

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

Student perceptions of Grammarly, teacher's indirect and direct feedback: Possibility of machine feedback

Tomoyuki Kawashima

Gunma University

tkawashima@gunma-u.ac.jp

Empirical data on automated writing evaluation (AWE) has been accumulating over the past several years, but previous research has concentrated on the impact of using machine feedback alone, and only a little is known about how learners view the usage of AI-assisted tools when combined with teacher feedback. More empirical evidence from instances in which machine and teacher feedback is used in combination is necessary to determine optimal approaches to incorporating AWE tools in classroom instruction. A study was conducted to assess Japanese college students' perceptions of three types of feedback when they experienced them in a certain order. Participants in two groups received Grammarly feedback on their first drafts, the teacher's indirect feedback on their second drafts, and the teacher's direct feedback on their third drafts. The questionnaire administered to the first group examined participant perceptions of the respective type of feedback, whereas the questionnaire given to the second group verified the strengths and weaknesses of each type of feedback and investigated the type of feedback participants believed was most effective in improving their writing in English. The content analysis approach was used for textual analysis of the responses to open-ended questions, and a chi-square test was performed for scale responses. The quantitative and qualitative analysis of questionnaire data revealed that, though machine feedback was favorably received, participants appreciated human direct and indirect feedback more highly for its appropriateness and reliability. Moreover, the largest number of participants rated the teacher's direct feedback as the most beneficial to their writing progress. The findings imply that machine feedback should not be used alone in writing instruction but as a supplement to teacher feedback. In addition, the participants' response suggests that using Grammarly feedback in a "form then content" sequence might be one feasible pedagogical approach in the classroom.

Keywords: automated writing evaluation, written corrective feedback, Grammarly, direct feedback, indirect feedback



The
JALT CALL
Journal
vol. 19 no.1

Introduction

In recent years, automated writing evaluation (AWE) systems, such as Criterion, Pigai, Grammarly, and many others, have become widely available, and it is hoped that such tools would assist not only students in being able to self-correct their writing but also teachers in providing corrective feedback more efficiently. AWE is a form of software that “uses artificial intelligence to evaluate essays and generate feedback” (Warschauer & Grimes, 2008, p.22). As early as 2006, Hyland and Hyland predicted that AWE software would be developed one after another and that feedback would be a focus of research. To support their prediction, many studies on AWE programs, including grammar checkers, have been conducted in recent years, for example, Criterion (Dikli & Bleyle, 2014; Li, 2021; Liao, 2016; Link, Mehrzad & Rahimi, 2022; Tsuda, 2014), Ginger (Daniels & Leslie, 2016), Grammark (Parra & Calero, 2019), Grammarly (Brown, 2017; Cavaleri & Dianati, 2016; Daniels & Leslie, 2016; Dembsey, 2017; Dizon & Gayed, 2021; Ebadi et al., 2022; Guo, Feng & Hua, 2021; John & Woll, 2018; Koltovskaia, 2020, 2022; Lazic, Thompson, Pritchard & Tsuji, 2020; Nang & Nikolov, 2021; O’Neill & Russel, 2019; Ummah & Bisriyah, 2022; Yousofi, 2022), MY Access (Mohsen & Abdulaziz, 2019), PEG Writing (Wilson & Czik, 2016), Pigai (Huang & Renandya, 2020; Jiang et al., 2020; Tian & Zhou, 2020; Zhang, 2020), Virtual Writing Tutor (John & Woll, 2018), and Write and Improve (Brown, 2017). While the initial focus of the research was on the feedback generated by AWE tools and learners’ responses to such feedback, though still limited in number, studies on how AWE applications would be used in English language teaching and learning came to be conducted, and more recently, the influence of AWE use on teacher feedback has been actively examined.

This paper focuses on Grammarly and explores the possibilities of using it in classroom instruction in Japan. Of the many AWE tools, Grammarly is likely the AWE technology with the greatest educational application in the future. This is because individuals, as well as educational institutions, can license Grammarly, and a free online version is available in addition to the paid version, albeit with limited functionality. In addition, while a few studies compared the use of AWE with its use in combination with teacher feedback, the procedures for giving AWE with teacher feedback have received little attention. The author believes that gaining insight into specific instructional approaches will contribute to the recent AWE research trend toward a greater focus on learners and teachers. The following sections review the literature regarding the advantages and disadvantages of Grammarly, the use of AWE with teacher feedback, direct and indirect feedback, and the Japanese context.

Grammarly feedback

Much has been argued about the advantages and disadvantages of Grammarly feedback. Grammarly's advantages for students include detailed and thorough feedback targeted at high-frequency errors (O'Neill & Russel, 2019), the convenience of receiving feedback immediately and as often as students like (Woodworth & Barkaoui, 2020), and an enhanced incentive for students to adjust their writing (O'Neill & Russel, 2019). On the other hand, some limitations have been identified, such as omissions or oversights of errors (John & Woll, 2018), inaccurate or flawed suggestions (John & Woll, 2018), too much feedback (Dembsey, 2017; Woodworth & Barkaoui, 2020), inadequate explanations and user knowledge required for applying feedback (Dembsey, 2017), and a narrow range of feedback (Parra & Calero, 2019).

Table 1 summarizes the research focus of the studies conducted with Grammarly, albeit it is not an exhaustive list. Of the three often contrasted writing qualities, complexity, accuracy, and fluency, or CAF for short, accuracy is the area most widely researched. Daniels and Leslie (2016) explored the correctness of Grammarly feedback itself, and Barrot (2021), Dizon and Gayed (2021), Ebadi et al. (2022), Guo et al. (2021), John and Woll (2018), and Nang and Nikolov (2021) investigated the impact of Grammarly feedback on writing accuracy. Ebadi et al. (2022) were unique in that they compared the impact on the accurate use of articles when Grammarly feedback was supplied alone versus when teacher feedback was offered in addition to Grammarly feedback. Three studies compared Grammarly feedback with feedback from other AWE tools: Brown (2017) with Write & Improve, Daniels and Leslie (2016) with Ginger and Microsoft Word, and John and Woll (2018) with Virtual Writing Tutor and Microsoft Word. The table shows that student perceptions of Grammarly feedback have been a popular research focus. However, all studies except for Koltovskaia (2020) and Nang and Nikolov (2021) investigated student perceptions when receiving feedback from an AWE system alone. The comparison of Grammarly feedback with teacher feedback was made only in four studies.

Table 1. Grammarly studies

Study	Context	Focus of study				
		Accuracy of feedback	Comparison with other AWE systems	Effects on CAF	Student perceptions	Comparison with teacher feedback
Barrot (2021)	Philippines			Y		
Brown (2017)	Japan		Y		Y	
Cavaleri & Dianati (2016)	Australia				Y	
Daniels & Leslie (2016)	Japan	Y	Y		Y	
Dembsey (2017)	US					Y
Dizon & Gayed (2021)	Japan			Y		
Ebadi et al. (2022)	Iran			Y		
Guo, Feng & Hua (2021)	China			Y		
John & Woll (2018)	Canada		Y	Y		
Koltovskaia (2020)	US				Y ¹	Y
Lazic et al. (2020)	Japan				Y	
Nang & Nikolov (2021)	Myanmar			Y	Y	Y
O'Neill & Russel (2019)	Australia				Y	Y
Ummah & Bisriyah (2022)	Indonesia				Y	
Yousofi (2022)	Afghanistan				Y	

Note. CAF stands for complexity, accuracy, and fluency. Y¹ = student engagement.

