How Teacher and *Grammarly* Feedback Complement One Another in Myanmar EFL Students' Writing

Nang Kham Thi[1] · Marianne Nikolov[2]

Abstract Providing feedback on students' writing is considered important by both writing teachers and students. However, contextual constraints including excess workloads and large classes pose major and recurrent challenges for teachers. To lighten the feedback burden, teachers can take advantage of a range of automated feedback tools. This paper investigated how automated feedback can be integrated into traditional teacher feedback by analyzing the focus of teacher and Grammarly feedback through a written feedback analysis of language- and content-related issues. This inquiry considered whether and how successfully students exploited feedback from different sources in their revisions and how the feedback provisions helped improve their writing performance. The study sample of texts was made up of 216 argumentative and narrative essays written by 27 low-intermediate level students at a Myanmar university over a 13-week semester. By analyzing data from the feedback analysis, we found that Grammarly provided feedback on surface-level errors, whereas teacher feedback covered both lower- and higher-level writing concerns, suggesting a potential for integration. The results from the revision analysis and pre- and post-tests suggested that students made effective use of the feedback received, and their writing performance improved according to the assessment criteria. The data were triangulated with self-assessment questionnaires regarding students' emic perspectives on how useful they found the feedback. The pedagogical implications for integrating automated and teacher feedback are presented.

Highlights

- This study investigated how automated feedback can be integrated into traditional teacher feedback.
- Characteristics of teacher and *Grammarly* feedback differ in terms of feedback scope.
- Students were able to successfully revise their errors regardless of the source of feedback.
- Provision of feedback led to statistically significant improvement in language and content aspects of writing.
- Effective integration of *Grammarly* in writing instruction might increase the efficacy of teacher feedback, affording it to focus on higher-level writing skills.

Introduction

Writing is an essential component of language learners' literacy development in school curricula, as well as a catalyst for personal and academic advancement. Providing feedback to students' written texts is a common teaching practice for improving students' writing skills. Investigating the effectiveness of written feedback on writing performance is a burgeoning field of inquiry, and many

✉ Nang Kham Thi
nang.kham.thi@edu.u-szeged.hu

Marianne Nikolov
nikolov.marianne@pte.hu

1 Doctoral School of Education, University of Szeged, 32-34, Petőfi Sgt., Szeged H-6722, Hungary

2 University of Pécs, 6 Ifjúság u., Pécs H-7624, Hungary

researchers (e.g., Ferris, 2004, 2007; Karim & Nassaji, 2018; Lee, 2009) have stressed its importance. Ferris (2004) suggested that feedback helps bridge the gap between students' present knowledge, which indicates areas of potential improvement, and the target language that they need to acquire.

Providing feedback on students' writing requires a great deal of time and effort on the teachers' part (Zhang, 2017). Contextual issues, including time constraints, excess workloads, and large classes, further increase the feedback burden. In response, automated writing evaluation (AWE) tools have come to be used to complement teacher feedback in writing classes (Wilson & Czik, 2016). In line with the favorable evidence of the reliability of AWE feedback (Li et al., 2015), L2 writing researchers (e.g., Koltovskaia, 2020; Ranalli, 2018) recommend integrating automated feedback into writing instruction to increase the efficacy of teacher feedback by freeing up teachers' time to focus less on lower-order concerns (e.g., grammar and mechanics) and turn more to higher-order concerns (e.g., content and organization).

Therefore, it is of great importance to investigate the ways in which automated feedback can be used as a support tool in a class setting. This study investigated the potential to integrate Grammarly into writing instruction to support teacher feedback. To this end, we examined the feedback provided by a teacher and by Grammarly through a written feedback analysis of language- and content-related issues and the impact of feedback from three sources (teacher, Grammarly, and combined feedback) on students' revisions. We further scrutinized the general impact of feedback on students' writing performance over 13 weeks. We probed students' attitudes toward the usefulness of each of the aforementioned feedback modes.

Efficacy of Teacher Feedback in L2 Writing

In L2 writing, writing scholars, researchers, and teachers have emphasized the importance of teacher feedback for developing students' writing (Tang & Liu, 2018). Providing such feedback, ranging from error correction to commentary feedback regarding rhetorical and content aspects of writing (Goldstein, 2004), is part of daily teaching practice (Lee, 2008, 2009). In the dichotomy between feedback on form and content, written feedback can be classified into corrective and non-corrective feedback (Luo & Liu, 2017): corrective feedback (CF) promotes learning the target language by providing negative evidence and non-corrective feedback scaffolds English writing in aspects of content, organization, linguistic performance, and format. The focus of teacher feedback has been debated over the past 30 years, which have seen the proposal of Ashwell (2000) and Fathman and Whalley's

(1990) recommendations that there should be a balance between feedback on form and meaning when providing feedback on students' writing.

Many studies on teacher feedback have been concerned with the relative effectiveness of different strategies for written CF. Much work has examined whether and to what extent CF can help improve L2 learners' accuracy in revised and new pieces of writing (e.g., Karim & Nassaji, 2018; Suzuki et al., 2019) and confirmed the positive effects of written feedback on writing accuracy. However, investigations of the usefulness of non-corrective feedback have so far been limited (Ferris, 1997; Ferris et al., 1997). One of the earliest studies on the influence of teacher commentary on student revision, conducted by Ferris (1997), indicated that a significant proportion of comments led to substantive student revision and found that particular types and forms of commentary tended to be more helpful than others.

