

# Student engagement with automated written corrective feedback (AWCF) provided by *Grammarly*: A multiple case study

Svetlana Koltovskaia

*TESOL and Applied Linguistics, English Department, Oklahoma State University, United States*

## ARTICLE INFO

### Keywords:

Automated writing evaluation (AWE)  
Automated written corrective feedback (AWCF)  
Student engagement  
*Grammarly*  
L2 writing

## ABSTRACT

Despite the increased use of automated writing evaluation (AWE) systems and similar programs for assessment purposes in second language (L2) writing classrooms, research on student engagement with automated feedback is scarce. This naturalistic case study explored two ESL college students' engagement with automated written corrective feedback (AWCF) provided by *Grammarly* when revising a final draft. Following previous research, student engagement was operationalized using three interconnected dimensions: behavioral, cognitive, and affective. Behavioral engagement was explored through the analysis of QuickTime-based screencasts of students' *Grammarly* usage. Cognitive and affective engagement were measured through the analysis of students' comments during stimulated recall of the aforementioned screencasts and semi-structured interview. Findings suggest that students had different levels of engagement with AWCF. One showed greater cognitive engagement through his questioning of AWCF. However, he did little to verify the accuracy of feedback which resulted in moderate changes to his draft. The other's overreliance on AWCF indicated more limited cognitive engagement which led to feedback's blind acceptance. Nevertheless, this also resulted in moderate changes to her draft.

## 1. Introduction

The use of automated writing evaluation (AWE) systems and similar tools for assessment purposes in second language (L2) writing classrooms has rapidly increased due to their numerous advantages. AWE systems have been claimed to provide computer-generated quantitative and qualitative feedback (Dikli, 2006). They have been considered to offer multiple practice and revision opportunities (Warschauer & Ware, 2006) and be more consistent and objective than human raters (Wang, Shang, & Briody, 2012). Especially, AWE programs have been praised for their capacity to free up teachers' time to focus less on lower-order concerns (e.g., grammar, mechanics) and more on higher-order concerns (e.g., content, organization) (Ranalli, 2018) and other aspects of writing instruction (Warschauer & Grimes, 2008).

Notwithstanding AWE benefits, previous studies suggest that students do not make good use of automated feedback (Attali, 2004; Chapelle, Cotos, & Lee, 2015; Warschauer & Grimes, 2008). The bulk of research, however, has mainly focused on students' final written products rather than their revision process (Stevenson & Phakiti, 2014). Stevenson and Phakiti (2019) noted that this focus on product over process tells us little about to what extent L2 learners developed metacognitive skills to notice, evaluate, and consequently improve writing. Therefore, studies are needed on student engagement with automated feedback during the revision process to help understand the benefits of such feedback. Despite its importance, research on student engagement with automated feedback is

**Abbreviations:** AWE, automated writing evaluation; AWCF, automated written corrective feedback; CF, corrective feedback; L2, second language; WCF, written corrective feedback

**E-mail address:** [skoltov@okstate.edu](mailto:skoltov@okstate.edu).

<https://doi.org/10.1016/j.asw.2020.100450>

Received 23 September 2019; Received in revised form 19 February 2020; Accepted 19 February 2020

Available online 19 March 2020

1075-2935/ © 2020 Elsevier Inc. All rights reserved.

surprisingly scarce.

Motivated by a lack of systematic research on this subject, this case study has explored ESL university students' engagement with automated written corrective feedback (AWCF) provided by *Grammarly* when revising a final draft. This study offers new insight into the process through which students engage with AWCF by presenting a holistic narrative of two cases.

## 2. Literature review

### 2.1. Research on AWE feedback

The two central components of AWE are a scoring engine that generates automated scores and a feedback engine that provides automated written feedback (Bai & Hu, 2017), also known as automated written corrective feedback (AWCF) (Ranalli, 2018). AWE originated from automated essay scoring (AES) and was initially used in high-stakes testing to generate numeric scores for summative assessment (assessment of learning) based on such techniques as artificial intelligence, natural-language processing, and latent semantic analysis (Cotos, 2014; Stevenson & Phakiti, 2019). The most notable and cited among AES systems are *Project Essay Grade* (PEG) developed by Ellis Page, *Intelligent Essay Assessor™* (IEA) from Pearson Educational Technologies, *Electronic Essay Rater* (e-rater®) from Educational Testing Services, and *IntelliMetric®* from Vantage Learning. In recent years, various AWE systems, including *Criterion®* from Educational Testing Service, *My Access!* from Vantage Learning, *WriteToLearn* from Pearson Educational Technologies, and others have been developed not only for summative but also formative assessment (assessment for learning) purposes to be used in writing classrooms (Chen & Cheng, 2008). While earlier AWE research largely focused on the validity and reliability of its scoring system in testing contexts (Attali & Burstein, 2006; Burstein & Chodorow, 1999; Burstein et al., 1998; Elliot, 2002; Enright & Quinlan, 2010; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002), recent studies have addressed instructional use of AWE. These classroom-based studies have explored student perceptions of the usefulness of AWE quantitative and qualitative feedback (Chen & Cheng, 2008; Dikli & Bleyle, 2014; Grimes & Warschauer, 2010; Lai, 2010) and investigated the effects of automated feedback on writing (Attali, 2004; Chapelle et al., 2015; El-Ebyary & Windeatt, 2010; Li, Link, & Hegelheimer, 2015; Li, Feng, & Saricaoglu, 2017; Liao, 2015).

Student perception studies of AWE feedback report ambivalent findings. In their comparative study, Dikli and Bleyle (2014) found that ESL students generally perceived *Criterion®* feedback to be helpful while, at the same time, valuing instructor feedback. In Grimes and Warschauer's (2010) study, the U.S. middle school students rated *My Access!* favorably for its usefulness, fairness, and user-friendliness. The students reported that the system motivated them to write and revise their papers and increased their confidence. Conversely, Chen and Cheng (2008) found that Taiwanese EFL students perceived the use of *My Access!* unfavorably at large, which the authors noted was attributable to limitations inherent in the system's assessment and assistance functions. However, the students identified the use of the system positively if it was utilized with the instructor facilitation. Similarly, in her comparative study, Lai (2010) found that Taiwanese EFL students mostly held negative perceptions of *My Access!* feedback because it was too general for them to make revisions, thus preferring peer evaluation over automated feedback.