Use of AWE with teacher feedback

There has not been much research on the effects of instruction when AWE feedback was offered in addition to teacher feedback. Table 2 shows a small number of studies in which students received both types of feedback. Four studies, Dikli and Bleyle (2014), Ebadi et al. (2022), Link, Mehrzad, and Rahimi (2022), and Mohsen and Abdulaziz (2019), compared the effects of instruction in multiple conditions where students received both teacher and AWE feedback and where they received teacher feedback or AWE feedback alone. On the other hand, the remaining three studies, Koltovskaia (2020), Nang and Nikolov (2021), and Tian and Zhou (2020), explored the effects of instruction when teacher and AWE feedback was provided in a certain order. For example, in Koltovskaia (2020), teacher feedback on higher-order concerns was offered for the student's initial submission, followed by Grammarly feedback on lower-order concerns in their second submission. In Nang and Nikolov (2021), students received Grammarly feedback for the first essay, teacher feedback for the second essay, and both Grammarly and teacher feedback for their third and fourth essays. In Tian and Zhou (2020), students revised their drafts three times, first in response to Pigai, second to peer feedback, and third to teacher feedback.

Table 2. Studies in which students received both AWE and teacher feedback

Study	Context	AWE	Focus of study
Dikli & Bleyle (2014)	US	Criterion	teacher feedback/teacher & AWE feedback
Ebadi et al. (2022)	Iran	Grammarly	teacher feedback/AWE feedback/teacher & AWE feedback
Koltovskaia (2020)	US	Grammarly	teacher feedback => AWE feedback
Link, Mehrzad & Rahimi (2022)	Iran	Criterion	teacher feedback/teacher & AWE feedback
Mohsen & Abdulaziz (2019)	Saudi Arabia	MY Access	AWE feedback/teacher & AWE feedback
Nang & Nikolov (2021)	Myanmar	Grammarly	AWE feedback => teacher FB => AWE & teacher FB
Tian & Zhou (2020)	China	Pigai	AWE feedback => peer => teacher feedback

Students incorporate AWE feedback much less compared to teacher feedback. According to Tian and Zhou (2020), the uptake rate for AWE feedback was 24.2%, much lower than 55.3% for peer feedback and 85.3% for teacher feedback. Seen from the opposite angle, more AWE feedback was left unattended than teacher feedback. Link, Mehrzad, and Rahimi (2022) revealed that 24% of Criterion feedback was neglected compared to 12% of teacher feedback. Furthermore, as one of the benefits of using AWE feedback, it is expected to promote a division of labor, with AWE tools handling lower-order concerns and teachers addressing higher-order concerns (O'Neill & Russel, 2019; Woodworth & Barkaoui, 2020). In this regard, we have recently seen research on the impact of AWE feedback on teacher feedback, such as Jiang et al. (2020), Koltovskaia (2022), Li (2021), Link et al. (2022), and Wilson and Czik (2016). However, the results reported so far have not necessarily led to the division of labor. For instance, teachers offered comments both on higher- and lower-order concerns (Koltovskaia, 2022). The number of teachers' comments on higher-order concerns did not increase despite the inclusion of an AWE tool (Link et al., 2022).

Whether the feedback is beneficial depends on student engagement (Koltovskaia, 2020; Zhang & Hyland, 2018). Cavaleri and Dianati (2016) claimed that students must have a discerning eye for suggestions they accept because some of Grammarly's suggestions are erroneous. If Grammarly benefits more advanced writers, it is reasonable to anticipate that students with limited language skills will become over-reliant on feedback. Advanced learners, for instance, are more inclined to question AWE feedback, consult other resources to improve the accuracy of their drafts, and incorporate feedback selectively (Koltovskaia, 2020). Furthermore, the engagement of one student may change over time. Tian and Zhou (2020) discovered that five students' engagement with AWE, peer, and teacher feedback was dynamic, suggesting that it may change from time to time depending on the writing stages. Another factor that may be relevant when examining the effectiveness of AWE feedback is the learner's

English proficiency. According to Huang and Renandya (2020), Chinese EFL university students with low English proficiency perceived Pigai favorably. However, they did not necessarily find improvement in students' revision with the use of AWE feedback.

Though information about the methods was limited, Nang and Nikolov (2021) reported how useful students perceived feedback in grammar, vocabulary, content, organization, and other aspects of writing when they received teacher feedback, Grammarly feedback, or both teacher and Grammarly feedback (Table 3). It is worth noting that students perceived feedback as more useful in terms of grammar and vocabulary when teacher feedback was combined with Grammarly feedback than when teacher feedback was offered alone. However, the findings of this study may need to be viewed with care because Grammarly feedback was perceived more positively than teacher feedback in grammar and vocabulary, which contradicts prior research reporting student opinions in favor of teacher feedback (e.g., Dikli & Bleyle, 2014).

Table 3. Student perceptions of the usefulness of feedback (%)

	Grammar	Vocabulary	Content	Organization	Others
Teacher feedback	66.7	37.0	48.1	63.0	14.8
Grammarly feedback	88.9	77.8	0	0	7.4
Combined feedback	85.2	66.7	44.4	55.6	0

Note. Adapted from Nang and Nikolov (2021)

Direct and indirect feedback

Feedback is “input from a reader to a writer having the function of supplying information to the writer for modification” (Keh, 1990, p.294), and it is classified into direct and indirect feedback. Direct feedback entails supplying the correct form, whereas indirect feedback means suggesting the existence of an error in some way without offering the correction (Ferris & Roberts, 2001).

The research has yielded inconsistent results about student preferences for feedback. In a study conducted in Japan, Elwood and Bode (2014) found that even students with a high level of English proficiency preferred that all surface-level errors be corrected directly to the greatest extent possible. On the other hand, in a study of Lebanese college students, preferences for indirect feedback on the first drafts and direct feedback on the final drafts were reported (Diab, 2005). These inconclusive findings may be due to various factors impacting student feedback impressions. External influences may include, for example, culture, instructor abilities, the sequence in which feedback is offered, and the stage in the writing process. In addition, correction difficulty is likely to influence student perceptions of feedback. Ferris and Roberts (2001) revealed that students were more successful in editing errors in the “treatable” category (verbs, noun endings, and articles) than in the “untreatable” types (word choice and sentence structure). Similarly, internal aspects such as student personality,

motivation, proficiency levels, and familiarity with AWE technology may be considered.

Concerning short-term instructional effects, direct feedback is more effective in improving writing accuracy. Grammarly can reduce errors in vocabulary usage (diction), language use (grammar), and writing mechanics (spelling and punctuation) (Ghufron, 2019; Ghufron & Rosyida, 2018). Dizon and Gayed (2021) also noted that Japanese college students had significantly more lexical variation when using Grammarly to write. Nevertheless, it is documented that Grammarly is less efficient in improving the content and organization of writing (Ghufron, 2019), fluency in terms of the overall number of words, or syntactic complexity judged by the number of subordinate clauses (Dizon & Gayed, 2021). On the other hand, the long-term effects of error correction on accuracy improvement are inconclusive. Evidence in favor of direct feedback shows that it increased accuracy in the usage of the simple past tense and the definite article in new pieces of writing (Bitchener, Young, & Cameron, 2005) and in the number of form-related errors over three weeks (Van Beuningen, de Jong, & Kuiken, 2008). In recent years, Jamalinesari et al. (2015) observed that indirect feedback was more helpful than direct feedback in improving the learners' linguistic accuracy on new writing tasks. By counting the number of common local errors committed by the students over the ten consecutive class sessions, they revealed that students who received the teachers' indirect corrective feedback performed better than those who received direct feedback from teachers.

