

Pedagogic Exploration Into Adapting Automated Writing Evaluation and Peer Review Integrated Feedback Into Large-Sized University Writing Classes

SAGE Open
October-December 2023: 1–18
© The Author(s) 2023
DOI: 10.1177/21582440231209087
journals.sagepub.com/home/sgo
 Sage

Wei-Yan Li¹ , Kevin Kau¹ , and Yi-Jiun Shiung¹

Abstract

Automated Writing Evaluation (AWE) software has been viewed as a promising tool for assisting writing. This study integrated AWE and its combination with peer review discussion into writing practice in a large college writing class in the Asian context. Adopting a mixed-method approach, this study employed a quantitative questionnaire to investigate how students perceived the integration at different writing stages and a qualitative interview to examine what prompted their quantitative decisions. The results show that the integration of AWE and peer review feedback significantly reinforced the perceived usefulness of AWE at the revision stage, and students with different proficiency levels had notably different attitudes toward AWE. Their writing anxieties intensified as they became increasingly aware of their own weaknesses during the writing process. Therefore, student writing challenges involve not only language proficiency and writing skills but also psychological variables.

Keywords

automated writing evaluation, academic writing, feedback, peer review, pedagogical techniques

Introduction

Academic writing is one major teaching challenge faced by language instructors (Almubark, 2016). Academic writing requires more than linguistic basics; it encompasses considerations in style, flow and organization and demonstrations of skills such as critical thinking, persuasive power, and informational synthesis. In addition, teaching academic writing involves not only delivering content but also guiding students through an entire writing process from brainstorming outlines to reaching final drafts. In whichever step in this process, providing regular and ample feedback is instrumental in scaffolding writing instruction (Bitchener & Ferris, 2012; Ferris, 2003; Hu, 2007; Lai & Tu, 2020). Feedback ranges from higher-order concerns, such as writing logic and argument development, to lower-order concerns, such as grammar, spelling, punctuation, or conciseness. Owing to this wide range, feedback provision is effort-intensive and time-consuming. To exacerbate this already daunting task, there has been an upsizing trend in writing classes; with increased class size come heavier workloads for writing instructors. In one study, a single writing instructor

in a university English class in Hong Kong was expected to mark 1,128 written papers of 250 students within a 13-week semester (Tso & Ho, 2016). Therefore, their feedback quality would deteriorate were they not to receive pedagogical support.

To extend such support, Automated Writing Evaluation (AWE) has received considerable attention. AWE software provides computer-generated feedback on the quality of a submitted text and generates automated scores via artificial intelligence, natural language processing, and latent semantic analysis (Dikli, 2006; Philips, 2007; Shermis & Burstein, 2003). Moreover, AWE software provides written feedback in the form of comments and/or corrections. Hence, all these features considered, the successful integration of AWE software may alleviate the feedback burden of writing instructors,

¹National Taiwan University, Taipei

Corresponding Author:

Shiung, Yi-Jiun, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617.
Email: yjshiung@ntu.edu.tw



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

redirecting their effort into teaching and maintaining teacher-student interaction in large class settings.

Research has been conducted to assess the potential of AWE application as a supplementary writing tool (e.g., Burstein et al., 2004; Chodorow et al., 2010) and one recent meta-analysis examines how AWE application improves L2 writing skills (Mohsen, 2022). Despite their promising results, it remains under-explored how an AWE system can be effectively incorporated into the writing classroom in tertiary education in the Asian context and in particular how AWE application helps learners with different proficiency levels address both higher-order and lower-order concerns.

Literature Review

Automated Writing Evaluation

AWE systems have been employed since the 1960s in the US to help score student essays (Page, 2003). In addition to this original assessment purpose, AWE is currently being used in high-stakes tests, such as the Test of English as a Foreign Language and the Graduate Management Admissions Test (Stevenson, 2016). AWE software, such as *Criterion*, *Grammarly*, and *ProWritingAid*, uses algorithms to extract features (e.g., a count of grammatical errors, word length, or the number of words in a discourse element) from submitted essays and then determines via linear regression optimal weights of those features to predict a criterion. Different AWE models have been created to feature weights assigned to mimic human scoring of identical essays, and the quality of the machine scoring is assessed by how closely it matches human scoring in a cross-validation sample (Shermis & Burstein, 2013).

Grammarly AI system incorporates machine learning with several natural language processing approaches. Its editing affordances can be used at the editing stage. Real-time feedback is made available. Problems and errors are underlined in different colors. In addition to grammatical correctness, *Grammarly* also provides feedback on clarity, engagement, delivery, style, and plagiarism checking (The Grammarly Editor, 2019). Suggested

corrections and explanations are also provided. The overall text score is based on the number and types of suggestions in the document. Writing aspects and comment types on *Grammarly* are shown in Table 1, adopted from Barrot (2022).

AWE Classroom Applications

AWE has been explored at different levels of explicitness and depth regarding its roles in writing practice, feedback provision and classroom instruction. AWE integration into EFL and EAP writing practice has been documented in several studies (Chen & Cheng, 2008; Grimes & Warschauer, 2010). In one case, students were required to keep submitting and revising their papers through AWE software until they reached a threshold score set by the teachers, after which they would submit their papers to their instructors for feedback (Chen & Cheng, 2008). In another, students received feedback from both AWE and teachers simultaneously through the AWE system (Grimes & Warschauer, 2010); teachers would review and then modify AWE feedback. In these studies, AWE feedback was considered useful but required augmentation with human feedback, a point stressed in several studies (Bai & Hu, 2017; Miranty & Widiati, 2021; O'Neill & Russell, 2019a).

In AWE-EFL/EAP integration, assistance in feedback provision is a key component. Among different explorations of AWE as a means of feedback provision, Chen and Cheng (2008) examined AWE integration and perceived usefulness in three different classes. They found that the integration was multifold (e.g., AWE as an instructional tool during drafting and revising or throughout the entire writing process; AWE coupled with both teacher and peer feedback in later phases; AWE coupled only with teacher feedback) and that students adopted a positive attitude toward AWE when it was used in the initial stages and in combination with both peer and teacher feedback.

To scaffold students' use of AWE, others have focused on how AWE can be employed or applied as

Table 1. Writing Aspects and Comment Types on *Grammarly*.

Grammar	fragments or missing commas, run-on sentences, garbled sentences, subject-verb agreement, ill-formed verbs, pronoun errors, possessive errors, wrong or missing words
Usage	wrong articles, missing or extra articles, confused words, wrong forms of words, faulty comparisons, preposition errors, nonstandard word forms, negation errors
Mechanics	spelling, capitalization, missing initial capital letters in sentences, missing question marks, missing final punctuations, missing apostrophes, missing commas, hyphen errors, fused words, compound words, duplicates
Style	repetition of words, inappropriate words or phrases, sentences beginning with coordinating conjunctions, too many short sentences, too many long sentences, passive voice
Organization development	introductions, theses, main ideas, supporting ideas, transitional words and phrases, conclusions
Plagiarism	summaries, paraphrases

part of classroom instruction. In these discussions, teachers tried to compensate for AWE's limitations. In Ghufon and Rosyida's (2018) quasi-experimental study of 40 EFL Indonesian students, *Grammarly* was considered useful in reducing errors in diction, language use, and mechanics. In Choi (2010), AWE was embedded into the process approach. In Steinhart (2001), AWE was employed as a strategy for post-teacher feedback reinforcement. In Grimes (2008), AWE was incorporated into pre-writing, peer discussion, and revision. Even though teachers were reported compensating for limitations inherent in AWE by giving guidance on how to interpret and use AWE feedback (Dikli, 2006) or turning off analytic scoring categories that seemed inaccurate (Grimes & Warschauer, 2010), AWE in these studies has shown great potential as an additional instruction tool.

AWE Potential and Limitations

Potential

AWE assists students in assessing and enhancing writing quality, as evidenced in several studies (Bai & Hu, 2017; Cavaleri & Dianati, 2016; Chen & Cheng, 2008; Darayani et al., 2018; Miranty & Widiati, 2021; O'Neill & Russell, 2019a; Wang et al., 2013; Zhang & Hyland, 2018), as it provides timely, form-focused, and corrective feedback, such as grammatical accuracy, clarity, and mechanics. In addition to this perceived usefulness, AWE has been shown to strengthen writing motivation and facilitate autonomous learning (e.g., Liao, 2015; Miranty & Widiati, 2021; Nova, 2018; Parra G & Calero S, 2019) in that the quantitative and qualitative feedback enables students to "witness" their learning growth, a form of cognizance considered crucial for being an "autonomous learner" in Ghufon and Rosyida (2018, p. 128). Furthermore, AWE encourages self-reflection when students self-revise, and as a consequence of no involvement of others, students free themselves from judgment, hence decreasing writing anxiety (Dodgson et al., 2016).