The results of the previous research on the effects of AWE feedback on writing are also mixed. Some studies report positive effects on writing (El-Ebyary & Windeatt, 2010; Li et al., 2015, 2017; Liao, 2015; Wang et al., 2012). For example, El-Ebyary and Windeatt (2010) found that *Criterion®* feedback positively impacted on the quality of Egyptian EFL students' writing although some students achieved better scores by using the avoidance strategy. The authors also reported that unlike conventional writing/feedback modes, *Criterion®* encouraged students to revise their essays (100 % resubmission rate). Likewise, regarding the effects of AWE feedback on draft revisions, Li et al. (2015) found that *Criterion®* led to increased revisions, and its feedback helped ESL students improve their linguistic accuracy. Conversely, Attali (2004) found that 71 % of the U.S. sixth to twelve grade participants submitted their essays to *Criterion®* once, indicating that most students did not utilize the revision capabilities of the system. Other studies also report discouraging findings that the majority of students who submitted their drafts for scoring submitted them only once and made limited revisions upon receiving automated feedback (Warschauer & Grimes, 2008; Warschauer & Ware, 2006). Chapelle et al. (2015) found that ESL students disregarded nearly 50 % of *Criterion®* feedback despite the provision of both direct and indirect feedback, thus making limited changes to their drafts. The authors suggested this was due to inaccuracies in *Criterion®* feedback.

Since the findings of the aforementioned studies on automated feedback are contrasting, it is difficult to definitively conclude whether students make the most of it on their writing. Zhang (2017) claimed that to benefit from feedback, students need to be effectively engaged with it. Student engagement with feedback is also believed to be a key factor in the success of writing development and language acquisition (Zhang & Hyland, 2018). While previous studies on AWE have provided insight into students' perceptions of automated feedback and how students utilize it to revise their texts as it pertains to revision operations (e.g., accept feedback) and times of submission, the majority of these studies have focused on students' written products, with little attention paid to their revision process (Stevenson & Phakiti, 2019). According to Zhang (2017), without careful investigation of how students engage with automated feedback during the revision process, it is impossible to know what factors facilitate or inhibit their response to such feedback.

### 2.2. Construct of student engagement and empirical research

Perhaps the most well-known conceptualization of engagement was proposed by Fredricks, Blumenfeld, and Paris (2004). They viewed engagement as a multifaceted construct which encompasses three interrelated dimensions: behavioral, emotional (affective), and cognitive. However, their conceptualization of engagement was proposed for school engagement. Ellis (2010) applied Fredricks

et al.'s tripartite conceptualization to student engagement with both oral and written corrective feedback (CF) in which *behavioral perspective* concerned learners' uptake and revisions elicited by CF, *affective perspective* referred to learners' attitudinal response to CF, and *cognitive perspective* involved "how learners attend to the CF they receive" (p. 342). Han and Hyland (2015) furthered Ellis's framework to explore four Chinese college students' engagement with teacher written corrective feedback (WCF). In their study, *behavioral engagement* involved revision operations in response to WCF and observable strategies used in improving the accuracy of drafts, future writing, and/or L2 competence. *Cognitive engagement* referred to the depth of processing of WCF encompassing cognitive and metacognitive operations. *Affective engagement* concerned learners' immediate emotional reactions and attitudinal responses toward WCF. The results showed that although the students received similar in terms of scope, type, and frequency teacher WCF, they engaged with it differently due to individual differences and contextual factors. The authors concluded that students, as active agents of their own learning, can decide how and what they learn from teacher WCF. Zheng and Yu (2018) applied the developed framework to their study on low proficiency university students' engagement with teacher WCF. The researchers found that while the students' affective engagement with teacher WCF was relatively positive, their behavioral and cognitive engagement with WCF was at a limited level as it was negatively impacted by their low English proficiency.

Zhang and Hyland (2018) further strengthened the framework to investigate two Chinese university students' engagement with both teacher WCF and AWE feedback provided by the Chinese AWE system, *Pigai*. They viewed *behavioral engagement* as students' behavioral reaction to feedback, including revision actions and time spent on revision. *Affective engagement* involved students' emotional responses and attitudinal reactions to feedback, while *cognitive engagement* concerned how students attend to feedback and their use of revision operations and cognitive (metacognitive) strategies. The results showed that the highly engaged student preferred AWE feedback over teacher feedback because the former provided immediate feedback and allowed her to resubmit her essay 13 times; thus promoting her autonomy. Conversely, the moderately engaged student had limited engagement particularly with AWE feedback because he was overwhelmed by the amount of feedback provided and felt embarrassed and demoralized by the low score he received. Zhang (2017) also focused on a Chinese university student engagement with *Pigai* feedback. *Behavioral engagement* in his study, however, referred to the number of submissions and time spent on revisions. *Emotional engagement* involved affective reactions and motivational changes. *Cognitive engagement* concerned understanding the feedback information, monitoring the revision process, and self-regulating. The results revealed the student was engaged with *Pigai* feedback behaviorally, emotionally, and cognitively. The author noted that when engaging in multiple revisions, the student felt motivated by the prospect of getting higher holistic scores and felt demotivated when multiple revisions resulted in low scores.

In line with previous research, engagement with AWCF, in this study, is also seen as composed of three interrelated dimensions where:

- *Behavioral engagement* concerns revision operations, i.e. actual revisions carried out, revisions strategies used to improve the accuracy of the draft, and time spent on revision.
- *Cognitive engagement* concerns how deeply students process AWCF (noticing or understanding) and their use of metacognitive and cognitive operations.
- *Affective engagement* concerns students' immediate emotional reactions and attitudinal responses to AWCF.