Previous studies have measured the impact of teacher feedback on students' revision by observing either students' revision operations (Ferris, 2006; Han & Hyland, 2015) or revision accuracy developments (Karim & Nassaji, 2018). Ferris (2006) classified students' revision operations into three categories: error corrected, incorrect change, and no change, while others (e.g., Karim & Nassaji, 2018; Van Beuningen et al., 2012) calculated improvement in accuracy in students' revised texts with an error ratio. The long-term effectiveness of written feedback has been established by several studies (e.g., Karim & Nassaji, 2018; Rummel & Bitchener, 2015). Despite these differences in the tools used, most studies reported a positive influence of feedback on students' revisions and new texts.

Although a significant positive impact was found for teacher feedback on students' writing, providing feedback requires considerable time and effort (Ferris, 2007; Zhang, 2017). Time constraints, large class size, and teachers' workload pose major challenges that prevent them from giving adequate feedback. Consequently, teachers tend to offer feedback primarily on language-related errors rather than content-related issues in students' writing (Lee, 2009). Thus, to ease teacher feedback burden and to enhance the efficacy of teacher feedback, automated feedback may come to be used.

Affordances and Limitations of Automated Feedback in L2 Writing

As educational technologies and computer-mediated language learning have advanced during the twenty-first century, the integration of computer-generated automated feedback in writing instruction has increased in popularity due to its consistency, ease of scoring, instant feedback,

and multiple drafting opportunities (Stevenson & Phakiti, 2014).

Study of the effects of automated feedback on students' writing has increased in recent years, and its findings indicate a positive influence on the quality of texts (Li et al., 2015; Stevenson & Phakiti, 2014). Li et al. (2015) looked at how Criterion (https://www.ets.org/criterion/) impacted writing performance and found that it led to improved accuracy from first to final drafts. Nonetheless, its limitations include an emphasis on the surface features of writing, such as grammatical correctness (Hyland & Hyland, 2006), failing to interpret meaning, infer communicative intent, or evaluate the quality of argumentation, and the one-size-fits-all nature of the automated feedback (Ranalli, 2018). Despite these pitfalls, automated feedback lowers teachers' feedback burden, allowing them to be more selective in feedback they provide (Grimes & Warschauer, 2010).

Noting the supplementary role that instructors and automated systems can play, Stevenson and Phakiti (2014) called for more research on how automated feedback can be integrated into the classroom to support writing instruction. Recent studies have compared the characteristics and impact of teacher and automated feedback (Dikli & Bleyle, 2014; Qassemzadeh & Soleimani, 2016). Dikli and Bleyle (2014) investigated the use of Criterion in a college course of English as a second language writing class and compared feedback from instructor and Criterion across categories of grammar, usage, and mechanics. They found large discrepancies: the instructor provided more and better-quality feedback. Others focused on instructional applications of automated feedback (e.g., Cavaleri & Dianati, 2016; O'Neill & Russell, 2019a, 2019b; Ventayen & Orlanda-Ventayen, 2018). A study by O'Neill and Russell (2019b) found that the Grammarly group responded more positively and was better satisfied with the grammar advice than the non-Grammarly group. Another study, by Qassemzadeh and Soleimani (2016) found that both teacher and Grammarly feedback positively influenced students' study of passive structures. Within this framework, new research is needed to investigate the applicability of automated feedback in writing instruction, which is the impetus for our study.

The Present Study

Context of Myanmar

As a result of the country's political and educational situations, research in ELT in Myanmar, especially classroom-based research, is sparse (Tin, 2014). Given the scarcity of publications in the periphery (including Myanmar), this study took the form of a naturalistic classroom-based inquiry in a general English class at a major university in Myanmar. The aim of the course is to improve students' English language skills. While developing students' English writing ability is one of the foci, teachers do not have sufficient time to provide adequate feedback on students' writing due to their heavy workloads and large classes of mixed-ability students.

Research Questions

This study was guided by four research questions that are mentioned as follows:

1. What is the focus of teacher and Grammarly feedback in terms language- and content-related categories?
2. To what extent do the students make use of the feedback under three conditions (i.e., teacher, Grammarly, and combined) in their revisions?
3. To what extent does the provision of feedback lead to improvement in writing performance as assessed on a pre- and post-test over a 13-week semester?
4. What are the students' views of the usefulness of feedback from different sources in their EFL course?

Methods

Participants

The sample was an intact class of 30 first-year English majors. The students were placed in the course based on their English scores in the national matriculation examination before admission to the university. Their results were assumed to represent their level of English proficiency at the time of the experiment. Though their exam scores placed them at intermediate (B1) proficiency level, their English writing proficiency varied in terms of mastery of English grammar, familiarity with structures and vocabulary used in writing tasks, and in the formal EFL instruction that they had received. All of them were native Burmese speakers; 11 were male and 19 were female, and all were of typical university age, 17 and 18 years old, and participated on a voluntary basis. They were informed of their right to withdraw from the research at any time during data collection. Three students failed to complete one of the writing tasks, and their data were excluded. The class teacher had an MA degree in Teaching English to the Speakers of Other Languages and over nine years of experience in teaching English at higher education institutions in Myanmar.

