

## 研究ノート

### What Do EFL Students Experience While Editing?

Shoko OTAKE

#### Abstract

When writing in English, students of English as a Foreign Language (EFL) face a difficult task. They tend to make lower-level errors like simple grammar, leaving little time to focus on higher level errors like coherence. Thus, automated writing evaluation (AWE) tools like *Grammarly* have received considerable attention from language classrooms. While studies have reported the effectiveness of this tool, few have documented students' experience during editing. In this study, EFL students edited their first draft of one assignment, using both Grammarly and self-editing. They recorded how the tasks went, such as overall time taken and any errors they could not understand. They also reflected on the overall usefulness of each method, and how often they use each one. Students tended to spend more time on self-editing, but found more errors with Grammarly. First year students especially commented that they could not understand the feedback messages. Implications of these results are discussed with respect to future class instructional methods.

**Key words:** EFL, automated writing evaluation, self-editing, Grammarly, feedback

For students learning English as a Foreign Language (EFL) in Japan, writing in English can be a difficult journey. They tend to put a lot of time and effort in lower-level aspects of their writing, like grammar and vocabulary. This leaves little room for both EFL students and teachers to focus on higher-order aspects of writing, like meaning and overall coherence (Liao, 2015). Thus, the use of automated writing evaluation (AWE) technology has been of much interest for instructors in language classrooms. By using AWE in the classroom, the instructor and his/her students may be able to decrease the time spent on lower-level aspects of writing, and spend more time making effective arguments (Wang & Goodman, 2012). There are many different types of AWE; some can be used for a fee, such as *Criterion* from Educational Testing Service. Others, such as *Grammarly* from Grammarly, Inc. can be free for some basic editing (i.e., grammar, spelling, misused words), but have “Premium” features that require a fee (i.e., clarity, tone, word choice). Research on AWE has had mixed results, where some report that it does not reliably improve student writing by itself (Heffernan & Otsoshi, 2005; Chen & Cheng, 2008), whereas others report increased grammatical accuracy (Wang, Shang & Briody, 2013). In this study, we focused on EFL students' experience

with the free, website version of *Grammarly* (<https://app.grammarly.com/>), because it is a tool that is readily available to use, and does not require any downloading.

What are the strengths of *Grammarly* as a feedback tool? This question can be addressed in detail if we discuss the literature on the types of feedback effective for student achievement in general. For example, Marzano, Pickering & Pollock (2001) have pointed out that “timely” feedback is one of the keys to effective learning. That is, feedback that appears immediately after the task will lead to improved learning compared to feedback one week after the task. *Grammarly* can do just this, since feedback messages appear when students copy and paste their work into their accounts. This is much faster than other methods like peer feedback, teacher feedback, and self-editing. Past research has shown that Japanese EFL students can use *Grammarly* to decrease grammar errors and increase lexical variation (Dizon & Gayed, 2021), and Chinese EFL students benefit from immediate feedback (Li, Zhu & Ellis, 2016). Having computer-assisted editing programs such as this one can encourage students’ independency and critical thinking as a writer. On the other hand, this feedback may not be very easy for students who have difficulty understanding it (Chan & Cheng, 2006), or tend to react negatively towards technology because of previous negative experiences (Ware, 2004).

As with any kind of AWE technology, Grammarly does have its weaknesses. Ellis (2009) and Marzano et al. (2001) suggest that another indicator of effective feedback is if it gives “corrective” feedback. That is, the feedback should explicitly show why a certain response was wrong. For example, just telling a student if they were right/wrong does not improve their accuracy in a future exam. In addition, when the explanation of the error is given to students, it needs to be “tangible and transparent” (Wiggins, 2012). For example, if a student does not understand what the word ‘vague’ is, using the word in their feedback is not effective.

It is sometimes difficult to provide corrective feedback via *Grammarly* because the types of feedback it gives are not specially designed for EFL students. For example, one type of feedback message, “*Beautiful* seems to be the wrong part of speech for this context: Adjectives can only modify nouns and pronouns,” may not be a very familiar term for Japanese EFL students that have only gotten feedback in Japanese terms in the past. If students do not explicitly understand the feedback that *Grammarly* provides, it is not “corrective”. That is, by Ellis (2009) and Marzano et al. (2001)’s standards, this is not very effective feedback, and may not lead to improved grammatical skills in the long run. Therefore, it is important to investigate not only if students manage to fix grammar, but also if they understand the revisions during the editing process. Furthermore, *Grammarly* also gives types of feedback that are only available in the Premium version (we will call this “Premium feedback” in this paper). In the free version, this feedback is partially hidden, only giving students a message like “We found 1 additional issue in this text

available only for Premium users: 1 Word choice”. Needless to say, it is difficult for students to successfully fix errors just with this message.