While the aforementioned studies shed light on student engagement with teacher and AWE feedback, more studies are needed to explore student engagement with automated feedback to "unlock" its benefits (Zhang & Hyland, 2018, p. 90). Particularly, studies need to focus on student engagement with *Grammarly* feedback. As Ranalli (2018) noted, *Grammarly* is making important inroads into L2 classrooms for its capacity to provide more specific feedback. Besides, research on *Grammarly*, generally, report positive results which suggests its use in writing classrooms is worth considering. For example, O'Neill and Russell (2019) investigated students' perceptions of *Grammarly* when it is used together with academic learning advisor (ALA) feedback. They found the group that received *Grammarly* feedback along with ALA feedback was significantly more satisfied than the group that received only ALA feedback. The participants reported liking *Grammarly* feedback because it was detailed, thorough, line-by-line, and prompt. Qassemzadeh and Soleimani (2016) explored the impact of *Grammarly* and teacher feedback on learning passive structures. They found that both *Grammarly* and teacher feedback can positively influence learning of passive structures. However, the role of the former in retaining passive structures is more highlighted than the latter. It is noteworthy that unlike many AWE programs such as *Criterion* or *My Access!*, which provide both numeric scores and AWCF, *Grammarly* provides only AWCF. Additionally, *Criterion* and many similar AWE systems are (J. Ranalli, personal communication, March 25, 2019; Ranalli & Yamashita, 2019). Because of the nature and timing of the feedback *Grammarly* provides, it has recently been termed an AWCF tool (Ranalli, 2018; Ranalli & Yamashita, 2019) rather than AWE. In this study, *Grammarly* will be referred to as an AWCF tool. While research on *Grammarly* has focused on students' perceptions of *Grammarly* feedback and its effectiveness in retaining certain grammatical structures, no work has been done on how ESL students engage with AWCF provided by *Grammarly*. Therefore, the current case study employs the multi-dimensional framework of student engagement to answer the following research question:

- 1 How do students behaviorally, cognitively, and affectively engage with AWCF provided by *Grammarly* when revising their final draft?

**Table 1**  
Participants' profiles.

Name	Age	Country	First language	Major	Class standing
Alex	21 (1998)	Hong Kong	Cantonese, Mandarin	Journalism	Junior
Kelsey	22 (1997)	Saudi Arabia	Arabic	Computer Science	Freshman

### 3. Methodology

#### 3.1. Research overview

The research design used in this study was a case study, which provides an in-depth, holistic, and contextualized understanding of the phenomenon under investigation (Yin, 2009). In particular, a multiple-case study was employed to explore how two students engage with AWCF.

#### 3.2. Participants and classroom

The study took place at a large southcentral university in the U.S. Seventeen undergraduate students enrolled in the International Freshman Second Language Writing course (ENGL 1223) during the fall 2018 semester were recruited for two reasons: 1) the students were L2 learners of English and 2) the students took a writing course in which they were required to produce a multiple-draft assignment. Eight out of 17 students volunteered to participate of which only two adequately completed all aspects of the study. The two participants were Alex and Kelsey (pseudonyms). Table 1 shows the participants' profiles outlining their demographic information along with their major and class standing.

At the beginning of the semester, the students had an in-class diagnostic assessment for which they were required to write a summary of an article (see Appendix A for a diagnostic writing prompt). The researcher and two of her colleagues independently evaluated the two participants' texts using the slightly modified TOEFL iBT Test - Independent Writing Rubric - to determine their language proficiency and writing skills. The texts were evaluated based on a scoring rubric of 0-5. The mean rubric score given by the raters for the quality of the students' writing was then converted to a scaled score of 0-30. Alex received a scaled score of 25 (the rubric score of 4), which means he is an advanced L2 writer, while Kelsey received a scaled score of 14 (the rubric score of 2.7), which means she is a low-intermediate L2 writer (see <https://www.ets.org/toefl> for more details). Although it was Alex's first semester in the U.S., he was one of the best students in the class. Unlike his classmates, his writing skills were better developed. The level of Kelsey's writing skills was average and thus similar to that of the students taking this course.

ENGL 1223 was a 16-week, three-credit research writing course. It was restricted to undergraduate students whose native language was not English. The class met three times a week for 50 min in each session. The major assignments of the course were an annotated bibliography, a research proposal (introduction, literature review, methodology), and a research proposal presentation. The researcher was a teacher of the course. The researcher is a non-native speaker of English from Russia, and she was in her third-year of Ph.D. studies in TESOL and Applied Linguistics at the time of the study.

Creswell (2014) noted that in qualitative research, the inquirer, as the primary data collection instrument, should explicitly state his/her role as this may shape the direction of the study. The researcher's four years of teaching experience and five years of Writing Center work in the U.S. influenced the way she teaches writing at a college level and provides feedback. She believes teachers should give feedback on higher-order concerns at the early stages of writing and lower-order concerns at the last stage of writing. However, often due to time constraints, she provides feedback predominantly on global issues, thus leaving grammar and mechanics for students to revise on their own. Therefore, she believes that as a complement to teacher feedback, AWE systems could empower students to revise their own work because they are more computationally adept at providing feedback on low-order concerns (Ranalli, Link, & Chukharev-Hudilainen, 2017), particularly the use of the easily accessible *Grammarly*.

#### 3.3. Grammarly

In this study, a free version of *Grammarly* (<https://app.grammarly.com/>) was utilized. The free version of *Grammarly* provides feedback on spelling, punctuation, grammar, and conventions, including spacing, capitalization, and dialect-specific spelling. *Grammarly* instantly provides feedback for improvement once a paper is uploaded online. The uploaded paper appears on the left side of the screen with errors underlined in red (i.e. indirect feedback) while direct feedback appears on the right side of the screen (see Appendix B). Direct feedback contains the error type (e.g., grammar), possible error correction (e.g., Korean Peninsula -> the Korean Peninsula), and suggestion (e.g., It appears that an article is missing before the word **Korean**. Consider adding the article.). Suggestions can be expanded for a comprehensive explanation of a grammar rule, i.e. a metalinguistic explanation (see Appendix C).

#### 3.4. Three-stage revision process

For this study, the students worked on a literature review. The literature review was a multi-draft assignment for which the students were expected to use four peer-reviewed scholarly articles relevant to their research topic of interest. The information sheet

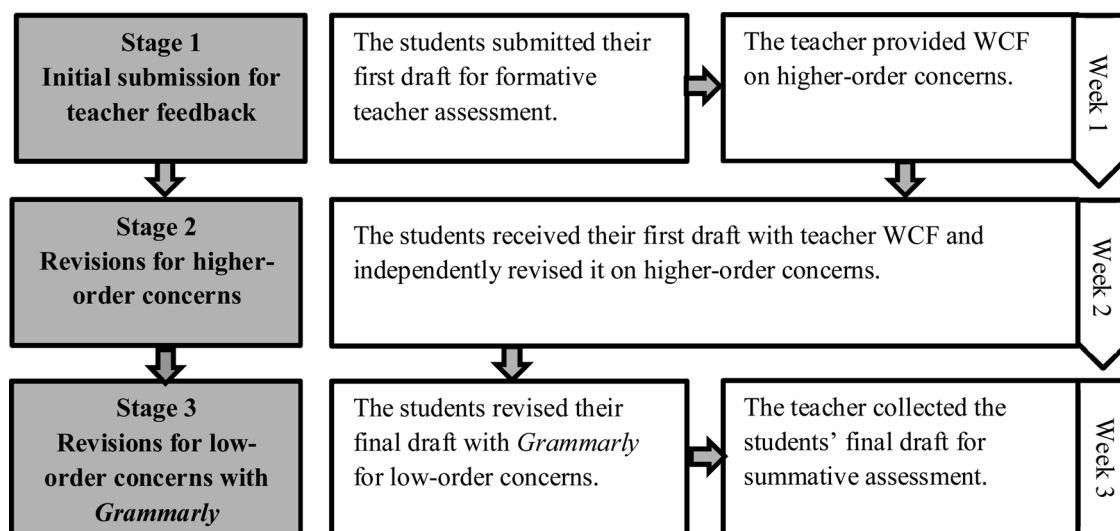


Fig. 1. The three-stage revision process of the literature review assignment.

