was the learning management system (LMS) used by the university. Due to the counterbalanced design, one first-year and second-year class wrote under the experimental condition for the first four weeks of the study while the remaining classes (one first-year and one second-year) wrote under the control condition. After four weeks, the students changed writing conditions. Writing topics were the same for all the students, and were related to themes that were pertinent to their everyday lives (see Appendix for a full list of topics). The weekly topics were assigned according to the counterbalanced design so that two classes each wrote about a specific topic under each writing condition. Put differently, there was an equal distribution between the topics amongst the two writing conditions. For these reasons,

any significant differences in L2 writing (lexical richness, syntactic complexity, grammatical accuracy, writing fluency) between the two writing conditions can be attributed to the AWE and text prediction features of Grammarly.



Data collection and analysis

Data was collected over the course of eight weeks through 15-minute guided freewriting tasks completed with the students' own smartphones. Compared to conventional freewriting in which no topic is provided and the focus is on creativity (Elbow, 1998), guided freewriting involves students writing about a specific theme (Harper, 2015). In addition, guided freewriting has been shown to increase students' writing fluency (Hwang, 2010), thus, it was chosen as the designated writing task in this study. Prior to writing, the prompt for that particular week was posted on the class's Google Classroom page. Each prompt was presented in both the students' L1 (Japanese) and the target language so that it was clear and understandable. After the 15-minute guiding freewriting task was over, the students were instructed to post their responses to Google Classroom and were assigned to reply to their classmates' writing as homework.

After data collection was complete, the students' written posts on Google Classroom were copied into an Excel spreadsheet and analyzed with respect to grammatical accuracy (human assessment), lexical richness (machine assessment), and syntactic complexity (machine assessment). The number of treatable grammatical errors made by the participants was marked by the researchers. The authors first checked 15 student responses together to build marking consistency and the remaining writing data was checked individually within a week's time. Inter-rater agreement was 90.8%. The number of grammatical errors marked by the researchers under each writing condition was averaged, and that figure was divided by the total number of words produced by each student in the Grammarly and non-Grammarly conditions. The resulting figures represented the grammatical accuracy ratio of each student under the two writing conditions. In terms of lexical richness, as assessed by LFP, and syntactic complexity, as determined by clause per t-unit, the students' posts were copied and pasted into web-based text analyzers developed by Ai and Lu (Ai & Lu, 2010; Lu, 2011, 2012; Lu & Ai, 2015) and the results obtained. Although only these two aspects were analyzed using the web-based tools, they evaluate lexical richness and syntactic sophistication under 25 and 14 different measures respectively. Prior to lexical analysis, all proper nouns and Japanese words were removed from the students' writing as their inclusion would affect the results. In total, 377 Japanese words and English proper nouns were removed. Furthermore, the students' writing was copied and pasted into an online lexical profiler (<https://www.lextutor.ca/>) to obtain a more in-depth analysis of how Grammarly influenced the participants' lexical richness.

Descriptive statistics (mean and standard deviation values) are provided to illustrate the effects that each writing condition had on the participants' writing quality. Independent t-tests were conducted to examine if there were statistically significant differences in grammatical accuracy, lexical richness,

writing fluency, or syntactic complexity between the non-Grammarly and Grammarly conditions. Effect size values are also given whenever a significant difference was found. Moreover, the number and percentage of words at six word-frequency levels (1k, 2k, 3k, 4k, 5k, and off-list) are given to provide a more nuanced look at the participants' lexical richness under the two writing conditions.

Results and discussion

Table 3 below depicts the students' writing performance under the different writing conditions (non-Grammarly and Grammarly) with respect to grammatical accuracy. Based on the descriptive statistics, students produced fewer treatable grammatical errors, i.e., mistakes related to verb, noun ending, and article usage, when they wrote with Grammarly compared to the control writing condition. This suggests that the intelligent writing tool had a positive impact on the students' ability to write with appropriate grammar in the L2. Results from an independent t-test indicate that the difference in grammatical accuracy between the two conditions was significant, $t(60) = 2.32$, $p = 0.02$. The effect size value was 0.56, which signifies an effect size between small (0.4) and medium (0.7) according to Plonsky and Oswald's (2014) effect size guidelines for L2 research. These findings support previous research which demonstrate that AWE software can positively influence the grammatical accuracy of L2 students' writing (Liao, 2016; Wang *et al.*, 2013). Although the effect size indicates that Grammarly only had a small to medium effect on the L2 students' grammatical accuracy, the results are promising and suggest that it may indeed help students produce higher levels of grammatical accuracy through the use of the tool.