The Japanese context

Feedback on form remains important partly due to gatekeeping mechanisms such as high-stakes exams emphasizing writing accuracy greatly (Dikli & Bleyle, 2014). This priority for lower-order concerns may be a universal phenomenon, but the tendency is especially evident in Japan. There are some reasons for this. Japanese college students lack experience writing a decent amount of English freely. According to an investigation of writing tasks in the ten most popular Japanese high school writing textbooks (Kobayakawa, 2011), free writing tasks (13.1%) were limited in comparison to translation tasks (41.4%) and other controlled writing tasks (33.2%). This shows that students are used to translation, dictation, fill-in-the-blank, sentence combining, and reordering, but not free compositions. Learners, in general, have been observed to focus on lower-order concerns, such as surface and mechanical errors, while overlooking higher-order concerns, such as developing ideas and organization of their writings (Ashwell, 2000; Keh, 1990). Furthermore, the majority of Japanese students without formal writing training in process or paragraph writing tend to be even more concerned with lexical and grammatical problems (Elwood & Bode, 2014).

Although it is unclear whether the emphasis of Japanese college students on form influenced the research findings, Ashwell (2000) reported that the content-then-form pattern of teacher feedback and the reversed order of

form-to-content pattern of teacher feedback produced no significant difference in formal accuracy or content. In his study, when participants revised their drafts, they received one of the following patterns of feedback: content then form, form then content, form & content then form & content, and no feedback. Moreover, Ashwell (2000) found that whether the form feedback was offered first or second, Japanese college students were primarily concerned with form and acted on it first.

There is also a reason for the teacher to focus on lower-order concerns in Japan. Forty students per class are still usual in high school. According to the author's observations as a former high school teacher, being in charge of 80–120 students, teachers usually emphasized addressing lower-order concerns due to the limited time they could spend with each student. An exception was when the teacher prepared speech scripts for a few students who would compete in speech contests because the content and organization of a speech were critical for better evaluation in the contest.

Table 4 shows AWE studies conducted with Japanese students to the author's best knowledge. These studies had a modest number of participants, with Tsuda (2014) having the most at 40. All studies analyzed student impressions of AWE systems, with the exceptions of Blake (2020), which evaluated an in-house tool, and Dizon and Gayed (2021), which studied the impact of four writing sessions on lexical variation and grammatical errors. No previous study in Japan has compared student perceptions of Grammarly feedback with teacher feedback.

Table 4. AWE studies in the Japanese context

Study	Participant	AWE system	Focus of study	
			Effects on CAF	Student perceptions
Blake (2020)	10 college Ss	In-house tool	–	–
Brown (2017)	17 college Ss	Grammarly ¹		Y
Daniels & Leslie (2016)	12 college Ss	Grammarly ²		Y
Dizon & Gayed (2021)	32 college Ss	Grammarly	Y	
Lazic et al. (2020)	37 college Ss	Grammarly		Y
Tsuda (2014)	40 college Ss	Criterion		Y

Note. Grammarly¹ = Grammarly, Write & Improve. Grammarly² = Grammarly, Ginger, Microsoft Word.

Research questions

The literature review shows two shortcomings in prior research. First, only a few studies examined student perceptions when AWE feedback was provided in conjunction with teacher feedback. Furthermore, no such studies have been undertaken in the Japanese context. Another area that has yet to be thoroughly investigated in previous research is the order in which teacher feedback and AWE feedback are introduced to students. To address these research gaps, the author designed an experiment with Japanese college students. A three-step

revision process was devised in which students experienced three types of feedback in the sequence of Grammarly, teacher's indirect, and teacher's direct feedback. Students revised their first drafts with Grammarly feedback and their second drafts with the teacher's indirect feedback. Student final drafts were returned with the teacher's direct feedback. Teachers' direct feedback was provided at the end to avoid students (mis)taking it for the final answer. Because of the teacher's expert and evaluator positions, Japanese students tend to accept teacher feedback uncritically (Muncie, 2000). The following two research questions were formulated using this instructional approach.

RQ1. How do students perceive three types of feedback when they experience them in the order of Grammarly, teacher's indirect and direct feedback?

RQ2. Which of the three types of feedback do students find most helpful for improving writing in English?

Study

Participants

The instruction was given online to two intact groups of students. A total of 189 Japanese non-English major second-year college students participated, and they were all learners of English as a second language. A non-probability sampling method was used, and the participants were a convenience sample. One-hundred sixty students in the department were in four classes, and the author taught the students in three classes every year. At the time of the study, participants had received formal English education in school for at least seven years. The department's average TOEIC IP test score was 462, whereas the average score for the 222,895 Japanese first-year college students who took the test in 2019 was 438 (Education Testing Service, 2020), suggesting the participants in this study had slightly higher proficiency than the average college students in Japan. Because of the covid-19, the test score for Group 2 was unavailable. However, student proficiency levels were deemed comparable because the college admission rate remained constant across the two years. No attempt was made to homogenize the participants, and thus, there was a disparity in their English proficiency levels. Group 1 ($n = 75$) attended the study in the second semester of 2020, and Group 2 ($n = 114$) participated in the study in the first semester of 2021. No student participated in both studies.

Instruction

The author taught the course in both semesters and provided instruction to two groups of students using the three-step revision process (Table 5). In the first and second weeks, students had a group discussion on a given topic for 12 minutes in class, and as an assignment, they prepared two 100 to 150-word-long reports of the discussions. Students were told to adhere to the following organization. They should begin the report with the sentence, "I had a group discussion with Student A, Student B, and Student C about the Given Topic." Then, it

should be followed by a topic sentence summarizing the whole passage. After presenting a few examples from the discussion, they were instructed to close the passage with the restatement of the topic sentence. Students made reports of four group discussions on the following topics in a semester: hobbies, club activities, embarrassing experiences, and best memory.

After collecting the first drafts electronically via Moodle, the teacher submitted student writing to Grammarly. Later, the teacher downloaded the feedback from Grammarly as a PDF file and returned it to each student. Students revised their reports based on the machine feedback in the third week and submitted their second drafts. Then, the teacher read over their writing and provided indirect feedback. He highlighted some parts in yellow, which needed to be changed, and the files were returned to students. Students updated their writing in the fifth week and submitted their third drafts. The teacher gave direct feedback this time, correcting student writing. Grammarly and the teacher's indirect feedback focused on low-order concerns such as sentence structure, grammar, word choice, and spelling. The direct feedback from the teacher, on the other hand, addressed both low- and high-order concerns, such as the purpose, organization, and development. This was a cycle of the three-step revision process, and participants experienced this cycle two times during a semester. Receiving feedback twice a semester seemed a low frequency. However, given the time required for the teacher to review a large number of student drafts, one instruction cycle could not be shorter than 6–7 weeks. As a result, two cycles were the maximum frequency in a 15-week semester.