of what the literature review is and the rubric are shown in Appendix D and Appendix E, respectively. For this study, the literature review assignment involved three stages (see Fig. 1).

#### 3.4.1. Stage 1-Initial submission for teacher feedback

At the beginning of the first week, the students uploaded the first draft of their literature review to a Dropbox in the D2L Brightspace Learning Environment to receive formative teacher assessment. When writing their first draft, the students were asked to focus primarily on higher-order concerns.

#### 3.4.2. Stage 2-Revisions for higher-order concerns

In the middle of the second week, the students received their first draft with teacher WCF on higher-order concerns. They were asked to independently revise their papers based on teacher WCF by the end of the second week. It is important to note that in Stage 1 and Stage 2, the students might have addressed low-order concerns. However, any revisions on low-order concerns that were completed at this time were not considered due to the study's emphasis on revision using *Grammarly*.

#### 3.4.3. Stage 3- Revisions for low-order concerns with Grammarly

In the middle of the third week, the students worked in the computer classroom on revising their final draft with *Grammarly*. At the end of the class, the students submitted their final draft for summative teacher assessment. At this stage, formal data collection was employed.

#### 3.5. Data collection

Data triangulation was included in the research design. The research data were collected by means of screencasts, stimulated recall, and semi-structured interview.

While revising their final draft with *Grammarly*, the students were asked to record their revision process with QuickTime player. Prior to that, the students were trained on how to use the QuickTime player screencast function and *Grammarly*. This function allowed the students to record video of their computer screen and capture their revision process. The revision process lasted one class period. The students' screencasts were collected at the end of the class.

The students then had an individual stimulated recall session in which they watched their screencast video on a TV monitor and were asked to recall their thoughts at the time of correction of each error for which they received AWCF. According to Gass and Mackey (2000), a recall should occur as soon as possible after the event. In this study, the recall occurred within 48 h of the students' error correction activity. The recall script and guiding questions can be seen in Appendix F. This introspective method was used to access students' thoughts as they were carrying out the revision activity (Gass & Mackey, 2000) to explore their cognitive, affective, and behavioral engagement with AWCF.

After the recall, the students had a semi-structured interview consisting of ten questions to explore their affective engagement with *Grammarly* and its feedback (see Appendix G). Additional follow-up questions were asked to build up a more complete picture. Before the interview, a pilot test was conducted with a volunteer student from the other section of the same course to make necessary modifications to the questions. Some adjustments were made, including clarification of ambiguous questions and simplification of the language. The entire reflective exercise (stimulated recall and semi-structured interview) lasted roughly an hour. It was conducted in English and audio-recorded.



### 3.6. Data analysis

#### 3.6.1. Analysis of screencasts

Behavioral engagement was explored through the analysis of the two students' screencasts. The analysis consisted of five phases. The first phase included reclassification of the error types detected by *Grammarly*. Although *Grammarly* automatically categorizes all errors into four error types (spelling, grammar, punctuation, and conventions), the error type *grammar* is too broad. Besides, its classification of the error types raises doubts. For instance, *Grammarly* suggested a participant changing the word *educational* to *education* and classified this error as a spelling error. However, the error has to do with word form rather than spelling. Therefore, to address the ambiguity of error categorization in *Grammarly*, the errors for which the participants received AWCf were categorized according to Han and Hyland (2015)'s taxonomy of error categories.

Additionally, because automated feedback is prone to be fallible (Chapelle et al., 2015), and this can ultimately affect student engagement with such feedback, the second phase involved diagnosis of the accuracy of AWCf the students received. For example, *Grammarly* suggested a student adding the indefinite article before *little* in the sentence: *Although there is many research that focus on the effect of insomnia on college students, only little literature investigated the relationship between insomnia and the college student's sleep hygiene in X University*. This AWCf was identified as inaccurate because adding *a* before *little* changes the intended meaning of the sentence, which was to show the gap by emphasizing that almost no literature exists on the topic (see Appendix H).

In the third phase, revision operations were identified. Revision operations were operationalized as any actions taken in response to AWCf. After repeatedly going through the screencasts, three categories of revision operations were determined: accept, reject, and substitute (see Appendix I). If the correction suggested by *Grammarly* was clicked by the student to automatically fix an error, this was regarded as feedback being accepted. If the correction was left unclicked or dismissed and consequently no changes were made to the text, this was viewed as feedback being rejected. If the correction was substituted by the student's own correction to address the error, this was considered as feedback being substituted. To enhance the validity in data analysis, the inter-coder agreement was calculated after a second coder, a trained colleague, independently coded 50 % of the data. The agreement rates for error categorization, accuracy/inaccuracy of AWCf, and revision operations were 96.1 %, 100 %, and 100 %, respectively.

The fourth phase involved identification of revision strategies. Any strategies taken to enhance the quality of the draft in response to AWCf were identified as revision strategies such as consulting the Internet or dictionary to verify the accuracy of AWCf. The stimulated recall data was further analyzed to reaffirm any observed strategies.

In the fifth phase, the amount of time spent by the students on draft revision was determined by first counting how much time they spent on each error for which AWCf was given. Then, the total time of working on all errors in the draft was calculated for each student. Simple statistical calculations were made for all quantitative data.