Table 3. Grammatical accuracy results

	M	SD	Total number of grammatical errors
Non-Grammarly	0.06	0.04	453.5*
Grammarly	0.04	0.03	310

*Note: 0.5 of an error was possible due to discrepancies in grammar marking between the authors

As shown in Table 4, Grammarly also seems to have had a positive effect on lexical richness, as student writing under this condition used a greater variety of different words (0.18 with Grammarly and 0.15 without). Results from an independent t-test confirmed that this difference in lexical richness was in fact significant, $t(60) = 3.83$, $p = 0.0003$. The effect size value was 0.84, which falls between a medium (0.7) and large (1.0) effect size (Plonsky & Oswald, 2014). Taking a deeper look at the data by examining word usage at different frequency levels also shows an increase in lexical richness. Specifically, the students wrote more higher frequency words or words at the 1K or 2K

word-frequency levels under the non-Grammarly condition (7243 or 96.4%) compared to the Grammarly condition (7135 or 95.4%). Accordingly, the students produced more 3K, 4K, 5K, and off-list vocabulary under the Grammarly condition (347 or 4.6%) compared to the non-Grammarly condition (267 or 3.6%). Based on the descriptive statistics and statistical tests, it is clear that Grammarly had a considerable impact on the students' capacity to write with greater lexical variation. While many AWE studies have examined the impact of these tools on grammatical accuracy and overall writing quality (e.g., Li *et al.*, 2015; Wang *et al.*, 2013), the results found in this study demonstrate that Grammarly may also promote greater lexical variation in L2 students' writing due to the predictive text feature.

Table 4. Lexical richness results

	M	SD	1K		2K		3K		4K		5K		Off-list	
			#	%	#	%	#	%	#	%	#	%	#	%
Non-Grammarly	0.15	0.03	6912	91.5	331	4.4	97	1.3	54	0.7	35	0.5	81	1.07
Grammarly	0.18	0.04	6800	90.1	335	4.4	78	1.0	65	0.9	84	1.1	120	1.59

= Total number of words per word frequency level

Unlike grammatical accuracy and lexical richness however, there does not appear to be a marked difference in writing fluency between the two writing conditions according to the descriptive statistics. As shown in Table 5, both the mean number of words written and the total number of words written were quite similar with or without the use of Grammarly. An independent t-test shows that the difference was not statistically significant, $t(60) = 0.07$, $p = 0.94$. While this is somewhat surprising given that predictive text can seemingly make English writing easier for L2 students, the research design of the present study may have been a factor in the students' inability to write more fluently with Grammarly. Specifically, they wrote under the Grammarly condition only four times in a one-month time period. Therefore, they may not have been accustomed to using Grammarly within this short time span. Perhaps if there were more writing tasks and/or an extended intervention under the experimental condition, then the students may have been able to produce more writing with Grammarly; however, that is for a future study to explore.

In terms of syntactic complexity, the participants actually performed better when writing without the help of Grammarly. That is, students' writing under the control condition exhibited more subordination compared to when they wrote using the Grammarly. With that said, results from an independent t-test indicate that the difference in syntactic complexity between the two writing conditions was not significant, $t(60) = 0.81$, $p = 0.41$. In other words, the use of Grammarly did not help nor hinder the students' ability to write with greater syntactic sophistication. This finding suggests that the use of the digital tool may not encourage L2 students to use more subordination in their writing.

