Journal of Response to Writing

Volume 9 | Issue 2 Article 1

2023

# Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation

Hooman Saeli
*The University of Tennessee, Knoxville*

Payam Rahmati
*Oklahoma State University, Stillwater*

Svetlana Koltovskaia
*Northeastern State University*

Follow this and additional works at: https://scholarsarchive.byu.edu/journalrw

Part of the Applied Linguistics Commons, Arts and Humanities Commons, Education Commons, and the First and Second Language Acquisition Commons

## Recommended Citation

Saeli, Hooman; Rahmati, Payam; and Koltovskaia, Svetlana (2023) "Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation," *Journal of Response to Writing*: Vol. 9: Iss. 2, Article 1.
Available at: https://scholarsarchive.byu.edu/journalrw/vol9/iss2/1

This Featured Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of Response to Writing by an authorized editor of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

https://scholarsarchive.byu.edu/journalrw/

RW
JOURNAL OF RESPONSE TO WRITING

# Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation

**Hooman Saeli**
*University of Tennessee, Knoxville*

**Payam Rahmati**
*Oklahoma State University, Stillwater*

**Svetlana Koltovskaia**
*Northeastern State University*

Abstract: The study explored six ESL university students' behavioral, cognitive, and affective engagement with e-rater feedback on local issues and examined any changes in students' engagement over two weeks. We explored behavioral engagement through the analysis of screencasts of students' e-rater usage and writing assignments. We measured cognitive and affective engagement by analyzing students' comments during the think-aloud protocol and reflection surveys. The findings indicated that the students had varying levels of engagement with the feedback. Behaviorally, all students used a range of revision operations to address errors based on the provided feedback. Cognitively, some students were more engaged than others. Affectively, students experienced both positive and negative reactions toward e-rater feedback. While some students' engagement with feedback did not change over two weeks, others' engagement grew more negative. We conclude that e-rater feedback could positively impact students' accuracy in local aspects of writing if students are actively engaged with the feedback.
**Keywords:** learner engagement, second language (L2) writing, e-rater, automated writing evaluation (AWE)

Saeli, H., Rahmati, P., & Koltovskaia, S. (2023). Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation. *Journal of Response to Writing*, *9*(2), 1–39.

Written corrective feedback (WCF) has been defined as commentary meant to increase the grammatical accuracy of student writing (Ferris, 2010). Earlier studies on the efficacy of WCF compared experimental and control groups to determine whether specific types of WCF caused higher accuracy gains among second language (L2) learners (Bitchener et al., 2005; Chandler, 2003; Ellis et al., 2008; Ferris & Roberts, 2001; Sheen, 2007). However, with an increasing focus on learners as feedback users, recent research has conceptualized and operationalized what learners do after receiving WCF, or what Ellis (2010) calls *engagement* with feedback. Studies (e.g., Han & Hyland, 2015; Saeli & Cheng, 2021; Zhang & Hyland, 2018) suggest that learner engagement mediates the effectiveness of feedback, including teacher-generated WCF (Han & Hyland, 2015), peer WCF (Saeli & Cheng, 2021), and automated feedback (Koltovskaia, 2020; Zhang, 2020). In the present study, we focus on the feedback an automated writing evaluation (AWE) system provides and learners' engagement with this type of feedback.

As Stevenson and Phakiti (2019) note, AWE systems were first implemented into L2 assessment contexts to address potential fallibility in human scoring procedures. Such an implementation has, however, been replete with debates on the nature of L2 learning, L2 writing development, and the efficacy of automated feedback (Stevenson, 2016). Some scholars view AWE-generated WCF as a way of alleviating teachers' workload, giving student writers the luxury of having limitless opportunities to receive and act upon WCF while advocating learner autonomy (Stevenson & Phakiti, 2019; Woodworth & Barkaoui, 2022). Other scholars, however, perceive AWE-generated WCF as dehumanizing the L2 writing process, promoting a product-oriented approach to teaching and learning L2 writing (Vojak et al., 2011).