#### 3.6.2. Analysis of the recall and interview data

The students' recall and interview data were qualitatively analyzed to profile their behavioral, cognitive, and affective engagement with AWCf. Prior to coding, audio recordings of the recall and the interview were transcribed with Trint (<https://trint.com>) and then checked for accuracy against the original recordings. Then, all transcripts were organized by an individual participant. To further sort and reduce data, each participant's recall transcript was segmented into language related episodes (LREs), and the researcher's questions and the participant's responses were placed into two different columns in Excel: the researcher's questions in the first column preceded the participant's responses in the second column (see Table 2). LRE was operationalized as any segment of the recall in which there is an explicit focus on a linguistic item (Swain & Lapkin, 1995), such as that one LRE corresponds to one error for which AWCf was given. The interrater agreement for identifying LREs was 100 %.

Only the students' responses from the recall and the interview transcripts were coded following three steps: open coding, axial coding, and selective coding as described in Corbin and Strauss (2008). Initial codes were closely related to the original data in the recall and the interview. For example, a comment such as *I decided to put it aside and I reselected after this, maybe, easier ones* was assigned the initial code of 'fixing easy errors.' A memo was written next to the code: *the student recognized that some errors cause more difficulties than others. Thus, he first dealt with errors that were easy for him to fix*. Informed by previous studies that offered insight into the labeling and categorizing of student engagement (e.g., Han & Hyland, 2015; Zhang & Hyland, 2018), this code was further refined to 'planning for cognition' in the axial coding, which is one type of metacognitive operation. This code then was attributed to the 'cognitive engagement' category in the selective coding. The codes were then compared across the two cases. After that, the case narratives were constructed. The same colleague again independently coded 50 % of the data to enhance the validity. While the inter-

**Table 2**

Example of Alex's data layout and coding sheet for stimulated recall.

LRE 13	Researcher	Alex
AWCF: Heavily relied; Has heavily relied	So, when you were looking at this (AWCF) and then kind of decided to move on to the next one, why did you decide to move on to the next one?	Because I can't think of which one I should use. I'm not sure (UNSURE <sup>a</sup> ). So, therefore, I decided to put it aside and I reselected after this, maybe, easier ones (FIXING EASY ERRORS <sup>a</sup> ).

<sup>a</sup> Initial codes assigned to the participant's comments during the open coding stage.

**Table 3**  
Alex's revision operations.

Error type	AWCF	Accurate AWCF			Inaccurate AWCF		
		Accept	Reject	Substitute	Accept	Reject	Substitute
Verb form	1		1				
Word form	1	1					
Articles	5	2	3				
Punctuation	2		2				
Spelling	1	1					
Prepositions	3	3					
Miscellaneous <sup>a</sup>	1	1					
Total number of errors in a 914-word text	14	8	6				
Percentage	100 %	57 %	43 %				

<sup>a</sup> Miscellaneous here refers to a spacing error.

coder agreement for behavioral and affective engagement reached 91.2 % and 93.4 %, respectively, the agreement rate for cognitive engagement was initially 67.4 %. After extensive discussion and going through data multiple times, the agreement rate for cognitive engagement reached 82.7 %.

## 4. Findings

### 4.1. Alex - advanced L2 writer

#### 4.1.1. Alex's behavioral engagement with AWCF

Behavioral engagement with AWCF involved revision operations and strategies used to enhance the accuracy of the draft and time spent on revision. Table 3 illustrates Alex's revision operations in response to AWCF. The table shows Alex made 14 errors in his 914-word draft for which he received AWCF. Seven error types were identified in his draft, including errors on the use of verb form (1), word form (1), articles (5), punctuation (2), spelling (1), prepositions (3), and spacing (1). Of the seven error types, articles were the most frequent errors identified by *Grammarly*. Of 14 received AWCF, which were all accurate, Alex correctly accepted eight and incorrectly rejected six; thus fixing 57 % of his total errors. This suggests Alex made moderate changes to his draft.

Regarding revision strategies, Alex once sought extra assistance from the Internet to verify the accuracy of AWCF. To illustrate, *Grammarly* suggested Alex using the preposition *in* instead of *on* in the sentence *They showed that the media plays an important role on this international relationship*. The screencast video showed how Alex typed 'play an important role' in the Google search engine to see if this phrase often appears with *in* or *on*. Alex scrolled down to see examples, and once he realized the correct preposition is *in*, he accepted the feedback. As for time spent on making necessary revisions in his draft, Alex spent a little over five minutes (333 s) on this despite having a 50-min class period.

#### 4.1.2. Alex's cognitive engagement with AWCF

Cognitive engagement concerned how deeply the students processed AWCF (noticing and understanding), their use of meta-cognitive operations to regulate their mental effort, and cognitive operations to process feedback and determine appropriate revisions. Regarding cognitive engagement with AWCF at the level of noticing, in the recall, Alex reported detecting all AWCF. He said feedback was conspicuous because errors were underlined in red and corrections were highlighted in green. Since Alex read *Grammarly* suggestions and because of AWCF's explicit and implicit nature, he said he relatively easily and quickly recognized the corrective intention of the majority of feedback.

In terms of cognitive engagement with AWCF at the level of understanding, the recall revealed Alex understood the cause/nature of eight errors and how to correct those errors. Therefore, Alex accepted eight AWCF on those errors. The following example demonstrates Alex's understanding of the error:

[Stimulated recall]		
LRE 12	Alex's text before revision	Alex's comments
AWCF	They pointed out that as most of the Americans did not have direct	It is North American English speaking. In 'judgment' I have 'e',
Judgements	experience about China, how they do their <u>judgements</u> on China was	'judgements.' So, that's why I'm thinking, oh, I am here in America.
-> judgements	heavily relied on the information covered from news media.	So, I should use the American version.

Alex appeared not to know the grammar rules of the six errors that were left uncorrected. For example, Alex incorrectly placed a comma in a compound object. In the recall, Alex reported reading *Grammarly* suggestion saying: *It appears that you have an unnecessary comma in a compound object. Consider removing it*. Despite the suggestion, Alex decided not to correct the error explaining this as follows:

---

[Stimulated recall]		
LRE 10	Alex's text before revision	Alex's comments
AWCF Opinion,	He tried to answer these questions by addressing the way the issue of a rising China affects public opinion, and the role of publicity in shaping public perceptions of US-China relations.	Because it should be asking me to delete the comma, and, I think, the sentence is too long and the comma is appropriate.