Table 5. Three-step revision process

Step	Assignment	Feedback	Type	Concerns
Step 1	Draft 1	Grammarly feedback	Mostly direct	Low order
Step 2	Draft 2	Teacher feedback	Indirect	Low order
Step 3	Draft 3	Teacher feedback	Direct	Low-and high-order

Grammarly feedback

In this study, the feedback generated by the Grammarly Premium version was shared with students in PDF format because the teacher did not have a license for all his students to use it. The PDF reports students received differed from the Grammarly feedback on the screen regarding the UI/UX designs. Take, for example, the feedback on two drafts of approximately 120 words each. Both the online and PDF versions of the feedback had the underlines in three different colors according to the type of feedback; red lines indicated accuracy problems, blue lines showed clarity suggestions, and green lines represented engagement issues. However, there were layout differences. Grammarly feedback on the screen is shown in Figure 1. From left to right, the online version was divided into three columns: text, feedback, and headings. By clicking the underline in the text, we could see the supplementary comment about the

corresponding feedback. Similarly, when we clicked the ‘overall score’ heading on the right, various metrics about the text, such as word count and readability, appeared. The PDF version, on the other hand, was seven pages long, with statistics on the first three pages. Pages 4–5 included the texts (Figure 2), and pages 6–7 had the comments (Figure 3). Suggestions were made in blunt language, such as “Faulty tense sequence,” as opposed to supplemental comments in sentences “Watch [in bold letters] seems to be in the wrong tense” in the web version. These differences in the UI/UX designs might have affected participants’ perceived usability.

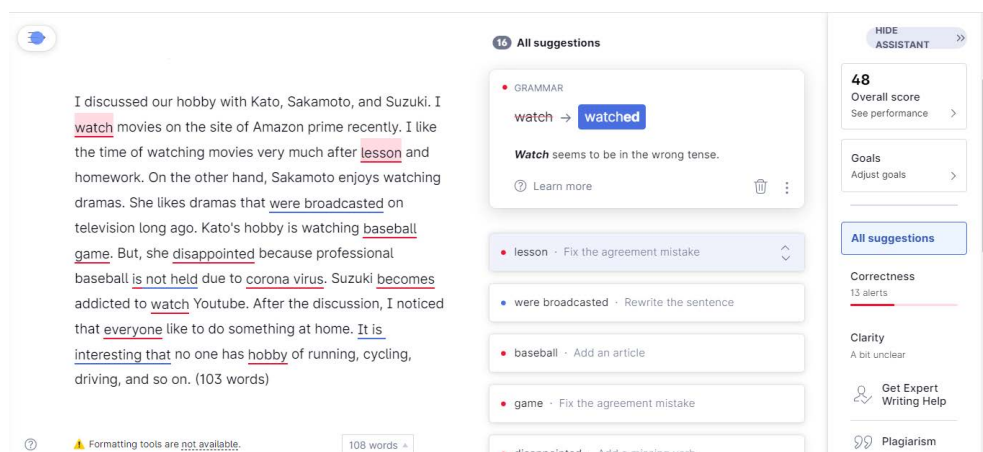


Figure 1. Grammarly feedback in the online version

I discussed our hobby with Kato, Sakamoto, and Suzuki. I watch¹ movies on the site of Amazon prime recently. I like the time of watching movies very much after lesson² and homework. On the other hand, Sakamoto enjoys watching dramas. She likes dramas that were broadcasted³ on television long ago. Kato's hobby is watching baseball⁴ game⁵. But, she disappointed⁶ because professional baseball is not held⁷ due to corona virus⁸. Suzuki becomes⁹ addicted to watch¹⁰ Youtube. After the discussion, I noticed that everyone¹¹ like to do something at home. It is interesting that¹² no one has hobby¹³ of running, cycling, driving, and so on. (103 words)

Figure 2. Grammarly feedback in the PDF version

1.	watch → watched	Faulty tense sequence	Correctness
2.	lesson → lessons	Incorrect noun number	Correctness
3.	were broadcasted	Passive voice misuse	Clarity
4.	a baseball, or the baseball	Determiner use (a/an/the/this, etc.)	Correctness
5.	game → games	Incorrect noun number	Correctness
6.	was disappointed	Incorrect verb forms	Correctness
7.	is → was	Faulty tense sequence	Correctness
8.	is not held	Passive voice misuse	Clarity
9.	the corona	Determiner use (a/an/the/this, etc.)	Correctness
10.	corona virus → coronavirus	Confused words	Correctness

Figure 3. Supplementary comments in the PDF version

During the orientation, the teacher informed students that Grammarly is AWE software operated by artificial intelligence. Furthermore, unlike Ashwell (2000), who gave teachers a 12-minute time limit to deliver three types of feedback on one draft to avoid the amount of feedback generating a difference in the comparison, no regulations were used to control the amount of feedback in this study. It should also be noted that only the teachers' direct feedback was able to comment on a wide range of issues, from word choice to organization.

Data collection

Two post-activity questionnaires were created. The first questionnaire with six open-ended questions (Appendix A) was administered to Group 1 in February 2021 to collect the participants' likes and dislikes about the three types of feedback. Questionnaire 2, with 15 items (Appendix B), was developed based on the responses collected by Questionnaire 1 and the test items used in previous research. It aimed to verify the advantages and disadvantages of the respective type of feedback more quantitatively and to examine which feedback participants found most helpful for improving their writing in English. Test items for feedback reliability (Q1 and Q6) and confidence enhancement (Q3 and Q8) were adopted from Cavaleri and Dianati (2016). Two questions asking the time students spent completing their drafts (Q11 and Q12) were adopted from Koltovskaia (2020). The questions about the strengths and weaknesses of Grammarly feedback (Q14 and Q15) were developed based on the questions used by Dembsey (2017) and the data obtained in Questionnaire 1. The second questionnaire was administered to Group 2 in July 2021. Both questionnaires were conducted online via Moodle at the end of the semester. It was one week

after students received the teacher’s direct feedback on their final drafts. The questionnaires were written in Japanese, and the author translated participant responses into English. Writing tasks were given as part of their coursework to all students. However, participation in the questionnaire survey was entirely voluntary. It was informed that failure to answer the questions would not affect their class grades and that they could opt out at any time.

Results

Student perceptions of three types of feedback (RQ1)

The responses obtained from Group 1 are represented in the first part of this section. Content analysis was used to code the answers to six open-ended questions. Questionnaire data were coded and categorized to extract themes by identifying keywords. The frequency of comments on each theme was then calculated. Table 6 shows the top two characteristics participants identified as the advantages and disadvantages of Grammarly feedback. The most common advantage participants felt was the provision of specific suggestions, followed by the ease of locating mistakes. Meanwhile, the two most common disadvantages were unnecessary or inappropriate suggestions that did not fit the context.

Table 6. Top two advantages and disadvantages of Grammarly feedback (N = 75)

Comments	Number of comments
Advantages (n = 75)	
Specific alternatives are provided	36
Easy to locate mistakes	11
Disadvantages (n = 58)	
Unnecessary comments	18
Suggestions do not fit the context	14

Similarly, outstanding benefits and limitations of the teacher’s indirect feedback with yellow highlighters are presented in Table 7. The responses indicated that almost half of the participants saw it as a chance to contemplate their writing. However, almost the same number of respondents cited the difficulty in determining what to change as a disadvantage. A subsequent examination revealed that participants who viewed it favorably as an opportunity to think for themselves and those who reported difficulties recognizing errors were not identical.