Limitations

The adoption of AWE into classroom settings is restricted due to uncertainties over accuracy and usefulness. In terms of accuracy, research on AWE typically focuses on the methodological concerns of AWE developers, addressing only a few linguistic "micro-features" (e.g., Han et al., 2006) or preposition errors (e.g., Chodorow et al., 2010). These studies represent a "system-centric (i.e., focusing on the system performance) rather than a user-centric (i.e., focusing on the user's interaction with the system) viewpoint" (Ranalli et al., 2017, p. 10). In system-centric studies, measurement

focuses on *precision* and *recall*. Precision measures how often a system is correct when something is identified as an error; recall measures the proportion of errors that the system has flagged out of the total number of errors present (Leacock et al., 2010). Data in Burstein et al. (2004) and Chodorow et al. (2010) showed that while feedback provided by AWE software was largely correct, there were still unflagged errors. Similarly, user-centric studies have shown that AWE software has relatively high precision but low recall in comparison to instructor feedback. In Dikli and Bleyle (2014), large discrepancies were found between AWE software and instructors; the latter had much higher precision and gave feedback of much higher quality. AWE software was also found to under-identify certain error types commonly observed in L2 writing, such as pronoun and verb-form errors (Ferris, 2011) and show inconsistency in accuracy among different types of errors (Lavolette et al., 2015).

The uncertainty over usefulness is fourfold. First, AWE feedback is treated with ambivalence. In several studies, students had positive views of AWE feedback but preferred the feedback from their instructors (Dikli & Bleyle, 2014; Gao, 2021; Thi & Nikolov, 2022). Second, much AWE feedback goes unused, as indicated in research on revision behavior (e.g., Bai & Hu, 2017; Cavaleri & Dianati, 2016). In a study of *Criterion* involving thousands of students across the US, more than 70% of 33, 171 essays were submitted only once for feedback, demonstrating that most students did not exploit the revision capabilities of the *Criterion* system (Attali, 2004). In Bai and Hu (2017) and Li et al. (2015), many students made no changes possibly due to frequent receiving of inaccurate feedback. This phenomenon is further evident in Cavaleri & Dianati (2016), where a majority of respondents disregarded some feedback even though they acknowledged the usefulness of AWE in providing detailed and helpful suggestions. Third, AWE feedback engenders comprehension difficulties. In Dikli (2006), non-native speakers found it difficult to understand feedback on grammar and usage errors. In Warschauer and Grimes (2008), few students took time to read through AWE generic feedback, and students who did either experienced difficulties understanding it or failed to act upon it. Although students were found in one study to understand and accept AWE feedback, the feedback was provided in graphic form (Wade-Stein & Kintsch, 2004). Forth, AWE feedback is deficient in flexibility. It allows for little interaction and meaning negotiation (Chen & Cheng, 2008) and provides advice much more on lower-order concerns and surface-level errors (Ghufon & Rosyida, 2018; Miranty & Widiati, 2021; Thi & Nikolov, 2022). In other words, higher-order concerns and content-related issues remain unresolved or unflagged. Moreover, AWE generic feedback did not provide meaningful support and instead encouraged students

to use fixed writing patterns, resulting in less creativity (Li et al., 2015).

Given these limitations, much aforementioned AWE literature has suggested the incorporation of teacher language- and content-related feedback (Grimes & Warschauer, 2010). In addition, addressing higher-order concerns is effort-intensive and time-consuming; relying on only one approach may not effectively enhance meaning negotiation and lighten feedback burdens of writing instructors. Peer review may thus be put into play to consolidate AWE feedback.

Peer Review in Writing Practice: What & How

Peer review, an approach where students give one another constructive written or oral feedback in pairs or in groups, has been commonly implemented in writing classrooms (Alnasser, 2018; Anderson et al., 2020; Chaktsiris & Southworth, 2019; Cui et al., 2022; Hentasmaka & Cahyono, 2021; Kusumaningrum et al., 2019; Nusrat et al., 2019; Weissbach & Pflueger, 2018). The successful conduct of peer review requires careful designing (Cui et al., 2022; Hentasmaka & Cahyono, 2021; Kusumaningrum et al., 2019; Weissbach & Pflueger, 2018). To structure peer review sessions, approaches that provide specific guidance to students, such as worksheets with a set of guiding questions (Cui et al., 2022), benchmarks showing groups of writing features (Anderson et al., 2020) or grading rubrics in aid of assessment (Anderson et al., 2020; Jackson et al., 2018; Law & Baer, 2020; Mahmood & Jacobo, 2019; Nagoshi et al., 2019), are often adopted. Moreover, if students are well-trained, the quality of peer review feedback will improve (Jackson et al., 2018).

Potential

Peer review enhances the writing and learning performance of students. Peer feedback, comparable to teacher feedback, leads to writing improvements in higher education settings (Huisman et al., 2020), for structured sessions enable student writers to discern their individual strengths and weaknesses and evaluate their writing performance. Through the peer review process, student writers benefit from constructive criticism (Huisman et al., 2018; Weissbach & Pflueger, 2018), receive support that may reduce their writing anxiety, and develop into autonomous writers (Chaktsiris & Southworth, 2019). In addition, the use of sliding scale rubrics may offer less motivated students an incentive to continue learning tasks and seek help (Mahmood & Jacobo, 2019). With the help of explicit grading rubrics, students generally adopt positive attitudes toward notes anonymously graded by their peers (Nagoshi et al., 2019).

Limitations

Despite the abovementioned strengths, peer review as a sole form of feedback provision exhibits doubtful usefulness. Although Lundstrom and Baker (2009) shared with other studies their finding that feedback-givers gained more in writing performance, other studies found otherwise, as summarized in Yu and Lee (2016). Moreover, students were found to question the quality of peer feedback because of their presuppositions about the competence of their peers (Alnasser, 2018; Tsui & Ng, 2000) and demonstrate negative attitudes toward the review process (Mulder et al., 2014). Students also expressed preferences as to whether peer review should be conducted in groups or in pairs. In Mulder et al. (2014), students demonstrated diminishing expectations of peer review after being exposed to the process in courses taught in different disciplines, year levels and class sizes.

Purposes and Research Questions

The purpose of our study is to increase the compatibility of peer feedback with AWE; in particular, we assess the extent to which student change their attitudes when these two approaches are adopted. The research purpose (RP) is threefold: (1) to integrate AWE software into writing practice to explore its role during pre-writing, writing, and self-editing in post-writing in a large writing class; (2) to consolidate AWE software with peer review during post-writing; and (3) to investigate the perceived usefulness of different AWE integrations at four different writing stages through a mixed-methods approach. To this end, our study answers two research questions (RQs):

RQ1: How do L2 student writers perceive the integration of AWE software during pre-writing, writing, and self-editing in post-writing?

RQ2: How do L2 student writers perceive the combination of AWE software and peer review in writing practice and instruction?

Methodology

Research Design

The combination of methodological approaches is justified by the pragmatic and balanced approach, which aims to “improve communication among researchers from different paradigms as they attempt to advance knowledge” and “offer the best opportunities for answering important research questions” (Johnson & Onwuegbuzie, 2004, p. 16). It is argued that the mixed-methods design allows diverse views and standpoints while enriching a comprehensive understanding of the studied phenomenon (Creswell & Plano Clark, 2011). To address the research

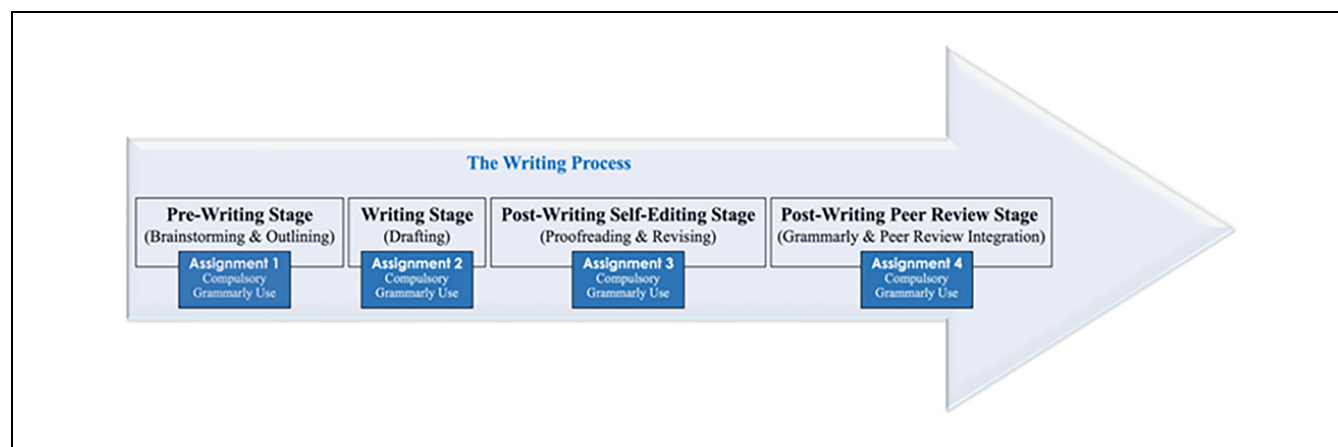


Figure 1. Grammarly use at four different writing stages.

questions posited, this study adopted a mixed-method approach—a quantitative questionnaire instrument and a qualitative interview instrument—to first investigate the perceived usefulness of AWE from the perspective of students and then explore the reasons and evidence explicating their quantitative decisions. This adoption aims to involve a diversity of perspectives, enrich the understanding of the RQs, and facilitate inter-researcher communication (Ågerfalk, 2013).