Regardless of the above controversies, the use of AWE systems to provide WCF on L2 writing has witnessed significant growth within the last few decades as technology has advanced (Zhang, 2020; Zhang & Hyland, 2018). Nonetheless, the effectiveness of AWE-generated WCF is still a hotly debated topic, and one area of concern is inaccurate

Saeli, H., Rahmati, P., & Koltovskaia, S. (2023). Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation. *Journal of Response to Writing*, *9*(2), 1–39.

WCF and, thus, the need for teachers' verification of WCF accuracy (Dikli & Bleyle, 2014; Stevenson & Phakiti, 2019). To this end, several studies have explored AWE systems' potential in providing WCF on student writing: Grammarly (Koltovskaia, 2020), Pigai (Zhang, 2020), Criterion (El Ebyary & Windeatt, 2010), and MY Access! (Grimes & Warschauer, 2010). These systems are more computationally geared toward providing feedback on local aspects of writing (Ranalli et al., 2017).

The existing studies have primarily explored the validity and reliability of AWE systems' scoring (Burstein & Chodorow, 1999; Elliot, 2002), learner perceptions of the usefulness of AWE feedback (e.g., Chen & Cheng, 2008; Dikli & Bleyle, 2014), and the effectiveness of AWE feedback on the quality of student writing (e.g., Attali, 2004; Chapelle et al., 2015; El Ebyary & Windeatt, 2010). Nevertheless, several shortcomings in these studies merit further investigation regarding the effects of AWE-generated WCF on student writing. In their review of AWE-generated feedback in L2 writing studies, Fu et al. (2022) conclude that compared with feedback from teachers and peers, AWE-generated WCF has led to relatively low effect sizes on the quality of student writing. In this paper, we hypothesize that this relative ineffectiveness may be partially rooted in how L2 learners engage with AWE feedback.

## Learner Engagement With AWE Feedback

Earlier research on e-rater, an AWE platform that provides feedback on local writing-related issues, focused on the validity and reliability of its scoring within the context of high-stakes tests (Burstein & Chodorow, 1999; Powers et al., 2002). More recent studies (e.g., El Ebyary & Windeatt, 2010) have explored the instructional potential of the AWE platform Criterion, which uses the e-rater engine. Some classroom-based studies have examined learners' perceptions of Criterion feedback. Dikli and Bleyle (2014) found that ESL

students perceived Criterion feedback as helpful despite missed and misidentified errors. Some other studies reported mixed results about the effects of Criterion feedback on the quality of student writing. El Ebyary and Windeatt (2010) found that Criterion feedback improved the quality of Egyptian students' writing. However, some students achieved better scores using the avoidance strategy, where they became aware of their flagged errors and refrained from using them in future writing. Nevertheless, the authors reported that Criterion encouraged students to revise their essays (100% resubmission rate). Likewise, Li et al. (2015) found that Criterion feedback helped ESL students revise their writing and improve grammatical accuracy.

While the above studies point to the positive effects of Criterion, others have highlighted the shortcomings in the feedback Criterion provides to students. Chapelle et al. (2015) found that ESL students disregarded nearly 50% of Criterion feedback and made limited changes to their resubmissions. The authors suggested that inaccuracies in Criterion feedback likely undermined these students' confidence in it. Ranalli et al. (2017) reported that Criterion was not accurate enough to provide valuable formative feedback, so its feedback did not lead to successful revisions. However, the authors claimed that Criterion could be useful because even inaccurate feedback may help students notice linguistic forms. Overall, these studies show that despite positive perceptions about AWE feedback, learners might not use this feedback to revise their writing. This leads us to the questions of what students do with AWE feedback after receiving it and why they do what they do, which is collectively what Ellis (2010) refers to as learner engagement with feedback.

Focusing on feedback in L2 studies, Zhang and Hyland (2018) define learners' engagement with feedback as "the extent [to which] students are invested or committed to their learning, embracing a complex of factors which can be seen in students' responses to texts and their attitudes to writing and responding" (p. 91). Learner engagement

with feedback comprises three dimensions: behavioral, cognitive, and affective (Ellis, 2010; Han & Hyland, 2015; Zhang, 2017; Zhang & Hyland, 2018; Zheng & Yu, 2018). We view learner engagement with AWE feedback as a multidimensional, dynamic, continuous construct, and we think that the above three dimensions can provide a richer understanding of the dynamic and complex nature of learners' engagement with e-rater feedback.