Table 7. Advantages and disadvantages of teacher’s indirect feedback (N = 75)

Comments	Number of comments
Advantages (n = 71)	
It allowed me to think	32
Easy to understand where to fix	17
Disadvantages (n = 38)	
Hard to know what to fix	34

Compared with the other two types of feedback, the number of comments submitted about the teacher’s direct feedback was small (Table 8). Of the 57 comments about the advantage, 55 were about the ease of understanding. On the other hand, there were as few as six comments on the disadvantage. Their topics were varied. For instance, one student wrote, “There were corrections that did not match up with others, and I wasn’t sure which one I should prioritize.” Another student was dissatisfied with the teacher’s extra feedback, saying, “I wished the teacher didn’t point out the parts that he hadn’t pointed out in the first round. I would have understood if he had corrected only the same parts as his first time.” Another student stated, “Dare I say the teacher’s direct feedback was harder to see visually than his marker-only indirect feedback.”

Table 8. Advantages and disadvantages of teacher’s direct feedback (N = 75)

Comments	Number of comments
Advantages (n = 57)	
Easy to understand	55
Disadvantages (n = 6)	
Miscellaneous topics	6

The data obtained from Group 2 using Questionnaire Two are provided in the following section. Table 9 presents participant responses to Questions 1 to 10. They were asked to indicate how strongly they felt Grammarly and teacher feedback were reliable, helpful in improving writing skills, and motivating to write in English on a four-point Likert scale. Cronbach’s alpha calculated to assess questionnaire reliability was .892, indicating that the responses to the ten questions were reliable. The number of each rating, the mean score calculated with ‘Strongly disagree’ as 1, ‘Disagree’ as 2, ‘Agree’ as 3, and ‘Strongly Agree’ as 4, standard deviation, p-values, and effect size in Cohen’s *h* are tabulated.

To determine whether there was a significant difference in student perceptions between Grammarly and teacher feedback, the responses “Strongly disagree” and “Disagree” were merged, as were “Agree” and “Strongly agree.” The Fisher’s Exact Test was then computed. A significant difference at the 1% level was present in student perceptions of teacher feedback being more motivating to write English than Grammarly feedback. Furthermore, student judgments

of whether the feedback is reliable or enhances confidence in writing English differed considerably at the 5% level. In either case, students perceived teacher feedback more favorably. An interesting disparity in perceptions was discovered. No significant difference was found in student perceptions of whether the feedback improves English composition. On the other hand, despite the fact that the difference was not statistically significant, students believed that teacher feedback will lead to better writing abilities more firmly than Grammarly feedback.

Table 9. How did participants perceive Grammarly and teacher feedback? (N = 114)

Feature	Strongly disagree	Disagree	Agree	Strongly agree	M	SD	<i>p</i>	Cohen's <i>h</i>
Reliable								
Grammarly feedback	0	9	71	34	3.22	0.57	.019*	.382
Teacher feedback	0	1	34	79	3.68	0.48		
Motivating to write English								
Grammarly feedback	2	14	75	23	3.04	0.63	.009**	.391
Teacher feedback	1	3	48	62	3.50	0.60		
Enhance confidence in writing English								
Grammarly feedback	2	19	71	22	2.99	0.66	.030*	.318
Teacher feedback	2	7	45	60	3.43	0.69		
Improve English composition								
Grammarly feedback	1	11	64	38	3.22	0.65	.128	.239
Teacher feedback	0	5	35	74	3.61	0.57		
Lead to improved writing skills								
Grammarly feedback	0	9	70	35	3.23	0.58	.059	.304
Teacher feedback	1	1	42	70	3.59	0.56		

Note. Significance, $p < .01^{**}$, $p < .05^{*}$; Effect size, large = .80, medium = .50, small = .20

Questions 11 and 12 dug into teachers' concerns that students would be extremely reliant on Grammarly. Students were asked to answer how many minutes they spent revising their drafts with Grammarly and the teacher's indirect feedback, respectively. Participants spent 43.8 minutes, on average, editing their drafts based on Grammarly feedback, whereas the revision using the teacher's indirect feedback took 49.9 minutes. This finding, however, should be treated with caution because this study did not examine the successful uptake of corrective feedback. Additional investigation is necessary to conclude

that both Grammarly and the teacher’s indirect feedback enhanced students’ self-correction.

Types of feedback participants found most helpful for improving their writing in English (RQ2)

The following two tables indicate participant responses to Questions 13 and 14 of Questionnaire Two, which asked about the feature of Grammarly feedback that participants thought would be most helpful for improving their writing in English (Table 10) and most annoying (Table 11). Of the four choices, the largest percentage of respondents (29.8%) considered it most beneficial for errors they were not aware of to be pointed out, followed by the presentation of the expressions and words different from the ones they used (27.2%). The offer of concrete suggestions to fix errors was welcomed by a somewhat smaller but close to a quarter of the participants (23.7%). To sum up, Grammarly’s ability to provide new knowledge in a tangible manner was appreciated by four out of five participants (80.7%) cumulatively. On the other hand, the remaining one-fifth of the respondents (19.3%) supported Grammarly’s unique feature of displaying errors in different colors based on error types.

Table 10. Grammarly’s features participants found most helpful for improving writing

Feature	Number	Percentage
Point out errors that I don’t notice	34	29.8
Suggest different expressions and words	31	27.2
Present a concrete way of correction	27	23.7
Present errors in different colors based on error types	22	19.3
	114	100

Meanwhile, as much as 30.7% of participants reported being irritated by unnecessary indications of errors that did not require correction. Similarly, 23.7% were annoyed by the inappropriate suggestions for the context. This implies that combining the two percentages mentioned above, the majority of respondents (54.4%) found Grammarly’s improper suggestions to be upsetting. Furthermore, a brief review of the drafts written by students pointing out these disadvantages revealed that most of the mistakes they felt unnecessary to modify were the proposals to replace Japanese names. An example of inappropriate suggestions for the context was the advice to change the word “supplements” to “accessories” and “take” to “carry” or “handle” in the sentence, “This is because *supplements* are simple to *take*.” The rest of the participants found the absence of practical assistance quite frustrating. Nearly a quarter of the respondents (24.6%) were bewildered by the indication of errors in lexical choice without understanding why specific phrases were unsuitable. For example, in a sentence, “So, in her case, every meal is provided since their parents prepare it for them,” one respondent received a suggestion to modify

the word “meal” to “dinner.” Moreover, 21.1% were at a loss regarding what to do with the writing because concrete suggestions were not provided. Students who made this point may have been unhappy with the blunt suggestion as follows; the sentence “By working part-time, the extra mental burden increases, and it is not possible to concentrate on study” was deemed “unclear sentences,” as was a sentence “Some passersby looked her bad luck. Ito [student name] was a volleyball player. She played in final round of tournament” was labeled as “monotonous sentences” with no explanation.