Materials

Three instruments were used for data collection: writing tasks, an assessment scale to assess improvement in students' writing performance, and self-assessment questionnaires.

Writing Tasks

Six writing tasks were developed (including a pre- and post-test) on topics familiar to the students. The tasks were ecologically valid, as they were retrieved from the prescribed curriculum. The genres included both argumentative and narrative essays, as these two genres prevail in the syllabus. Four guiding prompts, similar to that in Fig. 1, were provided in the writing tasks, and these were similarly structured to minimize possible linguistic differences.

Writing Assessment

The study adapted a B1 analytical rating scale (Euroexam International, 2019) to assess the students' English writing improvement. Euroexam International offers language tests in general, business, and academic English and German at levels A1 through C1. The writing assessment scale features four criteria: task achievement, coherence and cohesion, grammatical range and accuracy, and lexical range and accuracy. A description of the assessment criteria, together with definitions, is presented in "Appendix Table 3". All scoring of written texts (pre- and post-test) was done by the two authors independently, and the mean scores were calculated. The inter-rater reliability coefficients (Pearson r) between the two raters were 0.92 for the pre-test and 0.94 for the post-test on the assessment scale.

Questionnaire

A self-assessment questionnaire was developed to probe the students' emic perspectives of the effectiveness of feedback from three sources. Three closed items were presented to elicit information on the usefulness of the feedback, and five open-ended questions asking students to comment on the usefulness of the feedback.

Procedure

Data were collected over a period of 13 weeks from August to October 2020: the students completed six writing tasks, including a pre- and post-test (Fig. 2). In the first week, the research project was introduced, and then the pre-test was administered in the second week. The course was operated on a weekly basis: participants were given a writing task and received feedback from teacher, Grammarly, or both sources the week after the completion of the initial writing task. There were four treatment sessions in the whole program, and the students revised their texts in response to the feedback and sent the revised texts to the teacher via email in the same week. The process continued until Week 10, when the revised version of the fourth writing task was complete. In Week 13, students completed the post-test and the self-assessment questionnaire.

The provision of feedback was carried out by the class teacher, using the "Track Changes" functionality of

Write the essay according to its instructions. Your writing will be evaluated according to the following four criteria:
- the degree to which you have achieved the task;
- the organization and coherence of paragraphs;
- the range and accuracy of grammatical structures;
- the range and accuracy of vocabulary.

For the following essay, four prompts are given. Write about the four points in separate paragraphs. Write about 300-400 altogether.

Extroverts

Write an essay on your ideas about someone who is extroverted. In the text,
- characterize the person;
- describe the context where you met;
- tell a story about him/her and how he/she behaved with you and other people;
- discuss how his/her personality influenced his/her life.

Fig. 1 Sample writing task

Timeline	Experimental group
Week 1	Introducing the research project
Week 2	Pretest
Week 3	Essay 1
Week 4	Teacher feedback + Revision
Week 5	Essay 2
Week 6	*Grammarly* feedback + Revision
Week 7	Essay 3
Week 8	A combination of teacher feedback & *Grammarly* feedback + Revision
Week 9	Essay 4
Week 10	A combination of teacher feedback & *Grammarly* feedback + Revision
Week 13	Posttest
	Self-assessment questionnaire

Fig. 2 Data collection timeline

Microsoft Word, Grammarly, or a combination of both of these means. To keep the feedback process as natural as possible, the teacher was asked not to limit his feedback to language- or content-related issues. For automated feedback, a free version of Grammarly (https://www.grammarly.com/grammar-check) was utilized and the students uploaded their essays on the webpage, receiving instant feedback.

Data Analysis

Guided by Lee's (2009) work, a written feedback analysis was performed to investigate the focus of the teacher and Grammarly feedback. This involved error identification, categorization, and counting of feedback points: "an error corrected/underlined, or a written comment that constitutes a meaningful unit" (p.14). Feedback points marked on the students' first drafts were initially classified into language- and content-related issues and coded for analysis. Regarding language-related issues, linguistic errors in the students' drafts were identified and categorized based on Ferris's (2006) taxonomy, with adaptations. For content-related issues, in-text and end-of-text comments were classified into four categories: giving information, asking for information, praises, and suggestions according to the aim or intent of the comment suggested by Ferris et al. (1997). It should be noted that Grammarly feedback primarily relates to language-related errors, which is not the case in teacher feedback. Feedback points marked by the teacher and by Grammarly were cross-linked to students' revisions, and changes were analyzed based on their revision operations. This study partly followed the revision analysis categories of Ferris (2006) and Han and Hyland (2015) to classify revision patterns into three categories: correct, incorrect, and no revision (see Supplementary Data).

To examine the impact of feedback provision on students' writing performance, we calculated mean scores and standard deviations at the beginning and at the end of the course. Because the sample size was small, and the variables were not normally distributed, a bootstrap method was used to analyze the dataset. T-tests were administered using a bootstrap method in SPSS 22 (Corp, 2013) to estimate the difference between pre-and post-test performance. The self-assessment questionnaires included both quantitative and qualitative data. The frequencies of responses were calculated, and the students' perceived areas of improvements were reported. For open-ended questions, a qualitative analysis was conducted to better understand their perspectives on how useful they found the feedback. Their responses were summarized with the use of emerging common themes.