Table 5. Writing fluency and syntactic complexity results

	Writing fluency			Syntactic complexity		
	M	SD	Total number of words written	M	SD	Total number of subordinate clauses
Non-Grammarly	243.68	97.69	7554	1.31	0.45	1091
Grammarly	245.48	97.75	7610	1.23	0.24	1134

In summary, the L2 students performed significantly better in two out of the four variables examined in this study. While a significant difference was not found in regards to writing fluency or syntactic complexity, the L2 students wrote with higher levels of grammatical accuracy and lexical richness when using the Grammarly keyboard. There are some plausible explanations for these results. First, the automated corrective feedback feature likely helped students reduce the number of treatable grammatical errors, including common mistakes related to verb tense and singular vs. plural noun errors, as evidenced by the fact that when students wrote with Grammarly, they produced significantly more accurate texts. Additionally, it is possible that students made good use of the predictive text feature of Grammarly to incorporate less frequently used vocabulary in their writing, which in turn, also significantly increased the level of lexical richness. On the other hand, while these features promoted grammatical accuracy and lexical richness, they were not as helpful in encouraging more subordination or enhanced writing fluency. Predictive text and corrective feedback may be useful at the word-level, but the results from the present study suggest that they may not help L2 students produce more complex sentences nor increase writing output. As aforementioned, the short intervention might have played a factor in the non-significant results concerning these variables, so future research could explore the long-term use of AI-based writing assistants to determine if more experience with these tools leads to greater improvements in L2 writing among students.

Conclusion

Real-time corrective feedback and predictive text are two technologies that have received very little attention in L2 research (Frankenberg-Garcia, 2019). As a result, this study sought to address this gap in the literature by examining the use of an intelligent writing assistant, namely Grammarly, to improve different aspects of L2 writing quality. The results of the study demonstrate that the intelligent writing assistant had a significant, positive effect on the grammatical accuracy and lexical richness of L2 students compared to a control condition which did not allow for the use of any writing aids, digital or otherwise. As noted by several researchers (Schoonen *et al.*, 2009; Silva, 1993; Weigle, 2005), L2 students often struggle to produce quality writing in the target language. Thus, these findings are significant as they show that the combination of synchronous corrective feedback and predictive text can ease the cognitive

burden of L2 students and help them write more accurately and with greater lexical variety.

Some pedagogical recommendations can be made based on the study's findings. To begin, the authors recommend the use of Grammarly among beginner L2 students, especially those who struggle with grammatical accuracy. As shown in the results of this study, the use of Grammarly can help L2 students produce more grammatically accurate texts, which may also lead to more effective communication in the L2. In addition, Grammarly could be recommended to L2 students as a means to increase lexical variation in their writing, as this study's findings illustrate that lexical richness was enhanced in the Grammarly condition. As noted by Laufer and Nation (1995), a lack of lexical variation is one of the major weaknesses in L2 students' written production, so Grammarly could be used to alleviate this common issue among novice L2 students.

Although the results from this study highlight the potential of intelligent writing assistants to support L2 writing, as with any research utilizing convenience sampling, several limitations must be addressed. First and foremost, the small sample size makes it difficult to make broad generalizations from the findings. Therefore, future research should incorporate larger sample sizes from more diverse student populations. In addition, only four aspects of L2 writing were examined in the study. Because of this, researchers could utilize a holistic scale to evaluate whether or not intelligent writing assistants such as Grammarly can improve the overall quality of L2 writing in a future study. Moreover, a post-test was not administered in the study and as a result, it is not known if L2 students can become more effective English writers through the use of intelligent writing assistants. Hence, a future study could utilize a pre-posttest design to evaluate the effectiveness of Grammarly or another AI-based writing assistant to see if L2 students can make long-term improvements to their writing. Lastly, student perceptions of Grammarly were not explored; thus, it would be interesting to utilize surveys, reflective reports, and/or interviews to better understand views towards these kinds of tools to aid writing among L2 students.

References

- Ai, H., & Lu, X. (2010). A web-based system for automatic measurement of lexical complexity. *Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*, Amherst, MA. June 8–12.
- Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology, 18*(2), 124–150. <http://dx.doi.org/10125/44373>
- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology, 37*(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>