Table 11. Grammarly’s features participants found most annoying

Feature	Number	Percentage
Point out the part that doesn’t need to be fixed	35	30.7
Point out the error in the selection of phrases, but I don’t understand why it’s wrong	28	24.6
Suggest words that don’t suit the context	27	23.7
Point out the improvement of the text, but I don’t know what to do without specific suggestions	24	21.1
	114	100

Question 15 asked participants’ perceived efficacy of feedback. When asked which of the three types of feedback they found most beneficial in improving their writing in English and the reasons (Table 12), 49.1% of the respondents preferred the teacher’s direct feedback, followed by 36.4% choosing the teacher’s indirect feedback. Grammarly was supported by nearly 14.5% of the participants.

What was outstanding was that 30 of the 40 respondents who selected the teacher’s indirect feedback mentioned the opportunity to think as the reason. One student commented, “With the feedback from AI, corrections were made even on the parts that were not wrong, but with teacher feedback with markers, mistakes were pointed out appropriately. Knowing which parts were wrong and analyzing why they were incorrect helped me enhance my English abilities.” Another student responded, “AI’s feedback and [teacher] feedback with corrections showed where and how to make changes. I felt they were helpful, but marker-only feedback was better for improving my skills.” Only this group of participants mentioned the opportunity to think as the reason. Grammarly mostly provides direct feedback. As a result, students may have considered the teacher’s indirect feedback even more helpful in studying independently, which they assumed would result in improved writing abilities.

The fact that not a few respondents in favor of the teacher’s indirect and direct feedback cited a reason “the feedback is supplied after the contents have been understood” implies that many participants were doubtful of the reliability of Grammarly feedback. Some respondents expressed their distrust of Grammarly feedback because they had witnessed proper nouns in Japanese being changed unnecessarily and unsuitable suggestions that did not fit the context.

Though participants were not requested explicitly, three respondents mentioned their thoughts on the three-step revision process. They all agreed on the sequence of starting with Grammarly feedback, then moving on to the teacher's indirect feedback, and lastly, the teacher's direct feedback. One student choosing the teacher's indirect feedback wrote, "I think providing feedback only with markers, followed by feedback with fixes, is the best sequence." Another participant who chose the teacher's direct feedback as the most helpful added, "However, I think the present three-step process is necessary because if we had been given teacher's direct feedback from the beginning, we would not have come up with so many good sentences." This comment supports Muncie's (2000) argument and verifies how strongly Japanese students regard teacher feedback as absolute owing to the teacher's authoritarian status. Still another participant who found Grammarly feedback the most beneficial commented, "The corrections offered by the AI at the beginning helped me to correct basic problems in grammar and word use, etc. The teacher corrections allowed me to fix more detailed parts of the writing that the AI could not do, and I believe I was able to create better sentences as a result."

Table 12. Feedback participants found most helpful for improving their writing in English

Type of feedback & reasons	Number of reasons (N = 110)
Grammarly feedback (n = 16, 14.5%)	
It provides concrete suggestions.	8
It explains why it is incorrect	5
Others	3
Teacher's indirect feedback (n = 40, 36.4%)	
It gives me a chance to think for myself.	30
Feedback is given after understanding the contents.	6
Others	4
Teacher's direct feedback (n = 54, 49.1%)	
It provides concrete suggestions.	24
Feedback is given after understanding the contents.	13
It gives new insights and learnings.	9
It is easy to understand.	4
Others	4

Discussion

This research examined student perceptions of three types of feedback when students were provided in the sequence of Grammarly feedback, the teacher's indirect feedback, and the teacher's direct feedback, and the types of feedback students considered most beneficial for improving their writing in English. Participants' perceptions of Grammarly feedback illustrated the possibilities and limitations of its application in classroom instruction. Though they regarded both Grammarly and teacher feedback as effective in

improving English composition, only a small number of participants selected Grammarly feedback as the most helpful for improving their writing in English. Unnecessary indications of errors that did not require correction and unsuitable recommendations that did not fit the context were causes of participants' irritation. Participants seemed to believe teacher feedback is more trustworthy than machine feedback because they thought that the AI did not understand the meaning of their writing. Many participants preferred teacher feedback and mentioned that the teacher provided feedback after he understood the contents. Furthermore, it was suggested that Grammarly feedback could not match teacher feedback in increasing participants' motivation and confidence in writing English.

Looking at the teacher's direct feedback, only favorable remarks on the ease of understanding were received. Meanwhile, the teacher's indirect feedback received positive and negative evaluations from participants. Almost half of the participants regarded it as an opportunity to contemplate their writing. At the same time, almost the same number of participants cited the difficulty in determining what to change as a disadvantage. The teacher's direct feedback was chosen to be most beneficial for improving their writing in English by the greatest number of participants (49.1%), followed by the teacher's indirect feedback (36.4%) and Grammarly feedback (14.5%). Especially noteworthy in this study was a finding that more than one-third of the participants regarded the teacher's indirect feedback as most helpful, and three-fourths of the respondents mentioned the opportunity to think as the reason for their selection.

These findings provide us with some pedagogical implications. First, aside from Ebadi et al. (2022) and Mohsen and Abdulaziz (2019), more evidence of the advantages of using Grammarly feedback in addition to, rather than as a replacement for, teacher feedback (O'Neill & Russel, 2019; Woodworth & Barkaoui, 2020) was obtained. In contrast to teacher feedback, Grammarly feedback contains unnecessary or improper recommendations that might annoy students. Therefore, it is best to provide both Grammarly and teacher feedback. Second, in cases where multiple types of feedback are offered, Grammarly and the teacher's indirect feedback should be given before the teacher's direct feedback. It is conceivable from the participants' unfavorable reactions detected in this study that guidance ending with Grammarly and the teacher's indirect feedback may result in frustrating some students with incomprehensible feedback. As Dikli and Bleyle (2014) noted, Japanese students, including so-called strong students, have a strong desire to have their grammatical errors and inappropriate English use corrected. Some Grammarly feedback on sentences and the teacher's indirect feedback require students to think and react on their own. However, a certain number of students find it difficult to do so. The teacher's direct feedback, if provided at the end, may reduce such student dissatisfaction.

Some suggestions for further study can be made in relation to the sequence of feedback. Partly due to Japan's particular setting, this study adopted a 'form then content' feedback pattern. Although Ashwell (2000) found no significant difference in form accuracy and content score with the reversed pattern, a

replication study with a ‘content then form’ sequence (e.g., Koltovskaia, 2020) is worth considering. However, owing to the limitations of Grammarly and the teacher’s indirect feedback not being able to refer to the organization, the optimal sequence may be giving the teacher’s direct feedback after Grammarly and/or the teacher’s indirect feedback. Though participants in this study reported being annoyed with Grammarly feedback, no such reactions were reported in Koltovskaia (2020). This absence of negative reactions is most likely due to variations in the research context. Participants in this study were non-English major college students studying English as a required subject once a week. Meanwhile, participants in Koltovskaia (2020) were two L2 learners of English enrolled in an undergraduate writing course at a university in the US. Furthermore, only eight of the 17 students enrolled volunteered to participate, and only two completed all the requirements of the study. For this reason, if replication of her study concluding with Grammarly feedback were conducted with a large sample in Japan or in other EFL countries, it would be unlikely that the same favorable response are obtained.

Although many students perceived the teacher’s indirect feedback as the most beneficial to improve their writing abilities, the author speculates that they might have responded differently if they had not received the teacher’s direct feedback afterward. Moreover, giving Grammarly feedback after the teacher’s direct feedback may confuse students if the teacher and Grammarly offer conflicting suggestions for the same expression. One participant brought out the discrepancy between different types of feedback in his/her comment. This will be a timely research topic that should be investigated. Though Grammarly feedback was provided under teacher control in this study, students may be using its free version on their own at any time. Therefore, it is essential to ascertain how students will react if they receive inconsistent Grammarly and teacher feedback.