AWE Integration: Functions & Application of Grammarly

This study employed *Grammarly* as a subject of learning, a means to learning, and a catalyst to foster learning to explore in depth its function as a pedagogical technique and instructional component in the writing classroom. The employment is threefold: (1) selecting *Grammarly* as the studied AWE software and utilizing it as a segment of learning content, (2) requiring students to use *Grammarly* for feedback and reflect upon their revisions for each assignment, and (3) establishing a *Grammarly*-peer review integrated feedback mode (IFM) during post-writing peer review.

An elective academic English writing course with 82 enrolled students at an Asian university was studied. This writing class aimed to familiarize undergraduate students with fundamental writing skills. Topics included sentential structures and relationships, common rhetorical patterns, paragraph organizations, and idea development in argumentation. Students were assessed by their performance in collaborative writing tasks and four individual writing assignments. To investigate how *Grammarly* affected writing behavior and performance, students were free to utilize *Grammarly* throughout this 18-week course during different writing stages, as specified in Figure 1.

Table 2. Description of Participant English Proficiency Level ($n = 82$).

	Upper-intermediate or above*	Intermediate or below
Freshman	13	11
Sophomore	17	6
Junior	9	5
Senior	12	9
	51	31

*Upper-intermediate or above: (1) CEFR B2 or above; (2) TOEIC 750 or above; (3) IELTS 5.5 or above, or (4) TOEFL iBT 72 or above.

To realize the goal of AWE integration as a pedagogical technique, we (1) introduced *Grammarly* and gave drill workshops in Weeks 2 and 3; (2) phased the use of *Grammarly*; (3) introduced peer review and gave drill workshops in Week 13; and (4) introduced IFM and gave drill workshops in Week 14.

Participants

A class of 82 subjects from different academic backgrounds participated in this study. Their school year and English language proficiency are presented in Table 2. Among them, 51 (62%) had upper-intermediate or above levels of proficiency and 31 (38%) intermediate or below. Only 11 (13%) had used AWE in their writing before the time of the experiment. The teacher participant had taught academic English writing for 11 years and had 4 years of AWE user experience.

Eighteen non-AWE users and two students with some AWE experience from the class were invited for a pilot test during Week 2 to ensure the clarity of the questionnaire items. All participants completed the four-stage questionnaires, and a total of 20 students (10 from the

group of higher levels and 10 lower) were randomly selected and invited to participate in semi-structured group interviews.

Data Collection

Table 3 summarizes the research instruments used for data collection, the content of each instrument, the time of data collection, and the corresponding RP and RQ of each method.

5-Point Likert-Scale Questionnaires

A quantitative 5-point Likert-scale questionnaire was adopted and revised from Cavaleri and Dianati (2016) and Miranty and Widiati (2021). The original questionnaires comprised three domains: student perceptions about themselves in the writing process, student perceptions about *Grammarly* usefulness, and student perceptions about *Grammarly* drawbacks. Here, we designed two revised versions: Questionnaire 1 for the pre-writing, writing, and post-writing self-editing stages and Questionnaire 2 for the post-writing peer review stage. They shared Domains 1 and 2 but differed in Domain 3, where Questionnaire 1 examined student perceptions about the effectiveness of *Grammarly* in reducing writing difficulties and Questionnaire 2 IFM.

Each questionnaire comprised 34 items. To achieve the three RPs, we made four modifications. First, we added some new and revised statements into the three domains to address several challenges, such as language use and aspects (vocabulary, argument development, coherence and unity). Moreover, we substituted more specific and explicit descriptive statements (i.e., Domain 3, Statements 20–34) for some writing and revising behaviors overlooked in Cavaleri and Dianati (2016) and Miranty and Widiati (2021). The inclusion of the

Domain 3 items examined how the integration of *Grammarly* at the first three stages and IFM at the post-writing peer review stage affected writing behavior and performance.

Second, as per the pilot test results, we modified negative statements (5, 8, 9, 12, 16, 17, and 18) into affirmative ones to prevent confusion. We found that nine students had difficulties making judgments when statements were negative. For example, although four students verbally reported that Statement 8 (“I don’t always understand the feedback I get in my writing”) described their writing experience, they marked “2 Disagree,” as explained by one student subject, “I don’t always understand... so I don’t understand... so I disagreed.” This misconception may arise from the double negation in statements (e.g., “I don’t agree with the statement that I don’t always...”).

Third, we did not include statements regarding operating *Grammarly* (“I have technical issues with *Grammarly*” and “*Grammarly* is easy to use, especially in writing class”). *Grammarly* was mostly used in class as a pedagogical technique, and we provided early in the semester two 30-min drill workshops where students were offered sufficient time and practice to master the operation and familiarize themselves with the functions.

Lastly, we focused the items in Domain 3 on Questionnaire 2 upon the implementation of IFM, instead of *Grammarly* drawbacks, to scrutinize the perceived effectiveness of the integrated feedback of *Grammarly* and peer review.

The revised questionnaires were distributed during each of the four stages, and a total of 328 questionnaires, along with consent forms, were obtained.

Interviews

Semi-structured interviews, unlike a straightforward question-and-answer format between interviewers and

Table 3. Instrument Designs and Corresponding RPs and RQs.

Instrument	Content	Time of collection	Corresponding RP/RQ
5-Point Likert-Scale Questionnaire I	Investigation of student perceptions about themselves in the writing process, the usefulness of <i>Grammarly</i> in writing, and the effectiveness of <i>Grammarly</i> in reducing writing difficulties	Pre-writing (W4), writing (W7), and post-writing self-editing stages (W11)	RPs 1 & 3 RQ1
5-Point Likert-Scale Questionnaire II	Investigation of student perceptions about themselves in the writing process, the usefulness of <i>Grammarly</i> in writing, and the effectiveness of IFM in reducing writing difficulties	Post-writing peer review stage (W15)	RPs 2 & 3 RQ2
Student Group Semi-Structured Interview	Confirmation of student perceptions and exploration of reasons and evidence for student decisions in the quantitative data	W18	RP 3 RQs 1 & 2

interviewees, promote focused, conversational, two-way communication via guiding questions (Burgess, 1984). Semi-structured group interviews were conducted to probe into the causality of variables (i.e., writing difficulties, writing performance, revising behaviors, and *Grammarly* and IFM effectiveness) and seek more in-depth and explicit evidence of student perceptions. An interview guide and a list of pre-determined interview questions (IQs) were formulated from the quantitative data. The IQs are presented below:

- IQ1: How does *Grammarly* enhance your writing? Please provide explanations or examples.
- IQ2: Please name the benefits and disadvantages of using *Grammarly*.
- IQ3: How does IFM enhance your writing? Please provide explanations or examples.
- IQ4: Please name the benefits and disadvantages of implementing IFM.
- IQ5: What would you recommend if IFM were to be implemented in a large-sized writing class?

There were five interviewee groups of four (20 students in total). Each was interviewed for 1.5 hr, and each interview was audio-recorded. All the interviews were conducted in the same classroom and post-course. The guiding questions were sent to the interviewees 1 week beforehand.

Data Analysis

The quantitative results were calculated descriptively using software R version 4.1.0 to acquire the mean scores of each item, domain, and questionnaire. The one-way repeated measures analysis of variance (ANOVAs) using the *ez* package were gathered to examine student perceptions and the transformation of perceptions regarding integrating AWE via different means and during different stages. In addition, the Bonferroni-adjusted significance test for pairwise comparisons was respectively applied to examine the differences between stages.

All interview data were transcribed and coded thematically and inductively following the thematic analysis approach to generate research question-related categories. NVivo 12 was used for organizing and coding the qualitative data, because, as suggested by Malawi, “[T]he presence of nodes in NVivo makes it more compatible with grounded theory and thematic analysis approaches. Moreover, the nodes provide ‘a simple to work with structure’ for creating codes and discovering themes” (2015, p. 14).

The audios were transcribed, significant information was classified into nodes, and nodes formed into parent nodes (i.e., themes) that addressed the RQs. To be

specific, transcriptions were read, reread and analyzed line by line. Major points were first identified and coded with different nodes, such as accuracy uncertainty, useful tool, timely feedback, problems with explaining ideas, and formal expressions. After initial coding, all data were collated into groups by nodes. Amongst these nodes, patterns were identified to generate themes, such as enhancement of revision/language use/comprehension/writing quality, elevated anxiety, quality of feedback, psychological effects, and effectiveness with varied degrees. What emerged was the initial codes related to student perceptions about themselves during the writing process and their opinions regarding the usefulness and effectiveness of the AWE integration.

Furthermore, we conducted the coding consistency check to ensure credibility. After we each completed respective coding, we collaboratively deliberated on possible codes and explained and justified different codes for better categorization. We then compared the categorized data with notable similarities, differences, and recurring themes highlighted. We analyzed only the themes most relevant to the RQs for analytical convenience and simplicity (Creswell, 2007).