- Blackmore-Squires, S. (2010). An investigation into the use of blog as a tool to improve writing in the second language classroom. MA thesis submitted to the University of Manchester. Retrieved from <http://www.asian-efl-journal.com/Thesis/Thesis-Squires.pdf>
- Chen, H.-J. H. (2009). Evaluating two web-based grammar checkers – Microsoft ESL Assistant and NTNU statistical grammar checker. *Computational Linguistics and Chinese Language Processing*, 14(2), 161–180.
- Cheng, G. (2017). The impact of online automated feedback on students’ reflective journal writing in an EFL course. *The Internet and Higher Education*, 34, 18–27. <https://doi.org/10.1016/j.iheduc.2017.04.002>
- Chodorow, M., Tetreault, J. R., & Han, N.-R. (2007). Detection of grammatical errors involving prepositions. In F. Costello, J. Kelleher & M. Volk (Eds.), *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions* (pp. 25–30). <https://doi.org/10.3115/1654629.1654635>
- Clement, J. (2019). *Percentage of U.S. population who currently use any social media from 2008 to 2019*. Statista. Retrieved from <https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile>
- Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal*, 28(2), 420–459. <https://doi.org/10.11139/cj.28.2.420-459>
- Dizon, G. (2016). A comparative study of Facebook vs. paper-and-pencil writing to improve L2 writing skills. *Computer Assisted Language Learning*, 29(8), 1249–1258. <https://doi.org/10.1080/09588221.2016.1266369>
- Ebrary, K. E., & Windeatt, S. (2019). Eye tracking analysis of EAP students’ regions of interest in computer-based feedback on grammar, usage, mechanics, style and organization and development. *System*, 83, 36–49. <https://doi.org/10.1016/j.system.2019.03.007>
- Elbow, P. (1998). *Writing without teachers*. Oxford University Press.
- Fellner, T., & Apple, M. (2006). Developing writing fluency and lexical complexity with blogs. *The JALT CALL Journal*, 2(1), 15–26.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184. [https://doi.org/10.1016/S1060-3743\(01\)00039-X](https://doi.org/10.1016/S1060-3743(01)00039-X)
- Frankenberg-Garcia, A. (2019). Combining user needs, lexicographic data and digital writing Environments. *Language Teaching*, 53(1), 29–43. <https://doi.org/10.1017/S0261444818000277>
- Gamon, M., Leacock, C., Brockett, C., Dolan, W. B., Gao, J., Belenko, D., & Klementiev, A. (2009). Using statistical techniques and web search to correct ESL errors. *CALICO Journal*, 26(3), 491–511. <https://doi.org/10.1558/cj.v26i3.491-511>
- Grammarly. (2020). FAQ | Grammar check. Retrieved from <https://www.grammarly.com/grammar-check>

- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment*, 8(6), 4–43. Retrieved from <http://files.eric.ed.gov/fulltext/EJ882522.pdf>
- Harper, J. K. (2015). Combining movie viewing and guided freewriting to enhance student attitudes in a university academic writing class. *The Language Teacher*, 39(4), 3–7. <https://doi.org/10.37546/JALTTLT39.4-1>
- Heil, C., Wu, J., Lee, J., & Schmidt, T. (2016). A review of mobile language learning applications: Trends, challenges, and opportunities. *The EuroCALL Review*, 24(2), 32–50. <https://doi.org/10.4995/eurocall.2016.6402>
- Hwang, J.A. (2010). A case study of the influence of freewriting on writing fluency and confidence of EFL college-level students. *Second Language Studies*, 28(2), 97–134. Retrieved from <https://www.hawaii.edu/sls/wp-content/uploads/2014/09/Hwang.pdf>
- John, P., & Woll, Nina. (2020). Using grammar checkers in an ESL context: An investigation of automatic corrective feedback. *CALICO Journal*, 37(2), 169–172. <https://doi.org/10.1558/cj.36523>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Lee, B. (2020). Smartphone tapping vs. handwriting: A comparison of writing medium. *The EuroCALL Review*, 28(1), 15–25. <https://doi.org/10.4995/eurocall.2020.12036>
- Lee, L. (2017). Learners' perceptions of the effectiveness of blogging for L2 writing in fully online language courses. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 7(1), 19–33. <https://doi.org/10.4018/IJCALLT.2017010102>
- Lee, L. (2010). Fostering reflective writing and interactive exchange through blogging in an advanced language course. *ReCALL*, 22(2), 212–227. <https://doi.org/10.1017/S095834401000008X>
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology*, 17(3), 135–156. <http://dx.doi.org/10125/44343>
- Liao, H.-C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92. <https://doi.org/10.1016/j.system.2016.02.007>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writer's language development. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x



- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27.
<https://doi.org/10.1016/j.jslw.2015.06.003>
- Nakatsukasa, K. (2009). The efficacy and students' perceptions of collaborative blogging in an ESL classroom. *Developing and evaluating language learning materials* (pp. 69–84). Iowa State University.
- Nepomuceno, M. (2011) Writing online: Using blogs as an alternative writing activity in tertiary ESL classes. *TESOL Journal*, 5, 92–105.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 Research. *Language Learning*, 64(4), 878–912.
<https://doi.org/10.1111/lang.12079>
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 1, 8–25.
<https://doi.org/10.1080/01443410.2015.1136407>
- Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching*, 52(1), 1–39. <https://doi.org/10.1017/S0261444818000356>
- Rustam S., Wu-Yuin H. & Yueh-Min H. (2017). Review of research on mobile language learning in authentic environments. *Computer Assisted Language Learning*, 30(3–4), 284–303.
<https://doi.org/10.1080/09588221.2017.1308383>
- Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching and research* (pp. 77–101). deGruyter.
- Shang, H.-F. (2019). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, 1–13. <https://doi.org/10.1080/10494820.2019.1629601>
- Shih, R. C. (2011) Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, 27(5), 829–845. Retrieved from <https://eric.ed.gov/?id=EJ943540>
- Silva. T. (1993). Toward and understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657–677. <https://doi.org/10.2307/3587400>
- Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, 31, 275–288. <http://doi.org/10.1017/S0267190511000079>
- Wang, Y.-J., Shang, H.-K., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257. <https://doi.org/10.1080/09588221.2012.655300>

- Wang, S., & Vásquez, C. (2012). Web 2.0 and second language learning: What does the research tell us? *CALICO Journal*, 29(3), 412–430.
<https://doi.org/10.11139/cj.29.3.412-430>
- Wang, S., & Vásquez, C. (2014). The effect of target language use in social media on intermediate-level Chinese language learners' writing performance. *CALICO Journal*, 31(1), p. 78–102.
<http://dx.doi.org/10.11139/cj.28.2.345-368>
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Technical Report No. 17). National Foreign Language Resource Center.
- Yen, Y.-C., Hou, H.-T., & Chang, K. E. (2015) Applying role-playing strategy to enhance learners' writing and speaking skills in EFL courses using Facebook and Skype as learning tools: A case study in Taiwan. *Computer Assisted Language Learning*, 28(5), 383–406.
<https://doi.org/10.1080/09588221.2013.839568>

Appendix A

Guided freewriting topics and writing timeline

	Class A & C	Class B & D
Week 1	health (Grammarly)	university life (control)
Week 2	winter vacation (Grammarly)	work (control)
Week 3	hometown (Grammarly)	holidays (control)
Week 4	travel (Grammarly)	hobbies (control)
Week 5	university life (control)	health (Grammarly)
Week 6	work (control)	winter vacation (Grammarly)
Week 7	holidays (control)	hometown (Grammarly)
Week 8	hobbies (control)	travel (Grammarly)

Appendix B

Writing samples from a student in each writing condition

Student 1, Sample A (Non-Grammarly) – Topic: Hobbies

My hobby is playing (the) guitar. Because I like listening to music. When I *am* (an) elementary school student, I *listening* to rock music. So I like to play (the) guitar. I *join* a band club. In the club, I play *a* guitar and sing *a song*. But I can't sing *a song* well. I want to sing very well. But this is too difficult for me. Holiday I practice *a* guitar (a) long time. It's a chance to practice *a* guitar. But I don't do *a* voice training. Because my family *get* angry to me about my loud voice. So my voice training room is a Karaoke.

Writing sample details according to variables studied

Grammatical errors marked: 16

Number of words written beyond 2K word-frequency level: 3

Words written: 105

Number of clauses: 13



The
JALT CALL
Journal
vol. 17 no.2

Student 1, Sample B (Grammarly) – Topic: Travel

I went to Okinawa for a high school trip at the end of summer. This trip was so amazing. First, we went to Tyura aquarium. That place is very huge. So we can see a lot of time there. And there are so many kinds of fish and sea creatures. I couldn't all spot. Because we *don't* have too much free time. Next, we went to *the* camp place. In there we had a bbq. And then we had some events. For example, (there was) a bingo game, a birthday party, and a Karaoke tournament. Those events were

Writing sample details according to variables studied

Grammatical errors: 3

Number of words written beyond 2K word-frequency level: 6

Words written: 96

Number of clauses: 14

Note: Errors are marked in italics. Omissions are included in parentheses.