The available literature offers valuable insights into learners' engagement with AWE feedback. Zhang (2017) investigated how one university student engaged with holistic scores and written comments from the Chinese AWE system *Pigai*. Zhang concluded that positive engagement with AWE feedback will likely improve its effectiveness. Additionally, Zhang and Hyland (2018) explored two Chinese university students' engagement with both teacher-generated WCF and *Pigai* feedback. The results showed that the highly engaged student preferred AWE feedback over teacher feedback because *Pigai* provided immediate feedback and allowed her to resubmit her essay. However, the other student had limited engagement with AWE feedback because he was overwhelmed by the amount of feedback and felt embarrassed by his low scores. Zhang and Hyland concluded that AWE feedback is a valuable source of formative assessment because it offers frequent and process-oriented diagnostic assessment throughout the revision process. They also suggested that AWE feedback should be used as a supplement to teacher WCF. These studies provide valuable insights into how learners engage with AWE feedback behaviorally, cognitively, and affectively. However, we still do not know about the potentially dynamic nature of learners' engagement with this type of feedback; it is currently unclear whether, why, and how learners' engagement with AWE feedback changes over time.

## The Present Study

The existing literature suggests that L2 learners may not use AWE-generated feedback in future revisions, particularly because of their negative engagement with AWE feedback. We agree with Zhang (2017) and Zhang and Hyland (2018) that careful investigation of learners' engagement with AWE-generated feedback can provide important insights into how effectively this feedback improves learners' accuracy in local aspects of writing. In the present study, we address several gaps in the research on engagement with AWE feedback.

Few studies have investigated learner engagement with AWE-generated feedback, although this engagement is essential to maximizing the effects of feedback (Zhang & Hyland, 2018). Of the few studies on learner engagement with AWE feedback (e.g., Zhang, 2017), no previous research has examined the potential changes in this engagement over two drafts spaced out over time. As Zhang and Hyland (2018) noted, very few studies have focused on the process of writing in response to AWE-generated WCF. We believe that investigating L2 learners' engagement (i.e., affective, behavioral, and cognitive) with AWE-generated WCF can provide novel insights into the working of this WCF in the writing process. We also concur with Qi and Lapkin (2001) and Storch and Wigglesworth (2010) that learners' engagement with AWE-generated WCF should be examined within the classroom context to paint a clearer picture of the effects of this feedback on L2 learners' engagement and, eventually, their L2 writing development (e.g., accuracy gains). In addition, we are unaware of any prior study of learners' engagement with feedback on local aspects of writing generated by the Educational Testing Service's (ETS) e-rater. Therefore, our series of naturalistic case studies addressed these gaps, and we investigated the following questions:

1. How do university ESL students engage with e-rater WCF on local aspects of writing?

2. How does the engagement of university ESL students with this WCF change over a two-week period?

## Methodology

### Operationalization of Variables

Drawing on previous research (Ellis, 2010; Han & Hyland, 2015; Zhang & Hyland, 2018), we operationalized learner engagement with feedback:

- Behavioral engagement: learners' revision operations in response to AWE WCF (e.g., accepting feedback and employing revision strategies)
- Cognitive engagement: learners' use of cognitive and metacognitive strategies when using AWE WCF
- Affective engagement: learners' emotional reactions upon receiving AWE WCF and their overall attitudinal responses

We also operationalized WCF as commentary to increase learners' awareness of local errors and provide corrections to these errors.

### Study Content

This study took place in an L2 writing course at a US university during the fall semester in 2020. The course focused on expository composition, emphasizing technique and style in writing research papers from a genre perspective. The 16-week course met three times per week for 50-minute sessions. It included three major writing assignments: two descriptive and evaluative summaries and one research proposal. For each assignment, teachers provided feedback on global aspects of writing (e.g., content development and organization) for students' first drafts. Once students revised their drafts based on the teacher's feedback, they were required to upload their writing to Canvas, a virtual learning environment used at the participating university, to receive e-rater feedback integrated into Turnitin.