Procedures

Figure 2 outlines the experimental procedures. As illustrated, over the course of the 18-week semester, three types of data were collected, and four measures were taken. The student and teacher consents were obtained prior to the implementation of this research.

The post-writing peer review comprised four steps. To begin, the students were divided into groups of four. Next, each team was given the peer review discussion sheet (Table 4), and each member then took turns to present his or her work and the *Grammarly* feedback as hard copies. In the third step, each team discussed the *Grammarly* feedback and answered the guiding statements. After the completion of the 1-hr peer review discussion, the students had 30 min to reflect on peer feedback and revise their own work.

Credibility

Cronbach’s alpha is a statistic used by researchers to demonstrate a scale constructed or adopted for a study is appropriate for the purpose it sets out (Taber, 2018). As shown in Table 5, the Cronbach alpha values in this study indicate acceptable and good internal consistency. In addition, as the validity of a research instrument assesses the extent to which the instrument measures what it is designed to measure (Robson, 2011), the questionnaire used here were adopted from two previous literature, Cavaleri and Dianati (2016) and Miranty and

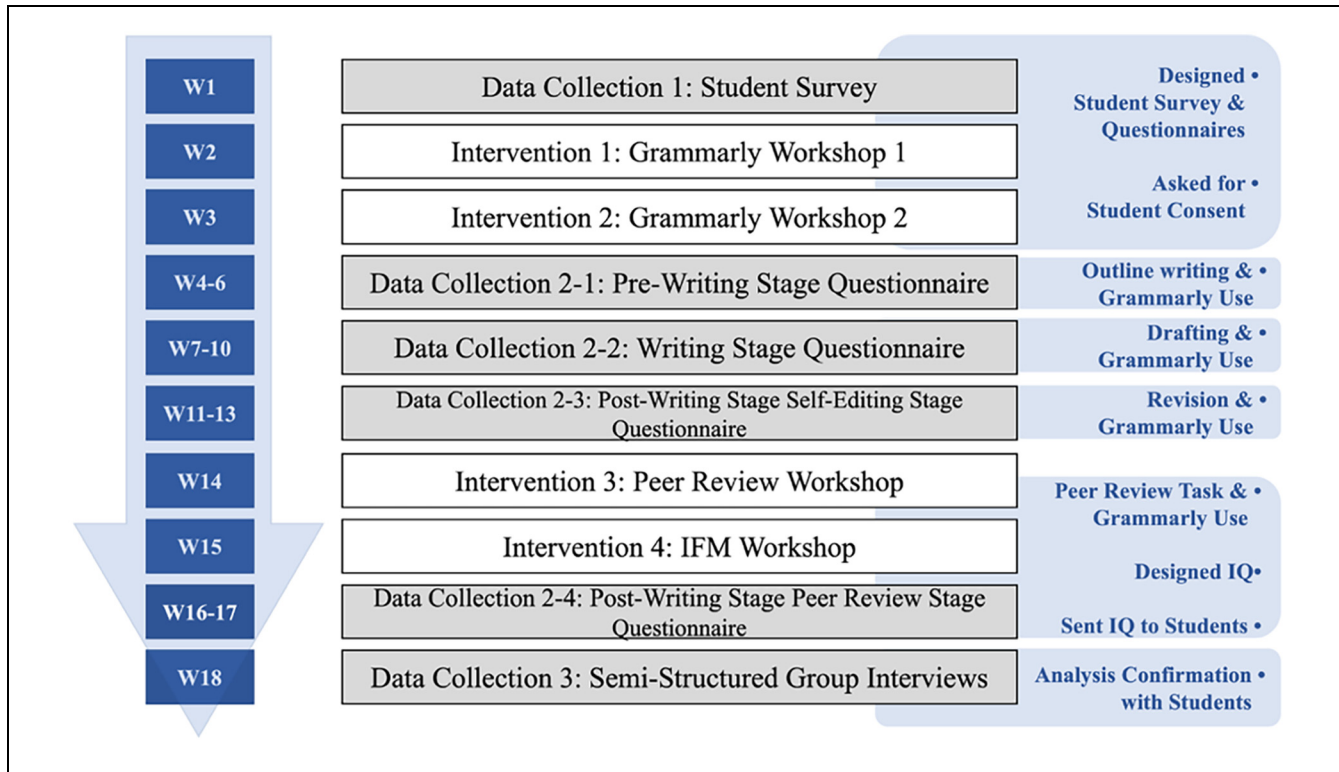


Figure 2. Procedures for data collection and intervention implementation.

Table 4. Peer Review Discussion Sheet.

Writer: teammates:

Part I. Organization, Argument, & Content

- The work has a topic sentence, supporting sentences, and a concluding sentence.
- The arguments are well-developed with relevant details and good logic.
- The work critically presents the writer's stance and idea(s).
- The persuasion of the arguments is achieved by using appropriate evidence and proper language and tone.
- The work is interesting to read.

The content successfully responds to the topic.

Part II. Overall Comment

What I like about the work

What I would like to see in the work

Table 5. Internal Consistency of the Questionnaires.

Domain	Sample size	Items	Cronbach's alpha
Perceptions about Themselves in the Writing Process	82	9	.823
Perceptions about the Usefulness of Grammarly	82	10	.782
Perceptions about Grammarly (IFM) Effectiveness in Easing Students' Writing Difficulties	82	15	.852
Internal consistency of data			
Cronbach's alpha range			
≥ .9	Excellent		
.8–.9	Good		
.7–.8	Acceptable		
.6–.7	Questionable		
.5–.6	Poor		
< .5	Unacceptable		

Widiati (2021), and it was further revised to form more holistic coverage of the writing challenges an L2 writer may encounter. The revised questionnaires were then piloted and final revisions were made.

As for the qualitative instrument, the guiding questions were piloted and sent to the interviewees 1 week beforehand, and the transcriptions were then sent to the interviews for confirmation. To ensure the credibility of the study, the coding consistency check was performed, where the coding was processed by the three authors independently and then collaboratively cross-examine the respective analysis.

Results and Analysis

This section comprises the presentation of quantitative questionnaire results clustered by domains and the explanation of the analysis of qualitative interview data.

Questionnaire Results

Tables 6 to 9 show the quantitative results gathered from the questionnaires. Results of domain comparison and item comparison are provided in Tables 6 and 7 and Tables 8 and 9, respectively.

Table 6 presents the comparison results between four stages using one-way repeated measures ANOVAs and pairwise comparison. As shown, there were statistically significant differences between stages within each domain ($p < .05$; $p < .001$), suggesting that at least one stage within each domain displayed a different mean score compared with the others.

To better understand the differences between stages in each domain, pairwise comparisons were performed, and p -values were adjusted using the Bonferroni correction (Figure 3). Significant differences were found between the pre-writing and post-writing self-editing stages and those between the post-writing self-editing and post-writing peer review stages in Domain 1 ($p < .05$ and $p < .0001$,

respectively). With regard to Domains 2 and 3, significant differences were observed in all stages ($p < .001$) with increasing mean scores.

Table 7 shows the results of the mean scores and standard deviation values of each statement item at four stages to examine student perceptions about themselves in the writing process. They all showed statistical significance ($p < .05$), suggesting that there was a statistically significant difference between the mean scores of the different levels of the score at least at one or more stages.

Most statements in Domain 1 were rated between “neutral” and “strongly agree,” except for Statements 8 and 9. The majority of the participants perceived Statements 3 and 5 to be the most agreeable at all stages ($m = 4.28, 4.48$), and both perceptions (difficulty in expressing ideas in writing and anxiety about being inaccurate) were intensified as time progressed. In contrast, Statements 8 and 9 were considered least agreeable ($m = 2.95, 1.63$) with decreasing degrees of difficulty in understanding feedback and perceived need for assistance in proofreading and English writing.

Table 8 presents the results of student perceptions about the usefulness of the *Grammarly* feedback at four different stages. Significant differences were observed ($p < .05$), suggesting a significant difference between the mean scores of the different levels of the score at least at one or more stages.

As shown, Statement 11 received the lowest average mean score among all the Domain 2 items ($m = 2.57$) and those at the pre-writing and writing stages ($m = 1.98, 2.04$). While the writing frequency of students was not or only marginally affected, their revision behavior was more affected by comparison (Statement 12).

Statements 15 (the quality of *Grammarly* explanations) and 18 (the comprehension level of *Grammarly* comments) received consistently lower ratings in the first three stages. This result suggests that some students might consider AWE feedback satisfactorily useful but the delivery dissatisfying. However, in Statement 15, the rating at the

Table 6. Comparison Results Between Four Stages Within Each Domain.

	Global mean	Pre-writing		Writing		Self-editing		Peer review		p -value
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	
1. Perceptions about Themselves in the Writing Process	3.35	3.34	0.46	3.37	0.45	3.38	0.37	3.33	0.35	.01*
2. Perceptions about the Usefulness of Grammarly	3.35	2.78	0.40	3.25	0.30	3.59	0.27	3.76	0.21	<.001**
3. Perceptions about Grammarly (IFM) Effectiveness in Easing Students' Writing Difficulties	3.58	3.29	0.33	3.52	0.21	3.63	0.22	3.88	0.24	<.001**

* $p < .05$. ** $p < .001$.