---

This example demonstrates Alex's lack of knowledge on the use of commas. If Alex had expended the suggestion for a meta-linguistic explanation on comma usage, this could have facilitated understanding of the error he incorrectly rejected. Instead he relied on his intuition, thus deploying the wrong revision operation. Overall, Alex's cognitive engagement was relatively extensive which was manifested in the use of several metacognitive and cognitive operations to regulate his mental processes. The metacognitive operation that helped Alex regulate his mental effort was planning for cognition:

---

[Stimulated recall]		
LRE 13	Alex's text before revision	Alex's comments
AWCF Heavily re- lied; has he- avily relied	They pointed out that as most of the Americans did not have direct experience about China, how they do their judgements on China <u>was heavily</u> <u>relied</u> on the information covered from news media.	Because I can't think of which one I should use. I'm not sure. So, therefore, I decided to put it aside, and I reselected after this, maybe, easier ones.

---

This operation demonstrates Alex was strategic. He first determined the difficulty level of an error. Then, he addressed easier errors, thus saving difficult errors for later. Another metacognitive operation Alex deployed was evaluating the outcome of the task by double- or triple-checking the revision operation he used. He often went back to the errors he accepted/rejected and read the entire sentence several times to ensure his decision was right. As for the use of cognitive operations, the recall revealed that to process feedback and figure out the appropriate revision operation, Alex reasoned:

---

[Stimulated recall]		
LRE 1	Alex's text before revision	Alex's comments
AWCF China - US -> China- US	Chen & Garcia (2016) examined in the perspective of US media behavior, and how they can influence the <u>China - US</u> relations.	Because I'm thinking what's the difference. There's a hyphen also for the 'China-US relations.' But the hyphen is maybe not using the method that <i>Grammarly</i> uses.

---

Additionally, Alex used context to determine the appropriate revision operation. He read the sentence where the error was detected to decide if the correction suggested by *Grammarly* worked in that sentence. Sometimes reading the sentence where the error was detected was not enough; thus, Alex needed more context to decide whether to accept or reject AWCF. In the recall, he stated, "I will first read a sentence with the error. If it is still not clear, I will read more sentences before the sentence with the error to check if this is correct or not."

#### 4.1.3. Alex's affective engagement with AWCF

Affective engagement referred to the students' emotional reactions toward AWCF upon receiving it and their overall attitude toward AWCF. Alex's emotional reaction toward AWCF was often distrust. Alex questioned AWCF because he appeared not to find it as authoritative as his teacher's feedback. Although this was the first time Alex used *Grammarly*, he knew such computer-generated feedback can be occasionally inaccurate. In the interview, Alex reported, "[...], sometimes I was not 100 % sure about what *Grammarly* recommended me. So, [...] I have to check it again and proofread whether feedback is correct or not."

This emotional reaction of doubt toward AWCF affected Alex's cognitive and behavioral engagement, which indicates that the three dimensions of engagement are linked to one another. Alex was quite extensively cognitively engaged with AWCF because he questioned its accuracy. However, instead of seeking extra assistance from other resources to verify the feedback's accuracy, Alex relied on his linguistic knowledge and intuition. His lack of knowledge of certain grammar and mechanics rules, however, resulted in rejecting six out of 14 accurate feedback.

Overall, Alex had a positive attitude toward AWCF and *Grammarly*. He found the tool to be helpful because, as he said in the interview, it helped him find mistakes he could not find on his own. According to Alex, *Grammarly* also "tells a little bit about the reason behind the error," and it is easy to operate. Alex said he would consider using *Grammarly* in the future for the following reason: "In the whole perspective, it is useful. At least, I can have more resources to proofread my paper" (the interview).

## 4.2. Kelsey - low-intermediate L2 writer

### 4.2.1. Kelsey's behavioral engagement with AWCF

Table 4 illustrates Kelsey's revision operations in response to AWCF. The table shows Kelsey made 26 errors in her 810-word draft for which she received AWCF. Ten error types were identified in her draft, including errors on the use of word choice (1), verb form



**Table 4**  
Kelsey's revision operations.

Error type	AWCF	Accurate AWCF			Inaccurate AWCF		
		Accept	Reject	Substitute	Accept	Reject	Substitute
Word choice	1		1				
Verb form	1	1					
Word form	3		1		1	1	
Articles	8	3		1	4		
Pronouns	1				1		
Punctuation	1	1					
Spelling	2	2					
Subject-verb agreement	4	3			1		
Prepositions	2	2					
Miscellaneous <sup>a</sup>	3	3					
Total number of errors in an 810-word text	26	15	2	1	7	1	
Percentage	100 %	57 %	8%	4%	27 %	4%	

<sup>a</sup> Miscellaneous here refers to a spacing error.

(1), word form (3), articles (8), pronouns (1), punctuation (1), spelling (2), subject-verb agreement (4), prepositions (2), and spacing (3). Like in Alex's case, the most frequent errors identified by *Grammarly* were articles. Of 26 received AWCF of which 18 were accurate and eight were inaccurate, Kelsey correctly accepted 15, incorrectly rejected two, incorrectly substituted one, incorrectly accepted seven, and correctly rejected one; thus also fixing 57 % of her total errors. This suggests Kelsey made moderate changes to her draft like Alex.

Unlike Alex, Kelsey did not refer to any external resources to enhance the accuracy of her draft. Similarly to Alex, Kelsey also spent a little over five minutes (323 s) on revision, despite having more errors in her draft than Alex.

#### 4.2.2. Kelsey's cognitive engagement with AWCF

Although Kelsey managed to detect all AWCF and recognize their corrective intention by just looking at *Grammarly* corrections as she reported in the recall, she did not always understand the cause/nature of the majority of errors, especially errors on article usage. To illustrate, *Grammarly* suggested Kelsey deleting *the* before *gender* in the sentence: *The results were analyzed according to the gender*. Kelsey uncritically accepted *Grammarly*'s accurate feedback. In the recall, she was unable to verbalize the underlying rule. Additionally, she admitted articles were her weakness and stated that when it comes to articles, "It's probably that I'm wrong, and *Grammarly*'s right." It is not surprising Kelsey accepted all accurate and inaccurate AWCF on articles except one. She appeared to substantially rely on AWCF on articles due to the lack of control over that form.

Interestingly, when *Grammarly* suggested Kelsey adding the definite article, she incorrectly substituted it with the indefinite article. The fact that Kelsey neither accepted the feedback nor rejected it could indicate she had some doubts about AWCF, but still trusted enough to consider *Grammarly* correction. Thus, prompted by AWCF and drawing on her intuition, she substituted the suggested definite article with indefinite:

LRE 20	Kelsey's text before revision	Kelsey's comments
AWCF The Bi- gger	However, data from bigger previous study were used for the second part.	When I read the sentence with 'the', it did not sound good. So I decided that I'm not going to accept that one but, probably, I need something else. So I thought it sounds better with 'a.'