Results and Discussions

Focus of Teacher and Grammarly Feedback

Figure 3 summarizes the focus of teacher feedback in comparison with Grammarly feedback and the percentage of each feedback category marked on the students' first drafts. In general, we found that the teacher focused on a broad coverage of writing issues, at the word, sentence, and text levels, while Grammarly indicated language errors: article/determiner, preposition, and miscellaneous errors including conciseness and wordiness issues.

The results of feedback analysis showed that the teacher provided 410 feedback points in 27 essays, targeting language errors (68.8%) and higher-level writing issues (31.2%). This sheds light on labor-intensive nature of teacher feedback. A more detailed analysis showed that teacher error feedback mainly concerns conjunction (10%),

Fig. 3 Feedback categories of teacher and Grammarly feedback
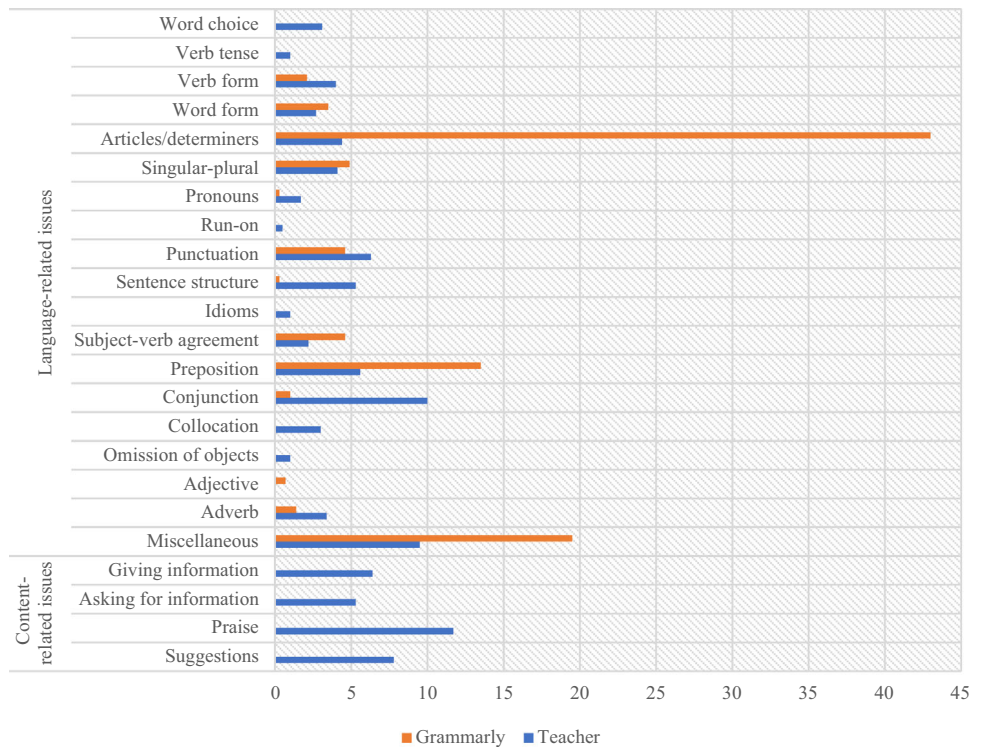
miscellaneous (9.5%), punctuation (6.3%), and preposition errors (5.6%). In the teacher's commentary on content, praise got the highest percentage (11.7%), followed by suggestion (7.8%), giving information (6.4%), and asking for information (5.3%). Our finding that praise accounted for only 11.7 per cent of the total written feedback contradicted that of Hyland and Hyland's (2001), but supported that of Lee's (2009). This might be due to differences in teachers' feedback beliefs about the role of praise in softening criticism when providing feedback on students' writing.

Grammarly predominantly provided feedback on errors of grammar, usage, mechanics, style, and conciseness. It detected 281 errors in 27 essays: the most predominant errors were article/determiner (43%), miscellaneous (19.5%), and preposition (13.5%) errors. Other less frequently indicated errors included with conjunctions (1%), sentence structure (0.3%), and pronoun use (0.3%) (Fig. 3).

All in all, it appears that Grammarly can be used as a learning tool to facilitate teacher feedback. This relates to the focus of each feedback type: the teacher's feedback covered both language and content issues, whereas Grammarly provided feedback on language-related errors. This finding may seem predictable, as Grammarly is understood to be a grammar-checking tool, this emphasis is to its advantage. In particular, its detection of article and prepositions errors was higher than those of teacher feedback. Thus, utilizing Grammarly effectively for offering

feedback on these errors would possibly save time and effort on part of teachers.

It is also fair to say that the use of Grammarly along with teacher feedback might also enhance the efficacy of teacher feedback. As in previous studies (e.g., Lee, 2009; Mao & Crosthwaite, 2019), the teacher feedback primarily attended to language errors (68.8%). Given time constraints and the large classes of mixed-ability students, providing effective and individualized feedback for students' writing is far beyond the capabilities of teachers. In this regard, using automated feedback as an assistance tool might become an outlet for coping with surface errors, lightening the teacher feedback burden: freeing teachers to focus on higher-order writing concerns such as content and discourse (Ranalli, 2018).