Table 7. Student Perceptions About Themselves in the Writing Process.

Perceptions about themselves in the writing process	Global mean	Pre-writing		Writing		Self-editing		Peer review		p-value
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	
1. My knowledge of English grammar is weak.	3.27	3.26	0.91	3.34	0.93	3.29	0.91	3.17	0.86	.002*
2. My knowledge of English vocabulary is very limited.	3.55	3.54	0.86	3.61	0.86	3.55	0.79	3.49	0.79	.009*
3. I am fine with English grammar, but I find it difficult to express my ideas in writing.	4.28	4.18	0.59	4.24	0.62	4.33	0.57	4.38	0.54	<.001**
4. I find developing arguments in writing very challenging.	3.78	3.71	0.68	3.74	0.66	3.83	0.56	3.85	0.55	<.001**
5. I am sometimes worried that I have written incorrect sentences.	4.48	4.22	0.61	4.46	0.55	4.62	0.49	4.60	0.49	<.001**
6. I am sometimes worried about the quality of coherence of my work.	3.14	3.12	0.69	3.10	0.70	3.21	0.60	3.15	0.57	.031*
7. I am sometimes worried about the quality of unity of my work.	3.10	3.06	0.69	3.05	0.66	3.17	0.58	3.11	0.52	.016*
8. I sometimes find it difficult to understand the feedback I receive about my writing.	2.95	3.10	0.68	2.99	0.73	2.88	0.66	2.82	0.59	<.001**
9. I need a proofreading service and help with writing in English.	1.63	1.84	0.37	1.76	0.43	1.55	0.50	1.37	0.48	<.001**

* $p < .05$. ** $p < .001$.

post-writing peer review stage increased, showing an improved appreciation for *Grammarly* feedback.

Positive perceptions of the *Grammarly* feedback were observed in the items examining the usefulness of *Grammarly* feedback (Statements 16 and 17) and its impacts on revision behavior (Statement 12) and revision quality (Statement 19). Notably, while it was reasonable to assume that feedback quality and student revision behavior might be closely associated, the data showed the opposite. Despite the global mean scores being positive, the ratings of Statements 16 and 19 at the pre-writing stages remained the lowest among their own four-stage results ($m = 3.29$ and 2.4 , respectively). Interestingly, while the mean score of Statement 16 at the post-writing peer review stage was the lowest among its four stages, Statement 19 had the highest.

Table 9 presents student perceptions of the effectiveness of the two different methods of AWE integration in addressing writing challenges. The first three stages examined the efficacy of the *Grammarly* feedback and the post-writing peer review stage the IFM. Overall, the former revealed mixed outcomes, yet the latter demonstrated positive perceptions of various degrees. The global mean score of each item regarding *Grammarly* effectiveness showed that most items received positive

responses, except for the ones relevant to the essay structure, writing logic and reasoning, and developments of topic sentences and arguments, in other words, the higher-order concerns. In contrast, Statements 22 and 23 were two of the three items rated “agree” among the 15, signifying the effectiveness of *Grammarly* in managing lower-order concerns.

Moreover, Statement 20 received the highest rating at all stages, confirming that the students responded to these questionnaires based on their true experiences with the tools. The majority of the students gave positive responses to Statement 33, where their progress in writing was agreed on. Most ratings fell between “agree” for Statement 34 ($m = 3.67$), where their perceived improvements in writing skills were confirmed.

Items examining the effectiveness of IFM at the post-writing peer review stage all showed the highest ratings of all stages; in other words, the students considered the IFM feedback more effective than sole *Grammarly* feedback. This might reflect their positive writing experience during the IFM-implemented phase, where they appreciated the effectiveness of the integration of *Grammarly* feedback and peer review discussion in resolving their writing challenges. Although the items regarding the assessments of higher-order concerns received lower ratings than the remaining items,

Table 8. Student Perceptions About the Usefulness of Grammarly.

Perceptions about the usefulness of Grammarly	Global mean	Pre-writing		Writing		Self-editing		Peer review		p-value
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	
10. Grammarly gives detailed feedback.	3.50	3.06	0.55	3.10	0.68	3.83	0.58	4.01	0.53	<.001**
11. I write more after using Grammarly.	2.57	1.98	0.47	2.04	0.69	3.06	0.71	3.22	0.63	<.001**
12. I revise my essays more after using Grammarly.	3.22	2.26	0.72	3.27	0.83	3.60	0.68	3.77	0.53	<.001**
13. I feel more confident about my writing after using Grammarly.	3.57	2.98	0.59	3.34	0.63	3.87	0.49	4.10	0.40	<.001**
14. Grammarly makes helpful suggestions for improving my work.	3.55	2.37	0.79	3.73	0.57	3.88	0.48	4.21	0.41	<.001**
15. Grammarly gives good explanations about my errors.	2.91	2.79	0.58	2.83	0.62	2.77	0.55	3.26	0.47	<.001**
16. The feedback is always helpful.	3.50	3.29	0.75	3.71	0.64	3.91	0.59	3.10	0.68	<.001**
17. I agree with many of the suggestions.	3.85	3.46	0.71	3.70	0.75	4.06	0.64	4.20	0.55	<.001**
18. I can understand the comments given by Grammarly.	3.21	3.23	0.48	3.04	0.43	2.95	0.44	3.62	0.60	<.001**
19. The advice given by Grammarly helps me revise my essays.	3.58	2.40	0.83	3.77	0.55	4.01	0.46	4.13	0.34	<.001**

* $p < .05$. ** $p < .001$.

their ratings were comparatively much higher than those of the same items at the other three stages. Sole Grammarly use and IFM implementation exhibited different degrees of impact on perceived effectiveness in overcoming their writing difficulties.

Domain 3, where all stages were compared, showed higher mean scores on the items at the two post-writing stages, an indication of greater perceived efficacy during the revision stages. This observation was also evident in the result of Statement 33, where the students acknowledged more significant progress at the latter stages. Notably, the new function, a Plagiarism Checker (Statement 32), received consistently high ratings at four different stages. Less proficient participants rated “agree” or “strongly agree,” and many more proficient participants rated “agree” or “neutral.”

Interview Findings

Several themes merged during the transcription coding. Here, to ensure a focused paper and analytical convenience and simplicity, only themes most relevant to the RPs are presented.

Intensified Anxiety

Nine out of twenty interviewees shared their intensified anxiety about being incorrect. As time progressed, the

experience of AWE practice offered the students opportunities to access their existing writing skills and knowledge, thereby providing insights into their own language needs and limits. Although providing the students with some guidelines on what and how to improve, this new grasp could have intensified their writing anxieties, as described by one of the interviewees.

“There were 11 comments related to the article usage in my second assignment. I knew the use of articles was one of my weaknesses, but I was surprised at how frequent I made such errors. So, now I always go back to check twice on the articles when I finish writing... In my last assignment, I carefully checked them during writing, used Grammarly to check again, asked my writing partner to check, and I went back to check the revisions again,” (Group D, Subject 19)

Two interviewees offered different insights. They praised the peer review discussion at the post-writing stage for facilitating peer support, as described as the “emotional support needed” by Subject 4 in Group D and portrayed as “an integral element in reducing stress” by Subject 15 in Group C.

Constructive, Yet Inadequate Comments

Most students recognized to some extent the usefulness of Grammarly feedback, but they criticized the inadequacy of comment explanations and the low recall rate

Table 9. Student Perceptions About the Effectiveness of *Grammarly* and IFM.

Perceptions about Grammarly (IFM) effectiveness in easing students' writing difficulties	Pre-writing		Writing		Self-editing		Peer review		p-value
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
20. I revise essays in Grammarly/IFM before submitting them to my teacher.	4.88	0.40	4.95	0.22	4.99	0.11	5.00	0.00	.005*
21. Grammarly/IFM can correct format errors.	3.17	0.58	3.41	0.54	3.48	0.55	3.60	0.56	<.001**
22. Grammarly/IFM can correct spelling errors.	3.93	0.64	4.23	0.45	4.40	0.49	4.57	0.50	<.001**
23. Grammarly/IFM can correct grammar errors.	3.95	0.63	4.10	0.54	4.18	0.42	4.34	0.48	<.001**
24. Grammarly/IFM can improve my wording.	3.73	0.59	3.94	0.40	4.01	0.46	4.09	0.36	<.001**
25. Grammarly/IFM can improve my essay content.	2.91	0.28	3.16	0.43	3.21	0.46	3.71	0.64	<.001**
26. Grammarly/IFM can improve my essay structure.	2.72	0.50	2.93	0.38	3.04	0.37	3.46	0.63	<.001**
27. Grammarly/IFM can improve my logic and reasoning skills in writing.	2.10	0.43	2.32	0.59	2.45	0.61	3.09	0.67	<.001**
28. Grammarly/IFM can help develop my topic sentence in writing.	2.72	0.50	2.90	0.40	3.00	0.42	3.51	0.59	<.001**
29. Grammarly/IFM can improve my argument in writing.	2.91	0.36	3.01	0.25	3.04	0.25	3.28	0.48	<.001**
30. Grammarly/IFM enhances the quality of coherence of my work.	2.94	0.62	3.40	0.49	3.55	0.50	3.71	0.51	<.001**
31. Grammarly/IFM enhances the quality of unity of my work.	2.91	0.59	3.10	0.49	3.23	0.45	3.40	0.63	<.001**
32. Grammarly/IFM helps me avoid plagiarism.	4.05	0.22	4.27	0.45	4.40	0.49	4.44	0.50	<.001**
33. I have made some progress in my writing.	3.15	0.55	3.34	0.50	3.48	0.53	3.76	0.53	<.001**
34. Grammarly/IFM helps me improve my writing skills.	3.22	0.61	3.76	0.60	4.02	0.54	4.21	0.41	<.001**