Our rationale behind using e-rater was twofold. First, this platform was already built into the course's Canvas page. This means that it was freely accessible to every student enrolled in the composition classes of the participating university. Second, because of the paradigmatic differences among the available AWE systems (Fu et al., 2022), we were interested in exploring whether working with e-rater would lead to any different engagement patterns compared with the existing studies which have utilized other AWE systems.

After receiving e-rater feedback on local aspects of writing and revising their errors accordingly, students were required to submit their second drafts to their teacher for summative feedback. This procedure aligns with the process approach, where students submit multiple drafts of essays, but only the final draft is graded. The multi-draft nature of the assignments allowed us to explore learners' engagement with e-rater WCF along with any changes in this engagement across the two-week period.

**Participants**

Six of 12 enrolled students in an intact class consented to participate in the study. The course instructor was a then-doctoral student in the Applied Linguistics program of the participating university. In his doctoral program, he had received training in teaching composition courses and providing feedback. At the time of data collection, the instructor had three years of experience teaching composition classes. The student participants' profiles can be seen in Table 1. All names are pseudonyms. The students' language proficiencies were determined based on standardized test results.

**Table 1**
*Participant Profiles*

| Name | Gender | Age | Country | L1 | Major | Level |
|---|---|---|---|---|---|---|
| Alex | Male | 22 | France | French | Human Resource Management | Upper-intermediate |
| Ali | Male | 21 | Kuwait | Arabic | Physics | Upper-intermediate |
| Eman | Female | 24 | Saudi Arabia | Arabic | Chemical Engineering | Upper-intermediate |
| Fahad | Male | 21 | Kuwait | Arabic | Physics | Upper-intermediate |
| Kim | Male | 23 | Korea | Korean | Mechanical Engineering | Upper-intermediate |
| Ram | Male | 24 | Thailand | Thai | Aviation Management | Upper-intermediate |

## WCF Provided by e-rater

The plagiarism detection service Turnitin has recently incorporated e-rater to expand itself beyond detecting plagiarism to facilitating feedback (Stevenson & Phakiti, 2019). In this study, Turnitin was integrated into Canvas, which the students had access to for all three major assignments. Specifically, e-rater directly indicated an error by highlighting it and providing a code (e.g., "S/V" for subject-verb agreement), and then provided WCF (e.g., "You may need an article before this noun."). The error categories were grammar, mechanics, usage, spelling, and style; each category had several subcategories. For example, grammar contained nine error subcategories, including S/V, fragment, run-on sentences, and proofread. Figure 1 shows the text and feedback from e-rater highlighted in purple.

**Figure 1**

*e-rater Interface*

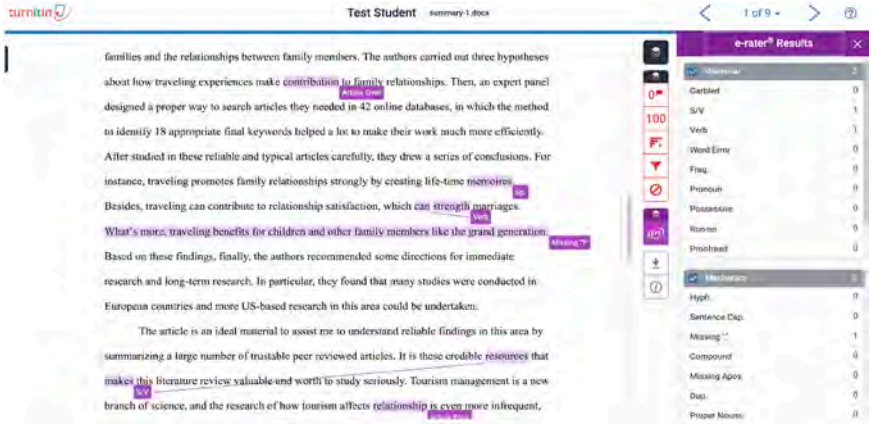



Figure 2 shows the typical WCF this system provided on our learners' writing.