Finally, this research is not free from limitations. Participants were chosen using convenience sampling, and they were students enrolled in the author’s course. This entails the possibility that participants may have found it difficult to criticize teacher feedback. Another limitation is the novelty of machine feedback. It was the first time that all the participants received Grammarly feedback, while they were used to receiving direct and indirect feedback from the teacher. Therefore, the unfamiliarity of machine feedback may have attracted more attention from participants, which may have affected the findings. Furthermore, participants in this study received Grammarly feedback in PDF format from their teacher twice a semester without using Grammarly themselves. Thus, they were not able to experience the usability of AWE tools to the greatest extent, which is considered a strong point of machine feedback (Woodworth & Barkaoui, 2020). This lack of frequency and direct experience may have influenced the outcome of this study.

Conclusion

This study described Japanese college students' perceptions of three types of feedback when they were offered in the order of Grammarly, teacher's indirect, and teacher's direct feedback. In addition, it provided the types of feedback students found most beneficial for improving their writing in English. While participants valued concrete suggestions offered by Grammarly, they rated the teacher's direct and indirect feedback as more reliable and helpful in improving their language skills than Grammarly feedback. Because of the lower reliability participants perceived on machine translation, it will be prudent to avoid relying solely on Grammarly feedback. Furthermore, based on the findings of this study, the author recommends specific sequences as better choices when Grammarly feedback is provided with teacher feedback. Grammarly feedback should be provided prior to the teacher's direct feedback. Moreover, if the teacher's indirect feedback is added, Grammarly and the teacher's indirect feedback can be included in any order, but both must come before the teacher's direct feedback.

Future research should investigate how the order of Grammarly and teacher feedback influences the successful uptake of feedback and overall evaluation of writing. The uptake rates for machine feedback reported by Tian and Zhou (2020) and Link, Mehrzad, and Rahimi (2022) will be good references. In addition, using machine feedback with teacher feedback, the impact of whether form or content is reviewed first would need to be examined. A replication study with Grammarly and teacher feedback using a 'content then form' sequence (e.g., Koltovskaia, 2020) will be worthwhile. Furthermore, how students react to feedback when Grammarly and their teacher offer conflicting feedback on the same expression will be an intriguing and timely issue worth researching. With the advancement of AWE systems, such cases will arise anytime and perplex students and teachers. Whether teachers like it or not, the day is coming when their students use online AWE technology. Therefore, answers to such a research agenda will provide teachers with new insight that will allow them to work in tandem with AWE software.

Acknowledgment

An earlier version of this paper was presented at JALT2021 Conference.

References

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9(3), 227–257. [https://doi.org/10.1016/s1060-3743\(00\)00027-8](https://doi.org/10.1016/s1060-3743(00)00027-8)
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>

- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14, 191–205. <https://doi.org/10.1016/j.jslw.2005.08.001>
- Blake, J. (2020). Genre-specific error detection with multimodal feedback. *RELC Journal*, 51(1), 179–187. <https://doi.org/10.1177/0033688219898282>
- Brown, M. H. (2017). English-learner perceptions of automatic writing evaluations. *The language and Media Learning Research Center Annual Report, Kanda University of International Studies*, 127–137. <http://id.nii.ac.jp/1092/00001458/>
- 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 & Learning*, 10(1), A223–A236. <https://journal.aall.org.au/index.php/jall/article/view/393>
- Daniels, P., & Leslie, D. (2016). Grammar software ready for EFL writers? *OnCue Journal*, 9(4), 391–401. https://jaltcue.org/files/OnCUE/OCJ9.4/OCJ9.4_pp391_401_AC_Daniels_Leslie.pdf
- Dembsey, J. M. (2017). Closing the Grammarly gaps: A study of claims and feedback from an online grammar program. *The Writing Center Journal*, 36.1, 63–100. <https://jmdembsey.com/2017/11/01/closing-the-grammarly-gaps/>
- Diab, R. L. (2005). EFL university students' preferences for error correction and teacher feedback on writing. *TESL Reporter*, 38(1), 27–51.
- Dikli, S., & Bley, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17. <https://doi.org/10.1016/j.asw.2014.03.006>
- Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, 17(2), 74–92. <https://doi.org/10.29140/jaltcall.v17n2.336>
- Ebadi, S., Gholami, M., & Vakili, S. (2022). Investigating the effects of using Grammarly in EFL writing: The case of articles. *Computers in the Schools*, 1–21. <https://doi.org/10.1080/07380569.2022.2150067>
- Education Testing Service (2020). *TOEIC Program Data & Analysis 2020*.
- Elwood, J. A., & Bode, J. (2014). Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System*, 42, 333–343. <https://doi.org/10.1016/j.system.2013.12.023>
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161–184. [https://doi.org/10.1016/s1060-3743\(01\)00039-x](https://doi.org/10.1016/s1060-3743(01)00039-x)
- Ghufron, M. A. (2019). Exploring an automated feedback program “Grammarly” and teacher corrective feedback in EFL writing assessment: Modern vs. traditional assessment. *Proceedings of the 3rd English Language and Literature International Conference, ELLiC, 27th April 2019, Semarang, Indonesia*. <https://doi.org/10.4108/eai.27-4-2019.2285308>
- Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Lingua Cultura*, 12(4), 395–403. <http://repository.ikipgribojonegoro.ac.id/id/eprint/543>

- Guo, Q., Feng, R., & Hua, Y. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 35(9), 2312–2331.
<https://doi.org/10.1080/09588221.2021.1879161>
- Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in language learning and teaching*, 14(1), 15–26.
<https://doi.org/10.1080/17501229.2018.1471083>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101.
<https://doi.org/10.1017/S0261444806003399>
- Jamalinesari, A., Rahimi, F., Gowhary, H., & Azizifar, A. (2015). The effects of teacher-written direct vs. indirect feedback on students' writing. *Procedia-Social and Behavioral Sciences*, 192, 116–123.
<https://doi.org/10.1016/j.sbspro.2015.06.018>
- Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, 93, 1–11.
<https://doi.org/10.1016/j.system.2020.102302>
- John, P., & Woll, N. (2018). Using grammar checkers in the ESL classroom: the adequacy of automatic corrective feedback. *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018*, 118–123.
<https://doi.org/10.14705/rpnet.2018.26.823>
- Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294–304.
<https://doi.org/10.1093/elt/44.4.294>
- Kobayakawa, M. (2011). Analyzing tasks in Japanese high school writing textbooks. *ARELE*, 22, 137–152. https://doi.org/10.20581/arele.22.0_137
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450.
<https://doi.org/10.1016/j.asw.2020.100450>
- Koltovskaia, S. (2022). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, 1–15. <https://doi.org/10.1017/S0958344022000179>
- Lazic, D., Thompson, A., Pritchard, T., & Tsuji, S. (2020). Student preferences: using Grammarly to help EFL writers with paraphrasing, summarizing, and synthesizing. In K.-M. Frederiksen, S. Larsen, L. Bradley, & T. Thouesney (Eds.), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 183–189). Research-publishing.net.
<https://doi.org/10.14705/rpnet.2020.48.1186>
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505. <https://doi.org/10.1016/j.system.2021.102505>