* $p < .05$. ** $p < .001$.

of the AWE feedback. According to three interviewees, the AWE feedback was satisfactorily useful for revision but questionable as a learning-reinforcement tool due to its very limited explanation of the writing issues. One interviewee explained this in more detail:

"A lot of time the comments are confusing. For example, this is my last assignment. [showing the assignment to the others]. Ther's one sentence here, 'Passionate people work hard to explore their potentials, and they have the urge to find mates to spend time together exchanging their ideas,' Grammarly suggested changing the word 'potential,' to singular form, and the explanation given was 'it seems that potentials may not agree in number with other words in this phrase,' I had no clue of what it meant, so I just went ahead click and accept." (Group A, Subject 1)

Providing unclear explanations could result in ambivalent attitudes toward the AWE feedback, as explained by one interviewee:

"I accepted some suggestions but ignored some, too. I do't think all of the suggestions were right. The explanations were not really clear or specific. So, I usually made my own decisions. But there're times that I felt I was the wrong one and just accepted the comments given by Grammarly." (Group C, Subject 14)

Grammarly Feedback as a Less Effective Tool at Earlier Stages

Although students appreciated the AWE feedback at the latter two stages, they were less impressed with it at the pre-writing and writing stages. This might result from the low frequency of *Grammarly* use at the pre-writing stage, hence less need, as explained by one interviewee:

"I spent most time brainstorming ideas and structuring information. Grammarly was seldom used," (Group C, Subject 13)

One more possible cause was the distinct disparity between the upper-intermediate level students (Groups A and C) and less proficient students (Groups B and D). The less proficient students perceived *Grammarly* as a reinforcement for boosting confidence. The more proficient students, by comparison, perceived it as an interruption in their writing flow, for functions like auto-corrects encouraged writers to frequently revise micro-issues, as explained by one interviewee:

"My writing was constantly interrupted when I was writing. I paused whenever a possible error was flagged. The 'unstopabl', error warning seriously distracted me, and I couldn't

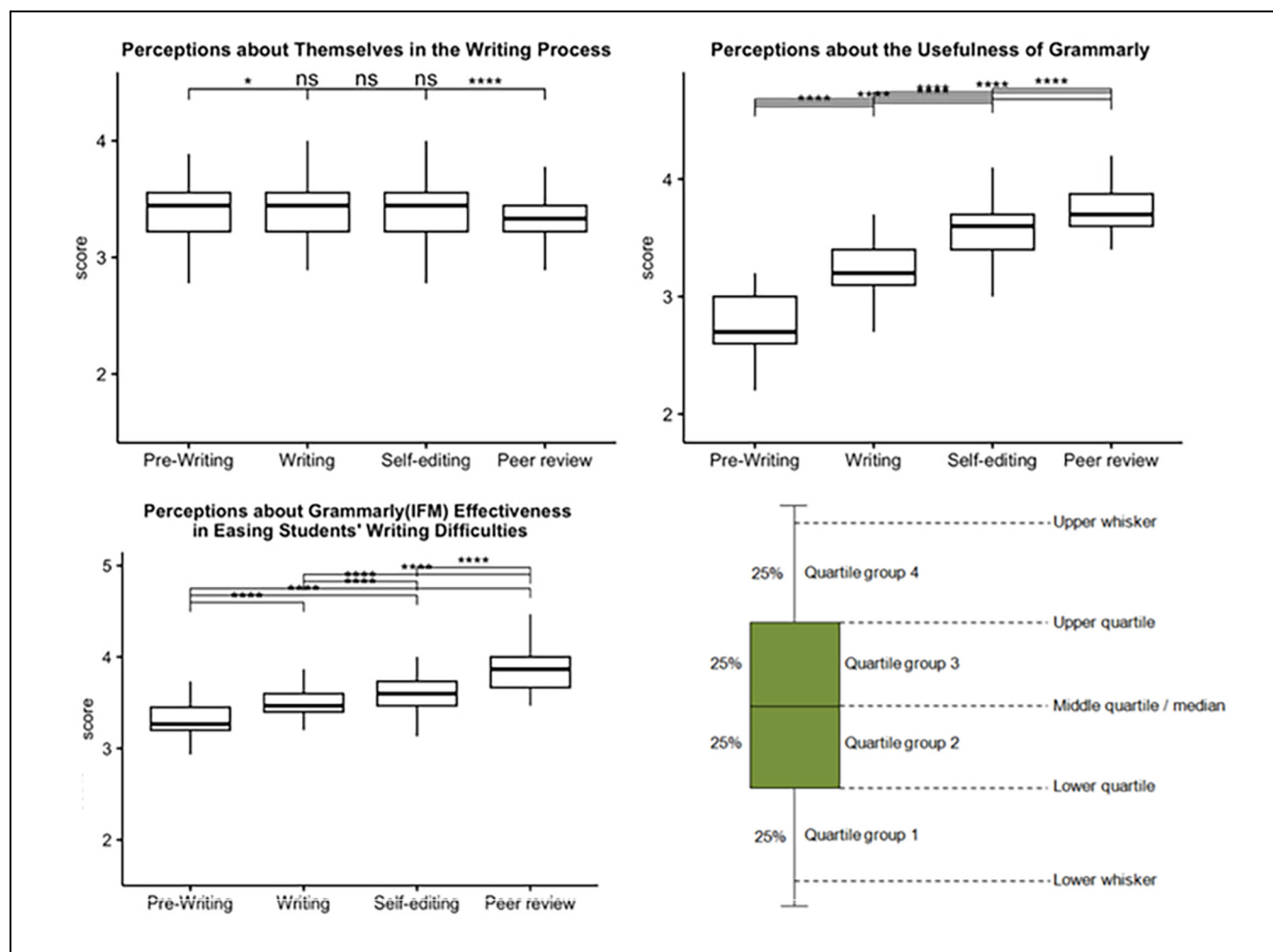


Figure 3. Pairwise comparison between the stage of the within-subjects factor.

stop going back to check what I had written. My writing pace and flow suffered.” (Group C, Subject 14)

Nevertheless, even though 11 interviewees considered Grammarly less effective in addressing higher-order concerns, six of them shared a mutual belief that the elimination of surface-level errors could strengthen the delivery of their argument:

“The feedback helped me revise smaller parts. Not much information was provided about the paragraph structure or argument presentation. But if the smaller errors could be fixed, the overall arguments would be more comprehensible and persuasive, I think,” (Group D, Subject 18)

Enhanced Confidence

Because the students confirmed the efficacy of the AWE feedback in correcting surface-level errors, they demonstrated higher confidence in their writing quality. This

was particularly evident in the post-writing stages, as explained by one interviewee:

“No matter how careful I am, I always manage to find errors in writing. Grammarly serves as my second pair of eyes. It can detect some careless errors like spelling, punctuation, and some grammar problems. I do’t have to mind those minor problems. I’m sure that those are ‘AI-fixable’, I only have to worry about my ideas.” (Group D, Subject 17)

IFM as a Comprehension Facilitator and a Writing Reinforcement

The students perceived the incorporation of the AWE feedback with peer review discussion as a facilitator. The incorporation improved their comprehension of the AWE feedback and strengthened their writing quality in terms of the organization and development of content, argument, and logic, as explained by two interviewees:

"We discussed actively and the outcomes were especially helpful for clarifying many of my doubts. We went over all the Grammarly feedback and revised about two-thirds of the flagged problems. And we disagreed with some of the comments thinking they were over-corrected," (Group B, Subject 6)

"My writing partner gave me some ideas about the development of my arguments and possible sources of more suitable evidence. The [Grammarly] comments suggested shorter and more concise sentences, so when I finalized my work, I adopted my partner's suggestion to use a compare-and-contrast pattern and tried to write shorter sentences to avoid run-on or wordy sentences." (Group A, Subject 3)

Notably, the higher appreciation for peer review discussion at the post-writing stage could provide some insights into why the students perceived *Grammarly* comments to be less helpful but more useful in enhancing revisions since the AWE feedback could be used as a guideline for the peer review discussion. However, this realization might turn attention to the insufficiency of *Grammarly* as a sole source of feedback, hence the less perceived helpfulness of the AWE feedback at the post-writing peer review stage.