We have discussed the strengths and weaknesses of *Grammarly* as a type of technology that gives feedback. In the past, researchers like Wiggins (1993) and Polio, Fleck & Leder (1996) have discussed how self-editing is also an effective way to give feedback. That is, looking back at their own work can also help students improve on their errors they often make while writing. Ferris (1995) had students find and record frequent error types (i.e., errors in nouns, verbs, punctuation/sentence structure, word forms, and prepositions) in their drafts. This significantly decreased ESL students’ errors over one semester. Ferris (1995) emphasized having students focus on global errors (i.e., errors that make the text incomprehensible). However, there have been few studies which have compared student experiences while self-editing and using *Grammarly* on the same text, especially because AWE is used as a supplemental tool in addition to peer/self-editing (Liao, 2015).

In the present study, we had students do two activities as part of an end-of-the-year reflection activity. First, they chose one “first” draft, and used both *Grammarly* and self-editing to edit it. Then, they wrote down their experiences while doing so, such as time taken, mistakes found, and feedback messages they did not understand. At the end, they reported their overall experience with editing and the method they would like to keep on using.

The student participants in this study were first and second-year EFL university students in Japan, who were writing a script in preparation for oral presentations about social issues. The class emphasized that students had to pretend they were speaking to an audience of native English speakers, because they are all required to study abroad in their third year. Since this class focused on oral activities and presentation structure, there was no time for teachers to comment on lower-level errors. In this particular semester, it was difficult to look at what students were doing to prepare, since classes were done online, via Zoom (from Zoom Video Communications, Inc.), due to COVID-19. Therefore, we introduced *Grammarly* to first and second-year students at the beginning of the semester as a tool to help them with assignments. It must be noted that students were encouraged to revise and submit their drafts (using both *Grammarly* and self-editing) as many times as they wished before their presentation. The peer- and self-editing was done in conjunction with conversation activities that students did over Zoom every week.

In order to assess students’ experiences with AWE and self-editing, we asked three research questions. The first question was, “How does the use of *Grammarly* affect the time Japanese EFL students spend editing, and the total number of errors they find, as compared to self-editing?”. We have already discussed how the “timely” fashion of *Grammarly* feedback could be beneficial to students. However, not many studies have examined the experience from a student’s point of view

as they edit. For example, it could be the case that many students are so unaccustomed to technology, or have difficulty in understanding the feedback given, that it takes more time to edit using *Grammarly*.

The second question we asked was, “What type of feedback message do students fail to understand? Do they tend to be errors that are indicated by the Premium version of *Grammarly*?”. Here, we went back to the fact that feedback has to be “corrective” in nature for it to work. If students do not understand what a *Grammarly* suggestion means, regardless of whether the suggestion was Premium/non-Premium, then it means *Grammarly* by itself is not “corrective” from their point of view. Alternatively, if students only report that they do not understand the Premium feedback, giving them Premium access in later semesters could help improve their learning process. As we described previously, for non-Premium users, the Premium feedback does not provide a lengthy explanation of the error.

The third question we asked was, “How often do students use *Grammarly*, as compared to self-editing, throughout one semester, and which method do they prefer?”. We included this because with an online class, we could not see what students were doing in preparation for their presentations. If students rarely used both methods, for example, it is problematic because it suggests students do not think editing is important in general. Alternatively, if students tend to use more self-editing more frequently, it could mean that they put emphasis on fixing global errors that are pointed out during group practice sessions, or that they are not used to *Grammarly* to the point where they can use it on a regular basis.

## **Method**

### **Participants**

We tested university students who were either first- ( $N = 115$ ) or second- ( $N = 106$ ) year university students studying English as a Foreign Language in the Kansai area. In terms of English ability, first year students had a L & R TOEIC score of  $M = 395.29$ ,  $SD = 99.24$ . Second year students' scores were  $M = 459.75$ ,  $SD = 119.93$ . In their English classes, they went through a series of 6 units about social issues, and presented on them. For these presentations, students were required to write and edit their own drafts, and were encouraged to edit them using both *Grammarly* and by themselves. Students completed the editing tasks and survey questions as part of an end-of-the-year review assignment.

### **Materials and Procedure**

We used Google Forms (<http://forms.google.com>) to present the task to students, which is the same way students had been doing their regular homework assignments. Students completed the

work individually at home on their own computers. They were asked to look at an instruction video if they were confused about any task.

During the task, students picked one “first draft” they wrote for a class presentation, from any unit they chose. Using the draft, they were asked to (1) Use *Grammarly* to check how many errors they could find, (2) record how much time this task took, and (3) record any error messages, if any, that they could not understand. Then, using the same draft, students needed to (1) self-edit the draft, (2) count how many errors there were, and (3) record how much time the task took. The order of the *Grammarly* and self-editing tasks were randomized across participants, by splitting each class in half and counterbalancing order of tasks between the halves. At the end, students were asked to rate, on a scale from 1 (I never use it) to 5 (I use it almost every day), how often they use *Grammarly* and self-edit their work without technology. Then, they wrote freely about what method of editing they most preferred.