Impact of Teacher, Grammarly, and Combined Feedback: Successful Revision

When examining the influence of feedback on students' revision, this study considered how the feedback was acted upon to facilitate comparability across feedback from three sources. A general pattern of students' revision operations led to successful revision, regardless of the source of feedback (Fig. 4), indicating their acceptance of feedback. The finding that teacher error feedback leads to effective revision is in agreement with the findings of Ferris (2006) and Yang et al. (2006). Moreover, the lowest percentage of

Fig. 4 Student revision operations

Correct revision ■ Incorrect revision ■ No revision
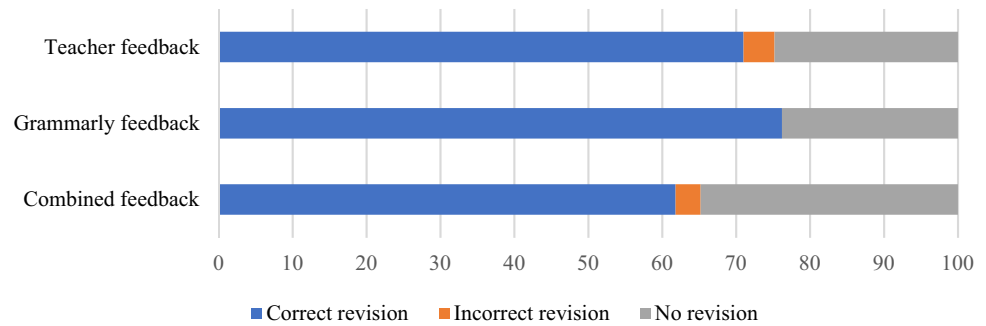
Excerpt 1

Student original: We need to know at least three languages in order to apply scholarships at foreign universities.

Grammarly comment: We need to know at least three languages in order to apply [apply for] scholarships at foreign universities. [The phrase following the intransitive verb **apply** seems to be missing a preposition. Consider adding one.]

Student revision: We need to know at least three languages in order to apply for scholarships at foreign universities. [correct revision]

Fig. 5 An example of Grammarly feedback and student's revision

unrevised errors reflects their beliefs and value regarding the importance of feedback in improving their writing performance. The results were interesting for Grammarly feedback which received the highest rate of correct revision (76.2%). The reason for this might be that Grammarly usually includes a concrete suggestion for revision that students can easily act upon. One example of this is shown in Fig. 5.

Further points of discussion concern how students responded to the combined feedback. It might be assumed that combined feedback resulted in more feedback points than the other conditions. However, the opposite was true: fewer feedback points were provided, and a lower ratio of correct revision was found than for the teacher and Grammarly, which had the highest ratio of no revision. A possible explanation of lower feedback points might relate to students' increased awareness of teacher and Grammarly feedback in previous essays or teacher's reliance on Grammarly feedback, instinctively assuming that it would handle grammar errors.

Although the students successfully revised their errors, it is worth exploring how well they revised individual error categories (Table 1). As the overall percentage of successful revisions was high, it was not surprising to see that the percentage of successful revisions in most error categories was also fairly high, regardless of the conditions. However, a closer examination of how students utilized feedback revealed stimulating new results. In connection with teacher feedback, while feedback on most error categories (e.g., conjunction, article/determiner, singular-plural, adverb, and word choice) was associated with correct revision, some feedback on idioms, pronoun, and sentence structure was left unattended. For example, 40.9% of errors in sentence structure led to no revision. This could be explained by the low number of error identifications in these categories and partial understanding of the instructions (Han, 2019). As Goldstein (2004) noted, reasons for unsuccessful/no revisions included: unwillingness to critically examine one's point of view, feeling that the teacher's feedback is incorrect, lack of necessary knowledge to revise, lack of time and motivation, and many others.

Despite the overall successful revision when acting upon Grammarly (76.2%) and combined feedback (61.8%), the results indicated that the students largely ignored feedback on miscellaneous errors. This finding is probably due to students finding the feedback in this category unhelpful or unnecessary in revision. Figure 6 demonstrates a typical example. This underlines how students selectively accept the feedback, filtering suggestions that are incorrect or unnecessary (Cavaleri & Dianati, 2016).

The question of whether Grammarly could be integrated into writing instruction can be answered by how the students responded to feedback in their revisions. The comparison of outcomes in the three conditions provided support for the potential to use Grammarly, along with teacher feedback. The reason for this is associated with the high percentage of successful revision in cases of feedback regarding the singular-plural (92.9%), subject-verb