**Figure 2**

*Sample WCF Provided by e-rater*

If learners seek additional information on a grammatical structure, e-rater contains a handbook that is accessible by clicking "View Handbook."

## Data Collection

We triangulated our data collection by utilizing students' writing with e-rater WCF, screencasts of their writing with think-aloud protocols, and reflection surveys. This enabled us to explore the learners' engagement with WCF and any changes in their engagement. For the students' writing, we chose the two summaries for two reasons. First, this genre is very common in the students' writing program. Second, the students were asked to write two summaries over two weeks, allowing us to explore any changes in their engagement with feedback within the same genre. For these summaries, the students were required to find two scholarly articles within their academic disciplines and write a descriptive and evaluative summary for each article. The average length of each summary was between 450–500 words. For the six participating students, the combined length of two summaries was as follows: Alex, 978; Ali, 721; Kim, 834; Fahad, 987; Eman, 856; and Ram, 967 words. Teachers provided no WCF on local aspects of writing, so the only official source of feedback on such aspects was e-rater. Henceforth, we refer to e-rater WCF as WCF.

After receiving e-rater feedback, the students revised their summaries. While writing their second drafts using this WCF, the participating students ($N$ = 6) were asked to record their revision process with Zoom and do a think-aloud protocol. Before this, all 12 students had been trained to use e-rater for revision and Zoom screencasting to record their revision process; the students were also instructed to do a think-aloud protocol per the course curriculum. The course instructor devoted two lessons to training students. The two sessions took place at the computer lab in the English Department of the participating university and lasted approximately 30 minutes each. Two similar tutorial videos were uploaded to Canvas for students wanting to revisit the in-person training.

The students uploaded their second drafts to Canvas for summative teacher feedback. Then, the students were asked to upload their

Zoom screencasts showing their revision process with think-aloud protocols to another Canvas folder. Immediately after revising the summaries, the students completed a reflection survey. This survey was in the form of a quiz on Canvas and included eight open-ended questions. The submission and revision of the descriptive summary and the evaluative summary involved the same process.

**Data Analysis**

We explored behavioral engagement by analyzing students' screencasts and their summaries in three phases. Since AWE feedback can be fallible (Chapelle et al., 2015; Fu et al., 2022), which can ultimately affect learners' engagement with it, we first determined the accuracy of the provided WCF. We utilized the same categories as e-rater to classify the flagged errors and manually labeled the WCF as accurate or inaccurate. In the second phase, we analyzed the screencasts and identified the students' revision operations. We defined these operations as any actions in response to the WCF and identified four revision operations used by the students: Correct revision, no change, incorrect change, and correct substitution. We operationalize these operations and provide some examples from the analyzed samples of student writing:

- *Correct revision* is when the student corrected an error per accurate WCF. For example, e-rater indicated that "then" may be a confused word in this sentence: "Parents and children participated in the reading aloud program benefited and showed greater cognitive development **then** those who didn't." The student (Eman) changed "then" to "than" per this suggestion.
- *No change* occurs when the student did not use (accurate or inaccurate) WCF and rejected it. For example, e-rater noted that the word "nomophobia" is misspelled in this sentence: "The authors used 3 types of

questionnaires with a general background, which were about addiction, anxiety, and **nomophobia**." The student (Kim) checked the correct spelling on the internet and left the word unchanged, thus rejecting the WCF.

- *Incorrect change* is when the student addressed an error incorrectly due to (accurate or inaccurate) WCF. For example, e-rater suggested checking subject-verb agreement here: "As they note, the study was conducted where study **were** the participants." The student (Fahad) changed the verb "were" to "are": "As they note, the study was conducted where study **are** the participants," thus making an incorrect change.
- *Correct substitution* is when the student, triggered by (accurate or inaccurate) WCF, substituted the flagged error with their own correct form. For example, e-rater indicated that a student may need an article before "sleep patient" in this sentence: "However, it is recommended that the best non-pharmacological treatement for **sleep patient** is through excercisng regularly . . .". Instead of responding to the WCF and adding an article before "sleep patient," the student (Ram) rewrote the flagged phrase: "However, it is recommended that the best non-pharmacological treatement for **a better sleep quality** is through excercisng regularly . . .".