Unlike Alex, Kelsey had limited cognitive engagement with AWCF which was displayed by the infrequent use of metacognitive and cognitive operations. Like Alex, the metacognitive operation Kelsey deployed at the very beginning of the revision process was planning for cognition. The screencast showed how Kelsey scrolled down to see "if there are a lot of mistakes" (the recall). By looking at how much feedback she received, it seems that Kelsey wanted to be mentally prepared for revision. Kelsey also double-checked the revision operation she used to ensure her decision was right; however, that happened only once. Regarding the use of cognitive operations, Kelsey also used context to determine the appropriate revision operation. However, unlike Alex, Kelsey did not read *Grammarly* suggestions nor she expanded them for a metalinguistic explanation, which indicates she did not take full advantage of *Grammarly* feedback to effectively respond to it.

Kelsey's minimal cognitive engagement was seen especially toward the end of the revision process as she uncritically started accepting AWCF. In the interview, she admitted: "At the beginning, I was thinking about every mistake. I used to read the sentence from the beginning, think about it a little bit but at the end, I just accepted."

#### 4.2.3. Kelsey's affective engagement with AWCF

Kelsey's emotional reaction upon receiving AWCF was surprise. She did not expect to have 27 errors in her final draft because, according to her, she addressed low-order concerns before submitting her draft. Thus when Kelsey saw how much AWCF she received, she immediately questioned feedback. In the recall, she said, "I was thinking if all of these are really mistakes or some of them

are just the computer thing that are mistakes but they're not.”

Once she started correcting errors, she experienced both positive and negative emotional reactions toward AWCF. For example, Kelsey liked AWCF when it confirmed her earlier doubts:

[Stimulated recall]		
LRE 10	Kelsey's text before revision	Kelsey's comments
AWCF for -> to	This study could lead for a better understanding of the sleep hygiene practices that college student could change on their rotten in order to decrease the insomnia severity.	When I was writing, I was asking myself if I have to use 'for' or 'to.' And then I decided to use 'for.' But then when I saw that <i>Grammarly</i> is telling me that I have to use 'to', I said "Oh! So I should have used 'to'."

Kelsey did not like AWCF when she recognized it was inaccurate. In the interview, she said, “If I know that the tool is wrong, I will not accept the feedback.” She also added, “If I am not sure, I will just accept feedback.” This demonstrates Kelsey's dependence on AWCF when she was unsure about how to correct errors.

Regarding Kelsey's attitudinal response toward AWCF, she found it to be useful. In the interview, she said she would use *Grammarly* in the future because it helped her correct some of her errors. She also noted, “[...] we all know technology sometimes makes mistakes. So, we should think about it before we make a decision.” This statement contradicts Kelsey's actions. She hardly ever critically thought about AWCF as she felt the need to eliminate all errors because, as she stated in the interview, “[she doesn't] want a teacher to look at [her] work and then immediately tell that [she's] international.”

## 5. Discussion and conclusion

Informed by the conceptual framework of student engagement with WCF and AWE feedback (Han & Hyland, 2015; Zhang & Hyland, 2018; Zhang, 2017; Zheng & Yu, 2018), this study focused on how two ESL college students behaviorally, cognitively, and affectively engaged with *Grammarly* feedback (AWCF) when revising their final draft. The findings offer insight into the complex process of student engagement with AWCF and provide implications for the use of automated tools for writing assessment in L2 classrooms.

The two students' behavioral engagement with AWCF involved revision operations, revision strategies, and time spent on revision. Both students focused on eliminating *Grammarly*-detected errors and corrected 57 % of their total errors, thus making moderate changes to their draft. This suggests the students did not effectively utilize *Grammarly* feedback to revise their final draft. This echoes previous studies' findings that students tend not to make good use of automated feedback (Attali, 2004; Chapelle et al., 2015; Warschauer & Grimes, 2008). Additionally, the two students barely used revision strategies to refine their draft, which indicates their behaviors remained at the surface level. Alex (advanced L2 writer) consulted the Internet once to verify the accuracy of AWCF. He primarily drew on his linguistic knowledge and intuition to correct errors. Because of his lack of knowledge about certain grammar and mechanics rules, however, he rejected accurate feedback. If Alex had used external resources or expanded *Grammarly* suggestion for a metalinguistic explanation, his behavioral engagement could have resulted in greater accuracy of the draft. Unlike Alex, Kelsey (low-intermediate L2 writer) did not utilize any external resources to enhance the accuracy of her draft nor did she read *Grammarly* suggestions and grammar rules. She appeared to substantially rely on AWCF and rarely thought critically about feedback. Nevertheless, this still resulted in moderate changes to her draft. This suggests that without cognitive engagement as in a multiple choice examination where a student can receive a score by randomly guessing the answer, behavioral engagement with AWCF alone can result in successful revisions if accurate feedback is accepted. However, mere behavioral engagement cannot lead to true learning. Regarding time spent on revision, both students spent just over five minutes. This is in line with Warden (2000), who found that L2 students spent an average of six minutes on draft revision after AWE feedback which was attributable to fewer errors. This could be the case in this study too. This could also be due to limited cognitive engagement with AWCF or automatic application of *Grammarly* correction with a single click that could have sped up the revision process. It is noteworthy, however, that little revision time does not necessarily indicate minimal cognitive engagement with AWCF. Although both students spent five minutes revising, Alex had more extensive cognitive engagement with AWCF than Kelsey despite having fewer errors than her.

From the cognitive perspective, which involved how deeply students processed AWCF (noticing and understanding) and their use of metacognitive and cognitive operations, both students noticed AWCF and recognized its corrective intention due to its implicit and explicit nature. However, unlike Alex who understood the cause/nature of the majority of errors and how to correct those errors, Kelsey had little awareness at the level of understanding. Her insufficient linguistic competence appeared to impact on her ability to effectively process feedback and determine appropriate revisions. Similar findings have also been uncovered in Zheng and Yu (2018). Additionally, unlike Alex, Kelsey used fewer metacognitive and cognitive operations which indicates her minimal cognitive engagement with AWCF. This consequently resulted in overreliance on feedback. Conversely, Alex had extensive cognitive engagement with AWCF which was manifested in the frequent use of metacognitive and cognitive operations. As a result, he could make independent judgments and selective incorporation of AWCF.