Table 1 Comparison of students' revision operations by error type

	Teacher (no./%)			Grammarly (no./%)			Combined (no./%)		
	CR	IR	NR	CR	IR	NR	CR	IR	NR
Word choice	11/84.6	0	2/15.4	–	–	–	6/85.7	0	1/14.3
Verb tense	2/50.0	1/25.0	1/25.0	–	–	–	7/70	2/20	1/10.0
Verb form	11/68.8	2/12.5	3/18.8	6/100	0	0	3/50.0	0	3/50.0
Word form	9/81.8	0	2/18.2	9/90.0	0	1/10	2/50	0	2/50.0
Articles/determiners	15/83.3	1/5.6	2/11.1	102/84.3	0	19/15.7	21/84	0	4/16.0
Singular-plural	13/76.5	1/5.9	3/17.6	13/92.9	0	1/7.1	3/37.5	0	5/62.5
Pronouns	4/57.1	0	3/42.9	1/100.0	0	0	4/66.7	0	2/33.3
Run-on	1/50.0	1/50.0	0	–	–	–	–	–	–
Punctuation	15/57.7	1/3.8	10/38.5	11/84.6	0	2/15.4	11/68.8	0	5/31.2
Sentence structure	11/50.0	2/9.1	9/40.9	1/100.0	0	0	8/61.5	3/23.1	2/15.4
Idioms	1/25.0	0	3/75.0	–	–	–	1/100	0	0
Subject-verb agreement	7/77.8	0	2/22.2	12/92.3	0	1/7.7	10/90.9	0	1/9.1
Preposition	16/69.6	1/4.3	6/26.1	32/84.2	0	6/15.8	16/69.6	0	7/30.4
Conjunction	33/80.5	1/2.4	7/17.1	2/75.0	0	1/25.0	13/54.2	2/8.3	9/37.5
Collocation	9/75.0	0	3/25.0	–	–	–	5/55.6	0	4/44.4
Omission of objects	4/100.0	0	0	–	–	–	3/100	0	0
Adjective	0	0	0	2/100	0	0	2/75	0	1/25.0
Adverb	10/71.5	1/7.1	3/21.4	4/100	0	0	5/100	0	0
Miscellaneous	28/71.8	0	11/28.2	19/34.5	0	36/65.5	6/20.0	0	24/80.0
Total	200/71.0	12/4.2	70/24.8	214/76.2	0	67/23.8	126/61.8	7/3.4	71/34.8

Percentages represent frequencies of revision categories within each error category. For instance, 84.6% of the word choice errors had a correct revision rating

CR correct revision, *IR* incorrect revision, *NR* no revision

agreement (92.3%), word form (90%), punctuation (84.6%), article/determiner (84.3%), and preposition (84.2%) following Grammarly feedback. Thus, it seems reasonable to say that using Grammarly to handle errors in these categories could be effective and allow time for teachers to focus on other higher-level writing issues. Specifically, although the teacher made 22 feedback points regarding sentence structure, 40.9 per cent of them were left unattended. This partly mirrors the indirectness or vagueness of teacher feedback that may be difficult for students to respond to (Tian & Zhou, 2020). What should be stressed is that teachers might be able to focus on these types of errors if they can efficiently make use of Grammarly to deal with surface-level ones.

Effect of Written Feedback on Students' Writing Performance

After receiving feedback over a semester, the students made improvement in their writing performance, as is shown in the significant increase in their post-test scores across four assessment criteria. As presented in Table 2, there was substantial improvement in *task achievement* and *coherence and cohesion* in the post-test scores. Similarly, in connection with *grammatical range and accuracy* and *lexical range and accuracy*, the students showed notable improvement from the pre- to the post-test. The effect sizes for all significant comparisons of learners' writing performance were medium to large. The positive impact of feedback provision on new writing tasks was in line with that found in previous studies (e.g., Karim & Nassaji, 2018; Rummel & Bitchener, 2015).


Excerpt 2

Student original: If we travel abroad, lacking proficiency in foreign language is really bothersome.

Grammarly comment: If we travel abroad, lacking proficiency in foreign language is <u>really</u> bothersome. [It appears that **really** may be unnecessary in this sentence. Consider removing it.]

Student revision: If we travel abroad, lacking proficiency in foreign language is really bothersome. [no revision]

Fig. 6 An example of Grammarly feedback on a miscellaneous error and student's revision outcome

Table 2 Comparison between pre-and post-test regarding the students' writing performance

Assessment criteria	Pre-test		Post-test		$t(26)$	p	Cohen's d
	Mean	SD	Mean	SD			
Task achievement	2.25	543	2.65	0.551	3.82	0.003	0.71
Coherence & cohesion	2.25	610	2.61	0.560	3.90	0.002	0.75
Grammatical range & accuracy	2.20	559	2.52	0.628	2.88	0.017	0.55
Lexical range & accuracy	2.26	685	2.69	0.483	3.55	0.003	0.68
Overall writing performance	8.98	2.091	10.46	1.965	3.14	0.006	0.61

Fig. 7 Students' perceptions of the usefulness of the teacher, Grammarly, and combined feedback
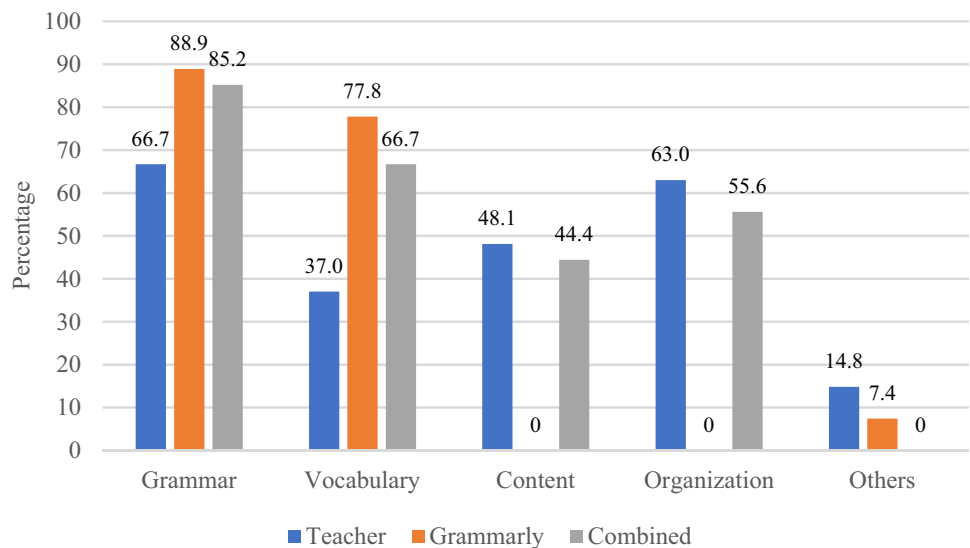
Students' Views on the Usefulness of Teacher, Grammarly, and Combined Feedback