We considered *correct revision*, *correct substitution*, and *no change* to inaccurately flagged errors as tokens of improvement in our learners' summaries. We considered *incorrect change* and *no change* to accurately flagged errors as a lack of improvement in our participants' summaries.

In the third phase, we analyzed the screencasts and identified the students' revision strategies. We focused on any strategies aimed at

enhancing the accuracy of the drafts in response to WCF (e.g., using an online dictionary or Google to verify the accuracy of WCF). The data from the think-aloud protocols were analyzed to reaffirm any identified revision operations and strategies. Overall, these data helped us explore the students' behavioral and cognitive engagement with WCF.

We analyzed the think-aloud protocols by downloading the screencasts from Zoom as an audio file and transcribing them. The entries from the reflection surveys were extracted from Canvas and transferred onto a Google Document. The data from these two sources helped us analyze the students' behavioral, cognitive, and affective engagement with WCF. The analysis followed three steps: Open coding, axial coding, and selective coding, as described by Corbin and Strauss (2008). We first took individual notes of important ideas, compared these notes to create larger thematic categories, and corroborated these categories with the think-aloud protocols and reflection surveys. These analyses helped shed light on the students' otherwise unobservable thoughts, rationales, and attitudes. Moreover, by comparing the data from the first and second summaries, we aimed to detect and explain any changes in the students' engagement with WCF.

## Results

### Learner Engagement with WCF

Table 2 presents the number of WCF instances the students received on their first drafts, but we leave out the categories with no WCF. The frequencies are for 12 summaries written by the six students.

Table 2 shows that the WCF was most frequently provided on article errors, followed by subject-verb agreement and spelling issues. We counted 46 instances of WCF on the 12 summaries.

**Table 2**

*The WCF on the First Drafts of Both Summaries*

| | | Alex | Ali | Kim | Fahad | Eman | Ram | Total |
|---|---|---|---|---|---|---|---|---|
| Grammar | Possessive | | | | | 1 | | 1 |
| | S/V | 1 | 2 | | 1 | | 2 | 6 |
| | Run-on | | | | | | 1 | 1 |
| | Pronoun | | | | 1 | | | 1 |
| | Fragment | | | 1 | | 1 | | 2 |
| Mechanics | Missing "," | | 1 | 1 | | | 1 | 3 |
| | Sentence capitalization | | | 1 | | | | 1 |
| Usage | Faulty comparison | | | | | | 1 | 1 |
| | Confused | | | | | 1 | 1 | 2 |
| | Preposition | | 1 | 1 | | | 1 | 3 |
| | Wrong article | | 1 | | | | | 1 |
| | Article error | 5 | 2 | 1 | 4 | 4 | 2 | 18 |
| Spelling | Spelling | | | 2 | 2 | 2 | | 6 |
| Total | | 6 | 7 | 7 | 8 | 9 | 9 | 46 |

To explore learner engagement with WCF, we focused on the learners' revision operations and strategies; their use of cognitive and metacognitive strategies; and their thoughts, attitudes, and perceptions while revising their first drafts. We also examined any changes in the learners' engagement with WCF during the two-week period.

*Behavioral Engagament with WCF*

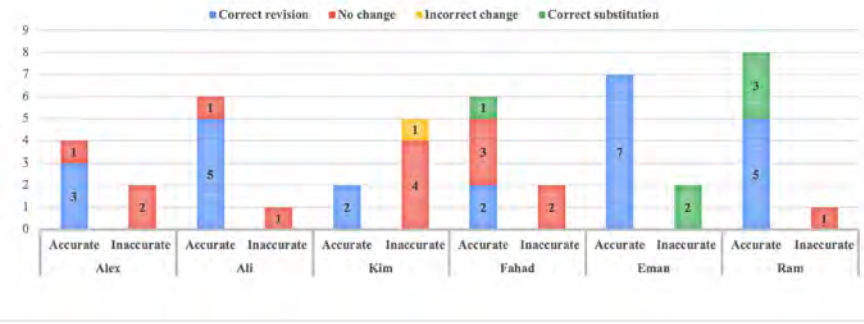
To examine behavioral engagement with WCF, we explored whether the learners incorporated this feedback in the second drafts of their summaries. For a total of 46 flagged errors, the learners used the following revision operations in response to both accurate and inaccurate WCF: Correct revision (24/46), no change (15/46), incorrect change (1/46), and correct substitution (6/46). All 24 errors that the learners correctly revised had been flagged accurately by e-rater. Of the 15 errors that the learners did not change, five had been flagged accurately, but 10 had not. The one error that a learner incorrectly revised in his second draft had been flagged inaccurately. Finally, four of the six errors the learners correctly substituted had been flagged accurately, but two had not. Overall, we noticed that, out of the 46 flagged errors, 33 had been flagged accurately, but 13 had been flagged inaccurately by e-rater.