Regarding affective engagement with AWCF, which involved students' emotional reactions and attitudinal responses to feedback, both students experienced different emotional reactions. Alex questioned each AWCF he received but did little to verify its accuracy. He relied on his L2 knowledge and intuition to determine appropriate revisions which led to rejecting accurate AWCF. This suggests questioning of feedback is not enough; it is what happens after that. In contrast, Kelsey seemed to excessively depend on AWCF, especially when she lacked knowledge about target forms. Overall, however, the two students had a positive attitude toward

*Grammarly* feedback which corresponds with findings of previous research that students generally tend to appreciate automated feedback (Dikli & Bleyle, 2014; Li et al., 2015, 2017).

This study could extend previous research on student engagement with automated feedback in the following ways. Compared to Zhang (2017) who claimed automated feedback can have a positive impact on student writing if active behavioral, cognitive, and affective engagement are in place, the findings of this study revealed behavioral engagement with AWCF alone could potentially lead to successful revisions if accurate AWCF is accepted. However, behavioral engagement alone cannot lead to true learning. Additionally, the findings showed that while one participant's negative affective engagement with AWCF (questioning) positively impacted on his cognitive engagement, the other's positive affective engagement with AWCF (trust) resulted in limited cognitive engagement. Students' language proficiency could be the cause of this. As Zheng and Yu (2018) claimed, limited linguistic knowledge could prevent students from fully processing feedback and making further revisions. Thus students may exhibit overreliance on automated feedback. The findings of this study also suggest that students with higher language proficiency are likely to question AWCF, spend more time processing it, and make selective incorporation of it.

The findings can provide several implications to enhance student engagement with automated feedback and better utilize *Grammarly* for assessment purposes in L2 writing classrooms. Based on the study's findings, students with low language proficiency may not be able to utilize *Grammarly* effectively as their lack of linguistic competence can prevent them from adequately understanding AWCF. Therefore, the use of *Grammarly* is recommended for students with more advanced English proficiency. To benefit from *Grammarly* feedback, students should receive proper guidance and training on how to effectively engage with it. First, for successful affective engagement with AWCF, students should be made aware of its strengths and weaknesses. *Grammarly* feedback contains a suggestion that could be expanded for a metalinguistic explanation that students should be encouraged to read to help them make appropriate revisions. Students should also be informed about the inaccuracies of AWCF to avoid excessive dependence on it. Second, for effective cognitive engagement with AWCF students should be guided to question and analyze it critically. Finally, for productive behavioral engagement with AWCF, students should be advised to confirm feedback with other sources or perhaps with other students or a teacher.

*Grammarly* and similar automated programs could serve as a useful resource for writing assessment in L2 classrooms if active engagement is in place. Teachers could incorporate them into writing curriculum as a supplemental tool to facilitate low-order concerns of student writing development. To enhance L2 writers' independent and critical thinking, students' reflections on their use of automated feedback could become a writing assignment. This could prove useful for helping both students utilize automated feedback for self-assessment of their own writing more effectively and teachers estimate what is already working well and what still needs to be improved in terms of developing students' writing skills.

Some limitations in this study should be acknowledged. Caution should be made when generalizing the findings to different contexts and a wider student population as they were based on two students' engagement with AWCF. Additionally, only one draft was analyzed in the study, so development or changes in student engagement with AWCF over time were not investigated. Future research may also consider including survey questions addressing students' self-confidence as this may also affect their engagement with AWCF.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

I would like to thank the anonymous reviewers for their insightful and encouraging comments and the journal editors, Martin East and David Slomp, for their helpful suggestions on the manuscript. I am also very grateful for Dr. Stephanie Link's advisory feedback and all the members of my doctoral committee's support and constructive feedback throughout my doctoral studies. Finally, I thank my colleagues who helped me determine participants' language proficiency and writing skills, a colleague who helped me analyze and code my data, and my participants who made this study possible.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.asw.2020.100450>.

## References

- Attali, Y. (2004). Exploring the feedback and revision features of criterion. April Paper Presented at the National Council on Measurement in Education in San Diego, CA.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3), 1–30.
- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, 37(1), 67–81.
- Burstein, J., & Chodorow, M. (1999). Automated essay scoring for nonnative English speakers. June *Proceedings of the ACL99 Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing*.
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., et al. (1998). Automated scoring using a hybrid feature identification technique. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*.

- Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385–405.
- Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94–112.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, Calif: Sage Publications.
- Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. New York, NY: Palgrave Macmillan.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, California: SAGE Publications.
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, 5(1), 1–35.
- Dikli, S., & Bley, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17.
- El-Ebary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, 10(2), 121–142.
- Elliot, S. (2002). Intellimetric™: From here to validity. In M. D. Shermis, & J. C. Burstein (Eds.). *Automated essay scoring: A cross-disciplinary perspective* (pp. 71–86). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32, 335–349.
- Enright, M. K., & Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater® scoring. *Language Testing*, 27(3), 317–334.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- 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), 1–43.
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31–44.
- Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432–454.
- 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.
- Li, Z., Feng, H., & Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *CALICO Journal*, 34(3), 355–375.
- Liao, H. (2015). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308–319.
- O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42–56.
- Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research*, 26(4), 407–425.
- 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), 1884–1894.
- Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674.
- Ranalli, J. (personal communication, March 25, 2019) claimed that Grammarly can be integrated into any word-processing environments and provide feedback in piecemeal and synchronously.
- Ranalli, J., & Yamashita, T. (2019, March). Synchronous automated written corrective feedback: Effects on L2 college students' revision behavior and text quality. Paper presented at the American Association for Applied Linguistics (AAAL) Annual Conference.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8–25.
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65.
- Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In K. Hyland, & F. Hyland (Eds.). *Feedback in second language writing: Contexts and issues* (pp. 125–142). (2nd ed.). New York, NY: Cambridge University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371–391.
- Wang, Y., Shang, H., & Briody, P. (2012). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 1–24.
- Warden, C. A. (2000). EFL business writing behaviors in differing feedback environments. *Language Learning*, 50(4), 573–616.
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies*, 3(1), 22–36.
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 157–180.
- Yin, R. K. (2009). *Case study research: Design and methods* (2 ed.). London: SAGE Publications.
- Zhang, Z. (2017). Student Engagement with computer-generated feedback: A case study. *ELT Journal*, 71(3), 317–328.
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102.
- Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13–24.

**Svetlana Koltovskaia** is a fourth-year Ph.D. student in the doctoral program in TESL and Applied Linguistics at Oklahoma State University in the USA. Her main research areas are technology in education, L2 writing, intercultural pragmatics, and genre analysis.