The results from the self-assessment questionnaires showed that most students perceived the feedback from both the teacher and Grammarly to be effective and useful for improving their writing (Fig. 7). Although most responded that Grammarly feedback helped them improve their grammar (88.9%) and vocabulary (77.8%), none reported improvements in content or organization. The teacher feedback was considered more valuable, as it facilitated improvement in different aspects of writing, and the combined feedback did this as well. Despite the students' positive impressions for both the teacher's and Grammarly feedback, their responses regarding specific areas of improvement for the combined feedback were considerably higher across different aspects of writing. This finding underlines the great potential for integrating Grammarly feedback into writing instruction, supplementing teacher feedback, as reported in previous studies by O'Neill and Russell (2019b), Ventayen and Orlanda-Ventayen (2018), and Ranalli (2021).

In the second part of the questionnaire, the students reported why they liked the feedback they received. Almost all students acknowledged the value and effectiveness of the teacher feedback. Their comments showed three emerging themes relating to the nature of the feedback, how it enhances their motivation, and their positive perception of teacher feedback. Almost all students stressed the value of the teacher's feedback, saying that his comments "guide me when my writing goes out of context" (Student 21), "show me both strengths and weaknesses of my writing" (Student 2), and are "short and clear" (Student 27).

Most comments regarding the usefulness of Grammarly feedback concerned its efficiency: "It is easy to use and available for free" (Student 3), and "I could use Grammarly at any time" (Student 9). However, a few students were dissatisfied with it: "To be honest, I don't feel very satisfied with it" (Student 15) and "Honestly, I didn't find Grammarly feedback useful" (Student 19). Further responses revealed how the combined feedback helped them revise their essays: "Teacher's feedback tells me my mistakes exactly and Grammarly fixes those for me" (Student 20), and "It's a perfect combination" (Student 25).

Implications

Our findings have pedagogical implications for the integration of Grammarly into teaching L2 writing. Considering the emphasis of Grammarly feedback on language-related errors as an advantage, writing teachers could use it as a supportive tool in their classes on a regular basis or encourage students to use it independently. In this way, teacher feedback burden could be reduced and challenges regarding time constraints and inadequacy of attention paid to individuals in large classes could be addressed to a certain extent. In particular, based on Grammarly's effective feedback on article/determiner and preposition errors and students' successful revisions of these errors reflect their acceptance of Grammarly as a provider of feedback in their EFL courses. Thus, teachers can exploit the affordances of Grammarly to maximize the efficacy of their feedback. However, teachers should be aware of the limitations of automated feedback and be sure to inform students of these limitations.

Additionally, writing teachers should be more selective and straightforward in providing feedback to improve students' writing performance and motivation. In our study, the students were not able to revise errors relating to sentence structure, leaving most of them unrevised. Moreover, because of the overlaps in teacher and Grammarly feedback in some language-related errors, teachers can identify the areas on which Grammarly can provide feedback effectively, allowing them to focus on higher-level writing skills including content development and elaboration, organization, and rhetoric.

Conclusion

This classroom-based study was conducted to examine the integrated use of Grammarly in a large class to support teacher feedback. The results showed the pedagogical potential of Grammarly for facilitating teacher feedback due to its effective feedback regarding surface-level errors and students' general acceptance of automated feedback. Moreover, it seems that students' successful integration of feedback in their revisions and increased performance scores on the post-test offer evidence that they successfully made use of feedback and that the provision of feedback led to an improvement in their writing performance. In addition, their positive attitude toward the usefulness of feedback provides further insights into how much they valued the feedback they received from the teacher and Grammarly.

Some limitations should be addressed, as we conducted the study in only one course at a university. Future research should involve more courses, teachers, and students at varying proficiency levels. The inquiry failed to include a control group because we considered it unethical to withhold feedback from students that they would typically receive in their course. Therefore, no comparison was made between the feedback group and a control group. However, we managed to examine how students applied feedback from three sources in their revisions and to track progress during the course. This investigation may offer insights into areas beyond how students use feedback in their revision and how feedback helps them develop their writing performance. We hope that the findings of this study indicate how Grammarly can be used as an effective feedback tool to help relieve teachers of a part of the burdensome task of responding to surface-level errors in students' writing.