## Results

Table 1 shows the average number of errors found in each task (*Grammarly* vs. self-editing). We conducted multilevel analyses with participant ID as the random factor, and task and university year (First or Second) as the fixed factors. These analyses were conducted using the *lmerTest* package in *R* (Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B., 2015). We found a significant main effect of task ( $b = -2.06$ ,  $t(220) = -6.38$ ,  $p < .001$ ), such that participants tend to find more errors with *Grammarly* than with self-editing, and another significant main effect of year ( $b = -.85$ ,  $t(219) = -2.07$ ,  $p < .05$ ), such that first year students found more errors than 2<sup>nd</sup> year students. Table 1 also shows the average time taken to finish each task. Here, we found a main effect of task ( $b = 2.31$ ,  $t(220) = 5.93$ ,  $p < .001$ ), where students needed more time to self-edit than to use *Grammarly*. This was true regardless of which year they were in, as there was no main effect of year.

Table 1  
*Average Number of Errors and Time taken (Minutes) to Finish Each Task (Standard deviations in Parentheses)*

Year	Errors		Time taken	
	Grammarly	Self-Editing	Grammarly	Self-Editing
First	6.09(4.81)	3.51(3.73)	4.97(.93)	7.08(1.04)
Second	4.70(4.06)	3.20(2.28)	4.33(.95)	6.84(1.00)

Table 2

*Average Frequency of Use and Proportion of Students that Chose Each Editing Method (Standard deviations and standard errors in parentheses, respectively)*

Year	Frequency of use		Final choice	
	Grammarly	Self-Editing	Grammarly	Self-Editing
First	3.56(1.33)	4.03(0.78)	.85(.04)	.15(.04)
Second	3.78(1.06)	3.75(0.90)	.93(.03)	.07(.03)

Second, we looked at the type of *Grammarly* feedback that students failed to understand. We first divided the types into (1) “non-Premium”, or feedback that provide the actual way to fix the error (i.e., supported by the free version of *Grammarly*), and (2) “Premium”, or feedback that does not give specific ways to fix the error for non-premium users. Several students were excluded from this part of the analysis (leaving  $N = 73$  for first years,  $N = 67$  for 2<sup>nd</sup> years) because they did not answer the question, or reported there were no feedback they did not understand.

Overall, the second-year students reported they had difficulty understanding more Premium feedback ( $M = .69$ ,  $SE$  (Standard Error) = .05) than non-Premium ones ( $M = .31$ ,  $SE = .05$ ). Using a binomial test, we found 2<sup>nd</sup> year students chose “Premium” feedback above the level for chance (.50),  $p < .05$ . However, there was no significant difference in first year students, because they reported similar numbers of Premium ( $M = .53$ ,  $SE = .06$ ) and non-Premium feedback messages ( $M = .47$ ,  $SE = .06$ ).

At the very end, students were asked what method of editing students preferred. The proportion of responses are reported in Table 2. Using a binomial test, we found both years chose “*Grammarly*” at above the level of chance (.50),  $p < .001$ . We also examined the frequency at which students used each editing method during the semester, which is also shown in Table 2. We conducted multilevel analyses with participant ID as the random factor, and task as the fixed factors. Here, year did not significantly improve the model. We found a main effect of task ( $b = .23$ ,  $t(220) = 2.39$ ,  $p < .05$ ), suggesting that in general, students use self-editing frequently.

### Discussion

Our study was designed to explore three different research questions, all of which address EFL students’ experiences with self-editing and *Grammarly*. Our first research question was, “How does the use of *Grammarly* affect the time Japanese EFL students spend editing, and the total amount of errors they find, as compared to self-editing?”. Students did tend to spend more time self-editing than editing using *Grammarly*. This is our expected result, given that research (Dizon & Gayed, 2021, Li, Zhu & Ellis, 2016; Marzano, Pickering & Pollock, 2001) has reported that the timeliness of feedback is a major advantage of *Grammarly*. The total amount of errors was greater in first

years, and for feedback given using *Grammarly*. We believe first years may have found more errors (especially grammar errors) than second years simply because of differences in English proficiency and writing experience. Students may have found fewer errors in self-editing because they were counting sentence-level or paragraph-level global errors, or failed to find errors in general. In contrast, *Grammarly* tends to detect many errors at the word- or phrase- level, leading to a larger total error count.