- Liao, H-C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308-319.
<https://doi.org/10.1093/elt/ccv058>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634.
<https://doi.org/10.1080/09588221.2020.1743323>
- Mohsen, M. A., & Abdulaziz, A. (2019). The effectiveness of using a hybrid mode of automated writing evaluation system on EFL students' writing. *Teaching English with Technology*, 19(1), 118-131.
<https://www.cceol.com/search/article-detail?id=737385>
- Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT Journal*, 54(1), 47–53. <https://doi.org/10.1093/elt/54.1.47>
- Nang, K. T., & Nikolov, M. (2021). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher*. <https://doi.org/10.1007/s40299-021-00625-2>
- O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42–56.
<https://doi.org/10.14742/ajet.3795>
- Parra, G. L., & Calero, S. X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, 12(2), 209–226. <https://doi.org/10.29333/iji.2019.12214a>
- Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, 91, 102247. <https://doi.org/10.1016/j.system.2020.102247>
- Tsuda, N. (2014). Implementing Criterion (automated writing evaluation) in Japanese college EFL classes. *Language and Culture: The Journal of the Institute for Language and Culture*, 18, 25–45.
<http://doi.org/10.14990/00000561>
- Ummah, L. K., & Bisriyah, M. (2022). EFL students' perception on Grammarly premium's feedback and dealing with inaccuracies. *JEES (Journal of English Educators Society)*, 7(2), 163–172.
<https://doi.org/10.21070/jees.v7i2.1687>
- van Beuningen, C. G., de Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *International Journal of Applied Linguistics*, 156(1), 279–296.
<https://doi.org/10.2143/ITL.156.0.2034439>
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36.
<https://doi.org/10.1080/15544800701771580>
- Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, 100, 94–109.
<https://doi.org/10.1016/j.compedu.2016.05.004>

- Woodworth, J., & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in ESL classroom. *TESL Canada Journal*, 37(2), 234–247.
<https://doi.org/10.18806/tesl.v37i2.1340>
- Yousofi, R. (2022). Grammarly deployment (in)efficacy within EFL academic writing classrooms: an attitudinal report from Afghanistan. *Cogent Education*, 9(1), 2142446. <https://doi.org/10.1080/2331186X.2022.2142446>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102.
<https://doi.org/10.1016/j.asw.2018.02.004>

Appendix A

Questionnaire One administered to Group 1

AIによるフィードバックについて

Q1 良かった点があれば、その理由も添えて書いてください。

Q2 改善したほうがよい点があれば、その理由も添えて書いてください。

教師による1回目のフィードバック(マーカーのみ)について

Q3 良かった点があれば、その理由も添えて書いてください。

Q4 改善したほうがよい点があれば、その理由も添えて書いてください。

教師による2回目のフィードバック(添削あり)について

Q5 良かった点があれば、その理由も添えて書いてください。

Q6 改善したほうがよい点があれば、その理由も添えて書いてください。

About AI's Feedback

Q1 If you think of any good points, please write them down with the reasons.

Q2 If you think of any points that could be improved, please write them down with the reasons.

About the teacher's first feedback with marking only

Q3 If you think of any good points, please write them down with the reasons.

Q4 If you think of any points that could be improved, please write them down with the reasons.

About the teacher's second feedback with corrections

Q5 If you think of any good points, please write them down with the reasons.

Q6 If you think of any points that could be improved, please write them down with the reasons.

Questionnaire Two administered to Group 2

それぞれのフィードバックについて、あなたはどのように感じましたか。選択肢の中からあなたの考えに最も近いもの、1つ選んでください。

選択肢

1. 全くそう思わない
2. ほとんどそう思わない
3. どちらかというと思う
4. そう思う

Q1. AIによるフィードバックは頼りになる

Q2. AIによるフィードバックは英作文のモチベーションを高める

Q3. AIによるフィードバックは英作文の自信を高める

Q4. AIによるフィードバックは作文の内容をよくする

Q5. AIによるフィードバックは英作文能力の向上につながる

Q6. 教師によるフィードバックは頼りになる

Q7. 教師によるフィードバックは英作文のモチベーションを高める

Q8. 教師によるフィードバックは英作文の自信を高める

Q9. 教師によるフィードバックは作文の内容をよくする

Q10. 教師によるフィードバックは英作文能力の向上につながる

Q11. あなたがAIによるフィードバックをもとに修正済みJournalを作成するのに要した時間を、分で答えてください。

Q12. あなたが教師によるフィードバック(マーカーのみ)をもとに校正済みJournalを作成するのに要した時間を、分で答えてください。

Q13. 今回は、1つの作文に対して、次の3種類のフィードバックを、①AIによるフィードバック、②教師によるフィードバック(マーカーのみ)、③教師によるフィードバック(添削あり)の順番で実施しました。3つのフィードバックのうち、あなたが英作文力の向上に最も役立つと感じたのはどれですか。理由と一緒に答えてください。

Q14. 次のAIによるフィードバックの特徴の中で、あなたの英作文力の向上に最も役立つと思うものを1つ選んでください。

1. 自分が気づかない間違いを指摘してくれる
2. 別の表現方法や単語を教えてくれる
3. 具体的な修正方法が提示されている
4. どこが間違っていて、どういう間違い方をしているかが色分けされている

Q15. 次のAIによるフィードバックの特徴の中で、あなたが最も困ったものを1つ選んでください。

1. 文脈に合わない単語が提案される
2. 文章の改善を指摘されても、具体的な改善案がないのでどうしたらよいか分からない
3. 語句の選択間違いを指摘されるが、なぜ間違っているのか分からない
4. 直さなくてよい部分が指摘される

How did you feel about each feedback? Please choose one of the options that best describes your opinion.

Options:

1. Strongly disagree
2. Disagree
3. Agree
4. Strongly agree

Q1. AI's feedback is reliable.

Q2. AI's feedback is motivating to write English.

Q3. AI's feedback enhances confidence in writing English.

Q4. AI's feedback improves English composition.

Q5. AI's feedback leads to improved writing skills.

Q6. Teacher feedback is reliable.

Q7. Teacher feedback is motivating to write in English.

Q8. Teacher feedback enhances confidence in writing English.

Q9. Teacher feedback improves English composition.

Q10. Teacher feedback leads to improved writing skills.

Q11. Answer the time you spent completing Draft 2 based on the AI's feedback in minutes.

Q12. Answer the time you spent completing Draft 3 based on teacher feedback in minutes.

Q13. This time, the following three types of feedback were given for each essay in the following order: 1) AI's feedback, 2) teacher feedback (with marking only), and 3) teacher feedback (with corrections). Which of the three types of feedback did you find most useful in improving your English writing skills? Please answer with your reasons.

Q14. Choose one from the following features of the AI's feedback that you think will be most helpful for improving your writing in English.

1. Point out errors that I don't notice.
2. Suggest different expressions and words.
3. Present a concrete way of correction.
4. Present errors in different colors based on error types.

Q15. Choose one from the following features of the AI's feedback that you found most annoying.

1. Point out the part that doesn't need to be fixed.
2. Point out the error in the selection of phrases, but I don't understand why it's wrong.
3. Suggest words that don't suit the context.
4. Point out the improvement of the text, but I don't know what to do without specific suggestions.