Appendix

See Table 3

Table 3 Operational rating scales for writing tasks at B1 level adopted from Euroexam International ([2019])

B1	Task achievement	Coherence & cohesion	Grammatical range & accuracy	Lexical range & accuracy
3	Task achieved at a high level			
	Rubrics: Followed completely in all 4 guiding points	Information: Well organized into a coherent text	Range: Good range of grammatical structures	Range: Good range of lexis to complete the task
	Content: Enough and relevant discussion and details are included on all 4 guiding points	Cohesive devices: Overall good use of cohesive devices	Accuracy: Grammatical structures used accurately with no or very few basic errors	Accuracy: Lexis used appropriately with no or little misuse
	*One mark will be penalized if some irrelevant discussion and details included			
2	Task achieved with minor gaps			
	Rubrics: Followed in 2 or 3 guiding points	Information: Part of the text is well organized	Range: Sufficient range of grammatical structures	Range: Sufficient range of lexis to complete the task
	Content: Enough and relevant discussion and details are included on 2 or 3 guiding points; Little or not relevant information is discussed on 1 or 2 points	Cohesive devices: Mostly good use of cohesive devices with minor gaps	Accuracy: Grammatical structures used mostly accurately with some errors that do not significantly impede meaning	Accuracy: Lexis used mostly appropriately with minor gaps
	*One mark will be penalized if some irrelevant discussion and details included			
1	Task achieved with major gaps			
	Rubrics: Followed in 1 or 2 guiding points	Information: Text is hard to follow	Range: Limited range of grammatical structures	Range: Limited range of lexis to complete the task
	Content: Enough and relevant details are included on 1 or 2 guiding points; Little or not relevant information is included on 2 or 3 points	Cohesive devices: Major gaps in use of cohesive devices	Accuracy: Grammatical structures used inaccurately interfering with meaning	Accuracy: Lexis often used inappropriately causing misunderstanding
	*One mark will be penalized if some irrelevant discussion and details included			
0	Task unachieved			
	Task unattempted/partially attempted	–	–	–
	Not enough language to make an assessment			

What each criterion is supposed to assess are as follows:

1. Task achievement concerns how well a candidate has fulfilled the task, addressing the guided prompt with relevant details while aiming at the general target reader, in other words, if he has done what he was supposed to do.
2. Coherence and cohesion focus on how well organized a text is, following a coherent structure to maintain the organization of the whole text while making good use of cohesive devices.
3. Grammatical range and accuracy focus on the accuracy of grammatical structures that a candidate uses, demonstrating a variety of grammatical structures available to him.
4. Lexical range and accuracy focus on the accuracy and lexical items that a candidate uses, displaying the appropriate choice and variety of words with an adequate range of lexis to complete the task.

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

Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning, 10*(1), A223–A236.

Corp, I. B. M. (2013). IBM SPSS statistics for windows, version 22.0. Armonk, NY

Dikli, S., & Bleyle, 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

Euroexam International. (2019). *Euroexam detailed specifications*. London: Euroexam International.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge University Press.

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339.

Ferris, D. (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*(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press.

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing, 16*, 165–193. https://doi.org/10.1016/j.jslw.2007.07.003

Ferris, D. R., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing, 6*(2), 155–182. https://doi.org/10.1016/S1060-3743(97)90032-1

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing, 13*, 63–80. https://doi.org/10.1016/j.jslw.2004.04.006

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1–44.

Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System, 80*, 288–303. https://doi.org/10.1016/j.system.2018.12.009

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31–44.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*, 185–212.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students ' writing. *Language Teaching Research*. https://doi.org/10.1177/1362168818802469

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*(March), 100450. https://doi.org/10.1016/j.asw.2020.100450

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69–85. https://doi.org/10.1016/j.jslw.2007.10.001

Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal, 63*(1), 13–22. https://doi.org/10.1093/elt/ccn010

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics, 07*(04), 197–215. https://doi.org/10.4236/ojml.2017.74015

Mao, S. S., & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. *Journal of Second Language Writing, 45*(November 2018), 46–60. https://doi.org/10.1016/j.jslw.2019.05.004

O'Neill, R., & Russell, A. M. T. (2019a). Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker. *Journal of Academic Language & Learning, 13*(1), A88–A107.

O'Neill, R., & Russell, A. M. T. (2019b). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1), 42–56. https://doi.org/10.14742/ajet.3795

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies, 6*(9), 1884. https://doi.org/10.17507/tpls.0609.23

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, 100816. https://doi.org/10.1016/j.jslw.2021.100816

Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact of Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics, 1*, 66–84. https://doi.org/10.1075/aral.38.1.04rum

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System, 81*, 135–145. https://doi.org/10.1016/j.system.2018.12.017

Tang, C., & Liu, Y. T. (2018). Effects of indirect coded corrective feedback with and without short affective teacher comments on

L2 writing performance, learner uptake and motivation. *Assessing Writing, 35*, 26–40. https://doi.org/10.1016/j.asw.2017.12.002

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

Tin, T. B. (2014). Learning English in the periphery: A view from Myanmar (Burma). *Language Teaching Research, 18*(1), 95–117. https://doi.org/10.1177/1362168813505378

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

Ventayen, R. J. M., & Orlanda-Ventayen, C. C. (2018). Graduate students' perspective on the usability of Grammarly® in one ASEAN state university. *Asian ESP Journal, 14*(7), 9–30.

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

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004

Zhang, X. (2017). Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research, 21*(2), 217–240. https://doi.org/10.1177/1362168815623291