The second question we asked was, “What type of feedback do students fail to understand? Do they tend to be those that are indicated by the Premium version of *Grammarly*?”. We found that many students do find trouble with the Premium feedback, which is expected given that those types of feedback do not provide an explicit explanation of the error at all. Interestingly, some students wrote that they were able to correct themselves in class Zoom sessions, because their peers suggested an alternative for their word choice. This may suggest that when students receive Premium feedback, class discussion can help them find a solution. One issue especially with first year students was that they expressed difficulty in understanding non-Premium feedback as frequently as Premium feedback. This is problematic because it suggests even simple explanations about grammar are not explicit in students’ eyes. In their comments, some students expressed frustration at the fact that they used English phrases that they learned in the Japanese school system, yet the *Grammarly* system told them they were incorrect. They were reluctant to make the suggested correction, and opted to change the phrase itself. This is similar to theories of L2 learning in general; Krashen (1982) suggested that error corrections cause students to be on the defensive, and switch to a simpler language construct. These results suggest that teachers should be more explicit about the nature of *Grammarly* feedback. That is, they do not always mean a students’ writing is *wrong*, but indicates how their ideas can be expressed more clearly in a native speaker’s point of view.

Our final research question was, “How often do students use *Grammarly*, as compared to self-editing, throughout one semester, and which method do they prefer?”. We found that students tended to self-edit more frequently than using *Grammarly*. Students commented that they self-edited more because they wanted to change the structure, or fix some word-choice issues when explaining about Japanese culture. This suggests that students had developed a way to think about how clear their ideas were, and realized that their self-editing is more helpful than using *Grammarly* in this aspect. It was surprising to see how many students opted to use *Grammarly* in future editing situations, when they did not use it as frequently as self-editing. It could be that some students liked the fast nature of the process and also want to ‘get better’ at using this tool. Alternatively, students could have thought that choosing *Grammarly* was the result their teacher was looking for.

The limitations of our study are manifold. Many of the measures we analyzed here (e.g., frequency of use, understanding of feedback messages, time) are based on self-reports. Students may have answered some questions based on what they believe is the socially desirable result (i.e., students should be editing often and be using *Grammarly*). Another issue is that the survey was given at the very end of the semester, which meant students' perceptions of editing across the semester may have been inaccurate. These two problems could be combatted if students are required to keep weekly logs of their editing behavior, such as frequency and the nature of errors they find. Researchers like Ferris (1995) and Marzano et al. (2001) have proposed already in the past how this log-keeping process can aid students in their learning.

Overall, the results of this study gave us a good picture of student experiences with editing. Despite the fact that classes were online, students still edited their work frequently. Student reports of errors they did not understand showed that just looking at how many errors decrease in a paper may not necessarily gauge their grasp of language concepts.

In future classroom lessons, it could be more helpful to have students focus on one type of error message from *Grammarly*, instead of editing each and every error. They could figure out what to focus on by keeping a log of the messages that they get most frequently for each assignment. Later on, they could discuss how this could be fixed, with peers that see the same feedback message. Ferris (1995) has indeed shown that having students focus on one concept at a time during editing can facilitate their acquisition of the language, and that this leads to less negative reactions than pointing out every error. Also, based on the fact that *Grammarly* feedback is quick and detects many local errors, it may be utilized effectively in writing classes with many students. This is because teachers have difficulty in providing detailed feedback in a timely manner in this situation. With English classes that emphasize student-to-student communication and clear presentation of ideas, global errors (i.e., structure, general wordiness) are very important to fix. *Grammarly* not may be as helpful, because detailed advice for these errors is often part of the Premium version. Instead, lengthy discussions with peers and teachers could be a better option to help students elaborate on their ideas.

### References

- Chen, C. F. E., & Cheng, W. Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12, 94-112.
- Dizon, G & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *JALT CALL Journal*, 17, 74-92.
- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 3-18.
- Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from



- here? (and what do we do in the meantime...?). *Journal of Second Language Writing*, 13, 49-62.
- Ferris, D. (1995). Teaching students to self-edit. *TESOL Journal*, 4, 18-22.
- Heffernan, N., & Otsoshi, J. (2015). Comparing the pedagogical benefits of both Criterion and teacher feedback on Japanese EFL students' writing. *JALT CALL Journal*, 11, 63-76.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, 2, 734.
- Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *The Modern Language Journal*, 100, 276-295.
- Liao, H. C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70, 308-319.
- Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43-68.
- Ware, P. D. (2004). Confidence and competition online: ESL student perspectives on web-based discussions in the classroom. *Computers and Composition*, 21, 451-468.
- Wang, M. J., & Goodman, D. (2012). Automated Writing Evaluation: Students' Perceptions and Emotional Involvement. *English Teaching & Learning*, 36.
- Wang, Y. J., Shang, H. F., & 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, 234-257.
- Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. (2012). Seven keys to effective feedback. *Feedback*, 70, 10-16.