Figure 3 presents the results of the learners' behavioral engagement. We have provided two bars for each student: one for accurately flagged errors and another for inaccurately flagged errors. The two color-coded bars also include the statistics for a given learner's revision operations in response to accurately or inaccurately flagged errors.

Figure 3 presents important results about each learner. Though we present the learners' improvements in terms of percentages, the underlying frequencies for these percentages can be found in Figure 3. In our data analyses, we considered correct revision, no changes after receiving inaccurate WCF, and correct substitution as tokens of improvement. In contrast, we thought of no change after receiving accurate WCF and incorrect change as tokens of lack of improvement.

**Figure 3**

*Behavioral Engagement with WCF on Both Summaries*




Below, we provide a descriptive account of the students' revision operations.

- The first learner, Alex, correctly revised three errors based on three accurate instances of WCF but did not change one error that was accurately flagged and two errors that were inaccurately flagged, thus improving the accuracy of his second drafts by 83%.
- Ali correctly revised five errors based on five accurate instances of WCF but did not change one error that was accurately flagged and one error that was inaccurately flagged, thus improving the accuracy of his second drafts by 86%.
- Kim correctly revised two errors based on two accurate instances of WCF and did not change four errors that were inaccurately flagged, but incorrectly changed one error that was accurately flagged, thus improving the accuracy of his second drafts by 86%.
- Out of six accurate instances of WCF, Fahad correctly revised two, correctly substituted one, and did not

change two. He also did not change two inaccurately flagged errors, thus improving the accuracy of his second draft by 63%.

- Eman correctly revised seven errors based on seven accurate instances of WCF and correctly substituted two inaccurately flagged errors, thus improving the accuracy of her second drafts by 100%.
- Ram correctly revised five errors based on five accurate instances of WCF and correctly substituted three errors based on three accurate instances of WCF. He did not change one inaccurately flagged error, thus improving the accuracy of his second drafts by 100%. These results show that the learners exhibited varying levels of behavioral engagement with WCF.

*Cognitive and Affective Engagement with WCF*

We examined each learner's cognitive and affective engagement with WCF by analyzing the screencasts, think-aloud protocols, and reflection surveys. Moreover, we compared the screencasts with think-aloud protocols and reflection surveys to see whether our learners' engagement with WCF changed between the two summary assignments in the two-week interval.

**Alex: Intuition Overriding WCF**. Alex was cognitively engaged with WCF and utilized cognitive strategies such as rereading his sentences and the WCF. The analysis of his screencasts showed that he addressed errors individually and sometimes returned to the errors he felt unsure about. However, the think-aloud protocols and reflection surveys suggested that Alex's intuition sometimes overrode the WCF. Alex shared his uncertainties regarding using the WCF:

> *Excerpt 1:* When I saw the feedback about subject/verb, I tried to think of why it could be wrong. The feedback said that the subject

> and verb may not agree, but I was pretty sure it was correct. I rejected the feedback because I could not see any other way to say it, and in my opinion it was correct. I think I did not really understand the feedback, because in my case I think it was wrong. (Reflection Survey 1)