In addition to reinforcing writing, IMF promoted negotiation, reflection, perspective-taking, critical analysis and self-reflection, as evidenced in the sharing of learning actions and writing responses by two interviewees:

"I always proofread my works after I finish them. When I used it [Grammarly] for my third assignment, I had clearer ideas where to pay attention to. So, this kept me focused. And for my last homework, in general, there was't much difference in the total number or in the decision of revisions I made between using it at the proofreading stage or at peer discussion. But I got the chance to see how people comprehended my work and interpreted my message. It's interesting to see my own work from the eyes of an engineer." (Group C, Subject 15)

"I read my teammate's work and I liked his 'Butterfly vs. Eagle,' analogy, so I tried to come up with my own analogy in my next assignment. It didn't turn out to be as well-received as I thought but I would give it another try again. Maybe start with IG share?"

Discussion and Implications

Disparities in the Perceived Usefulness of Grammarly are Found at Different Stages

The frequency of AWE adoption can be a determinant of its value (Chen & Cheng, 2008). In the pre-writing stage, *Grammarly* is perceived to be least useful, for it is less needed in outlining and planning phases. In the writing stage, *Grammarly* is perceived as a distraction for more proficient students but as a confidence-booster for less proficient students. In the post-writing self-editing stage, *Grammarly* is perceived more useful by enabling better revisions and self-reflections. In the post-writing

peer review stage, *Grammarly* is perceived to be the most useful. Not only does it act as a crucial guideline for peer review discussions, but *Grammarly* also enables self-reflection, meaning negotiation, perspective-taking and critical analysis—some merits fostered during peer review discussion as argued by some scholars (Weissbach & Pflueger, 2018).

Revision Behaviors and Performance are Enhanced at all Stages

Although Bai and Hu (2017) and Cavaleri & Dianati (2016) claim that much AWE feedback is not adopted by users, our findings favorably compare to those of Ghufroon (2019), where students find the feedback helpful for revisions. This affirmative attitude is even firmer at the post-writing peer review stage, as peer support enables better comprehension of the AWE feedback and stimulates further exploration and negotiation of ideas.

Writing Quality is Enhanced by Addressing Low-Order Concerns at Earlier Stages

This finding is consistent with previous literature examining the usefulness of *Grammarly* in language usage and writing mechanics (e.g., Bai & Hu, 2017; Zhang & Hyland, 2018). Although *Grammarly* leaves most higher-order concerns unresolved, the surface-level improvements contribute to the overall comprehensible level of arguments. This finding supports the argument of Azar (2007) and Cavaleri and Dianati (2016) that grammatical accuracy enhances writing intelligibility and ensures the clarity and precision of an intended idea.

Writing Quality is Enhanced by Addressing Higher-Order Concerns at the Peer Review Stage

The integration of *Grammarly* feedback with peer review discussion is found to be beneficial for redirecting student attention into global issues and fostering meaningful idea exchange via peer support. Peer review empowers integrated feedback by providing both-end negotiation opportunities for meaning-making of arguments, language use, and message delivery and reinforcing writing, communicating, and cognitive processing (Nagoshi et al., 2019).

Psychological Changes May Take Place Throughout the Grammarly Application

The more students practice writing, the more they realize their own limitations, and this awareness is fortified when they engage in critical peer review discussions, as argued by Zhang and Hyland (2018). Although students perceive the "user-friendly" nature of *Grammarly* and its

advantages, the perception of this merit does not lessen their writing anxiety. We observe elevated anxiety among students despite the enhancement of their confidence with timely AWE assistance. What students undergo during the writing process involves not only their issues with language and idea development but also complex psychological activities. These findings contradict those of Miranty and Widiati (2021), who suggest that the opportunity to self-reflect and self-revise through AWE reduces student anxiety due to the self-paced learning mode.

IFM Helps Alleviate Writing Anxiety and Reinforce Student Confidence

Although a majority of students may experience anxiety when being constantly reminded of their own writing weakness and insufficiency, some individuals benefit from the inclusion of peer review. Providing peer support to students in writing practice serves as a pillar of support that may lead to the alleviation of their writing anxiety (Chaktsiris & Southworth, 2019).

IFM Encourages Higher Cognition Activities and Complements Grammarly

The AWE feedback enables students to spend more time and energy on more complex tasks, and peer review encourages them to perform higher cognitive skills to complete tasks. IFM fosters meaningful interactions requiring negotiation, reflection, perspective-taking, and critical processing—all of which involve higher cognitive skills and are crucial for fostering autonomous learning. Hence, the introduction of the *Grammarly*-peer review discussion into writing instruction could maximize the utilization of the writing instructor force and minimize the possibility of excessive time and focus on a specific cognitive level, the higher cognitive level or the lower.

The Conjunction is a Complementary Combination

Content formation and argument development are two writing aspects in which most students need assistance. In addition to *Grammarly* feedback, supplementary assistance could be incorporated to reinforce AWE, such as teacher feedback (Cavaleri & Dianati, 2016; Gao, 2021). However, in a large writing class, providing teacher feedback may become a burden, wearing out teacher energy and enthusiasm. Therefore, the need for a less time- and energy-consuming alternative is called for.

This research does not directly measure explicit evidence of the development of autonomous learning, but as shown in the cases like the “Butterfly versus Eagle” analogy in the student interview, the combination of AWE feedback with peer review better enables students

to envision their learning growth, a finding that supports Miranty and Widiati (2021). In other words, IFM helps make up for the flaws of *Grammarly*, as IFM encourages higher-order concerns, provides opportunities for negotiation and further personal and specific comments, as well as encourages revisions based on self-reflection and negotiation for meaning. Also, IFM lays a good foundation of peer review (i.e., the AWE feedback provides direction and content for peer discussion, as well as offering immediate corrective feedback).

Suggestions and Conclusion

This study explores in a large university writing class different methods of *Grammarly* integration at the pre-writing, writing, and post-writing self-editing stages and *Grammarly*-peer review integration at the post-writing peer review stage. The AWE feedback receives perceived usefulness of varying degrees for revision at different stages. *Grammarly* may be ideal for reducing the writing anxiety of less proficient students in the writing stage; for more proficient learners, *Grammarly* is recommended in the post-writing stages. However, the use of *Grammarly* should be reconsidered if writing pace and idea generation are major concerns. Integrating *Grammarly* with peer review empowers feedback provision by: (1) improving higher-order concerns; (2) enhancing meaning negotiation and cognitive processing; (3) alleviating writing anxiety; (4) promoting self-reflection; and (5) fostering autonomous learning.

The results suggest that the teacher feedback burden may be alleviated if IFM were successfully implemented. The success of IFM implementation hinges on the “preparedness of measures,” the availability of sufficient preliminary knowledge and practice of the tools (i.e., *Grammarly* and peer review rubrics) at the designed workshops. In other words, IFM should be seen as a pedagogical technique and learning content rather than a supplementary tool. However, we do not recommend replacing teacher feedback with IFM. As suggested by the interviewees, teacher involvement can be modified in its means but should not be supplanted. Therefore, practitioners may consider either (1) substituting IFM for revision at the first two stages and teacher feedback at the last stage or (2) focusing on higher-order concerns such as critical argumentation, creativity, and overall logic and persuasion in peer review discussion and teacher instruction, leaving *Grammarly* to work its charm on lower-order concerns.

Acknowledgments

Special thanks to the faculty at the Academic Writing Education Center NTU, Mr. Marc Anthony, and Chong-Hsien Chiou for the kind and professional assistance.

Author Contributions

WYL, KK, and YJS contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. All authors read and approved the final manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education, Taiwan (MOE) [Project NO. PGE1080405].

An Ethics Statement for Animal and Human Studies

This study was exempted from IRB Review because it was conducted in commonly accepted educational settings and only involved the use of survey and interview procedures.

ORCID iDs

Wei-Yan Li  <https://orcid.org/0000-0002-1803-8088>

Kevin Kau  <https://orcid.org/0000-0002-7258-811X>

Yi-Jiun Shiung  <https://orcid.org/0000-0003-3951-1038>

Availability of Data and Materials Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Ågerfalk, P. J. (2013). Embracing diversity through mixed methods research. *European Journal of Information Systems*, 22, 251–256.
- Almubark, A. A. (2016). Exploring the problems faced by the teachers in developing English writing skills for the students in Saudi Arabia. *International Journal of English Language Teaching*, 10(4), 10–23.
- Alnasser, S. M. N. (2018). Exploring student-writers' views on replacing teacher feedback with peer feedback and computer-based feedback. *Arab World English Journal*, 9(3), 345–366. <https://doi.org/10.24093/awej/vol9no3.23>
- Anderson, O. S., El Habbal, N., & Bridges, D. (2020). A peer evaluation training results in high-quality feedback, as measured over time in nutritional sciences graduate students. *AJP Advances in Physiology Education*, 44(2), 203–209. <https://doi.org/10.1152/advan.00114.2019>
- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. The annual meeting of the National Council on Measurement in Education. San Diego, CA, United States.
- Azar, B. (2007). Grammar-based teaching: A practitioner's perspective. *TESL-EJ*, 11(2), 1–12.
- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychologist*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Barrot, J. S. (2022). Integrating technology into ESL/EFL writing through Grammarly. *RELC Journal*, 53, 764–768. <https://doi.org/10.1177/0033688220966632>
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Burgess, R. (1984). *In the field: An introduction to field research*. Routledge.
- Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing evaluation service. *AI Magazine*, 25(3), 27–36. <https://doi.org/10.1609/aimag.v25i3.1774>
- Cavaleri, M., & Dianati, S. (2016). You want me to recheck your grammar? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, 10(1), 223–236.
- Chaktsiris, M. G., & Southworth, J. (2019). Thinking beyond writing development in peer review. *The Canadian Journal for the Scholarship of Teaching and Learning*, 10(1), 1–22. <https://doi.org/10.5206/cjsotl-rcacea.2019.1.8005>
- Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology*, 12, 94–112.
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, 27(3), 419–436. <https://doi.org/10.1177/0265532210364391>
- Choi, J. (2010). Educating citizens in a multicultural society: The case of South Korea. *The Social Studies*, 101(4), 174–178. <https://doi.org/10.1080/00377990903284153>
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among fire approaches*. Sage Publications.
- Creswell, J. W., & Plano Clark, V. L. (Eds.). (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage Publications.
- Cui, Y., Schunn, C. D., Gai, X., Jiang, Y., & Wang, Z. (2022). Corrigendum: Effects of trained peer vs. Teacher Feedback on EFL students' writing performance, self-efficacy, and internalization of motivation. *Frontiers in Psychology*, 13, 878147. <https://doi.org/10.3389/fpsyg.2022.878147>
- Darayani, N. A., Karyuatry, L. L., & Rizqan, M. D. A. (2018). Grammarly as a tool to improve students' writing quality. *Jurnal Sains Sosial dan Humaniora*, 2(1), 83–89.
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5(1), 1–35.
- 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>
- Dodgson, A. N., Tariq, B., Alauyah, M., & Yusof, M. (2016). The secondary school students' usage of English learning websites to self-correct writing errors. *ASIAN TEFL Journal of Language Teaching and Applied Linguistics* 1(1). <https://doi.org/10.21462/asianteftl.v1i1.3>
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Lawrence Erlbaum Associates Inc.

- Ferris, D. R. (Ed.). (2011). *Treatment of error in second language student writing* (2nd ed.). The University of Michigan Press.
- Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system Pigai. *International Journal of Emerging Technologies in Learning (iJET)*, 16(11), 322–330.
- Ghufron, M. A. (2019, April 27). *Exploring an automated feedback program "Grammarly" and teacher corrective feedback in EFL writing assessment: Modern vs. traditional assessment* [Conference session]. *Proceedings of the 3rd English Language and Literature International Conference*, 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. <https://doi.org/10.21512/lc.v12i4.4582>
- Grimes, D. (2008). *Middle school use of automated writing evaluation: A multi-site case study* [Ph.D. thesis, University of California]. University of California.
- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment*, 8(6), 1–43.
- Han, N. R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12, 115–129.
- Hentasmaka, D., & Cahyono, B. Y.; Universitas Negeri Malang, East Java, Indonesia, d.hentasmaka@gmail.com, Universitas Negeri Malang, East Java, Indonesia, bambang.yudi.fs@um.ac.id. (2021). Peer feedback uptakes and outcomes across EFL students' proficiency levels: A study at tertiary education in Indonesia. *International Journal of Instruction*, 14(3), 271–286. <https://doi.org/10.29333/iji.2021.14316a>
- Hu, G. (2007). Developing an EAP writing course for Chinese ESL students. *RELC Journal*, 38, 67–86. <https://doi.org/10.1177/0033688206076160>
- Huisman, B., Saab, N., van Driel, J., & van Den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955–968. <https://doi.org/10.1080/02602938.2018.1424318>
- Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2020). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, 57(3), 328–338. <https://doi.org/10.1080/14703297.2019.1630294>
- Jackson, M. A., Tran, A., Wenderoth, M. P., & Doherty, J. H. (2018). Peer vs. Self-grading of practice exams: Which is better? *CBE Life Sciences Education*, 17(3), es44. <https://doi.org/10.1187/cbe.18-04-0052>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Kusumaningrum, S. R., Cahyono, B. Y., & Prayogo, J. A. (2019). The effect of different types of peer feedback provision on EFL students' writing performance. *International Journal of Instruction*, 12(1), 213–224.
- Lai, H. Y., & Tu, J. F. (2020). Technology university students' attitudes towards the English graduation benchmark (EGB): A comparison between English-major and non-English major students. *Sustainability*, 12, 2461. <https://doi.org/10.3390/su12062461>
- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning and Technology*, 12, 50–68.
- Law, S., & Baer, A. (2020). Using technology and structured peer reviews to enhance students' writing. *Active Learning in Higher Education*, 21(1), 23–38. <https://doi.org/10.1177/1469787417740994>
- Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2010). Automated grammatical error detection for language learners. *Synthesis Lectures on Human Language Technologies*, 3(1), 1–134. <https://doi.org/10.2200/s00275ed1v01y201006hlt009>
- 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>
- Liao, H. C. (2015). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308–319.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30–43.
- Mahmood, D., & Jacobo, H. (2019). Grading for growth: Using sliding scale rubrics to motivate struggling learners. *Interdisciplinary Journal of Problem-Based Learning*, 13(2), <https://doi.org/10.7771/1541-5015.1844>; <https://docs.lib.purdue.edu/ijpbl/vol13/iss2/10/>
- Miranty, D., & Widiati, U. (2021). Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts. *Journal of Education and Instruction*, 11(4), 126–137.
- Mohsen, M. A. (2022). Computer-mediated corrective feedback to improve L2 writing skills: A meta-analysis. *Journal of Educational Computing Research*, 60(5), 1253–1276. <https://doi.org/10.1177/07356331211064066>
- Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education*, 15(2), 157–171. <https://doi.org/10.1177/1469787414527391>
- Nagoshi, K., Zaidi, Z., Wright, A., & Stalvey, C. (2019). Peer assisted feedback: A successful approach for providing feedback on United States medical licensing exam-style clinical skills exam notes in the United States. *Journal of Educational Evaluation for Health Professions*, 16, 29. <https://doi.org/10.3352/jeehp.2019.16.29>
- Nova, M. (2018). Utilizing grammarly in evaluating academic writing A narrative research on EFL students' experience. *Journal of English Education and Applied Linguistics*, 7(1), 80–96.
- Nusrat, A., Ashraf, F., & Saeed, R. (2019). Effectiveness of instructors' and peers' oral feedback on the accuracy of English writing: A study of Pakistani ESL undergraduate learners. *International Journal of English Linguistics*, 9(4), 314.
- O'Neill, R., & Russell, A. (2019a). Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker. *Journal of Academic Language and Learning*, 13(1), A88–A107.

- Page, E. B. (2003). Projectessay grade: PEG. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). Lawrence Erlbaum Associates Inc.
- 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>
- Philips, S. M. (2007). *Automated essay scoring: A literature review*. Society for the Advancement of Excellence Education.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychologist*, 37(1), 8–25. <https://doi.org/10.1080/01443410.2015.1136407>
- Robson, C. (2011). *Real world research: A resource for users of social research methods in applied settings* (2nd ed.). John Wiley and Sons Ltd.
- Shermis, M. D., & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and new directions*. Routledge.
- Shermis, M. D., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Lawrence Erlbaum Associates Inc. (Eds.).
- Shermis, M. D., & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and new directions*. Routledge.
- Steinhart, D. (2001). *Summary street: An intelligent tutoring system for improving student writing through the use of latent semantic analysis* (Publication No. 3005103). [Doctoral dissertation, University of Colorado]. ProQuest Dissertations and Theses Global.
- Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers & Composition*, 42, 1–16.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296.
- The Grammarly Editor. (2019). *The Grammarly editor: New look, more suggestions, better writing*. <https://www.grammarly.com/blog/better-writing-with-grammarly/>
- Thi, N. K., & Nikolov, M. (2022). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher*, 31, 767–779. <https://doi.org/10.1007/s40299-021-00625-2>
- Tso, W. B., & Ho, S. Y. (2016). April 25). *Chances and challenges: Teaching academic writing to university students* [Conference session]. Seminar at Language Centre-Teaching & Learning, Hong Kong Baptist University, Hong Kong. <http://hdl.handle.net/10722/226274>
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147–170.
- Wade-Stein, D., & Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, 22(3), 333–362. https://doi.org/10.1207/s1532690xc2203_3
- Wang, Y. J., Shang, H. F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257. <https://doi.org/10.1080/09588221.2012.655300>
- 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>
- Weissbach, R. S., & Pflueger, R. C. (2018). Collaborating with writing centers on interdisciplinary peer tutor training to improve writing support for engineering students. *IEEE Transactions on Professional Communication*, 61(2), 206–220. <https://doi.org/10.1109/tpc.2017.2778949>
- Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, 49(4), 461–493. <https://doi.org/10.1017/s0261444816000161>
- 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>