Excerpt 1 shows Alex's uncertainty about the accuracy of WCF. This potentially led him to use his knowledge when revising the highlighted errors instead of consulting other sources. According to Figure 3, Alex's accuracy gains in his second drafts were more limited than some of the other learners (e.g., Eman). His uncertainty about feedback accuracy may explain why he did not change one accurately flagged error. Moreover, Alex's affective engagement with WCF was somewhat negative:

> *Excerpt 2:* I was more confused about what could have been wrong, which made me lose some times to think about it. However, since it is a software, I am always careful to not correct everything it tells me to change because sometimes it can be wrong. (Reflection Survey 1)

Excerpt 2 includes two interesting points. First, Alex seemed confused by the WCF because of the lack of explanation for his errors. Second, Alex shared his doubts about the accuracy of the WCF because this feedback was generated by a fallible software system. These points suggest that Alex's affective engagement with WCF was mostly negative and that this negative engagement might have overshadowed his more positive cognitive and behavioral engagement with WCF.

**Ali: Amenable to WCF.** Ali understood most of the provided WCF and reviewed his errors and the feedback. This suggests that his cognitive engagement with WCF was quite extensive. Ali's think-aloud protocol shows his reasoning:

> *Excerpt 3:* There is a title which does not need an article and e-rater makes it wrong as article error so in this part I think I will ignore that because I am right here. (Think-Aloud Protocol)

Additionally, Ali's affective engagement was positive since he preferred to receive explicit WCF, highlighted and color-coded errors, and feedback that could improve his grammatical accuracy. Excerpt 4 shows Ali's attitudinal response to the WCF:

> *Excerpt 4:* I like the e-rater feedback in general because it's like double checking your essay because all of us want to have a perfect essay and we would like to prevent mistakes as much as we can. (Reflection Survey 1)

Because of the explicit nature of the WCF and his intention to increase his accuracy in writing, Ali's preferences and the WCF he received were well-aligned, thus making his affective engagement with WCF mostly positive.

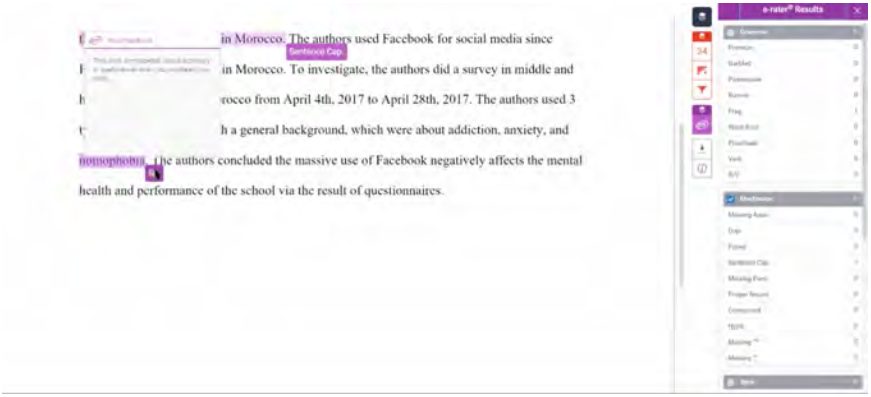
**Kim: Seeking Reassurance from Outside Sources.** The screencasts showed that, cognitively, Kim was highly engaged with WCF. For instance, he verified the spelling of Morocco and nomophobia on Google. He also used a dictionary to check the highlighted errors and spelling issues. Excerpt 5 demonstrates Kim's confusion with WCF during the think-aloud protocol:

> *Excerpt 5:* What? This is weird! What's wrong with "nomophobia." I am typing it in Google (reads the definition out loud). It's correct! (Think-Aloud Protocol)

Figure 4 provides a screenshot of the WCF Kim received.

**Figure 4**

*Screenshot of the WCF Provided to Kim*




Kim seemed to engage with WCF extensively because he felt confused about it. He also referred to the e-rater Handbook to learn about commas after accepting the feedback with some uncertainty:

> *Excerpt 6:* Because my paragraph was a bit short, there were only three things mentioned from e-rater. However, two of them were weird. It said there was a wrong usage of capitalization for the sentence "the school for the students in Morocco." Also, e-rater said I misspelled "nomophobia." I ignored both of them since I think there were some kind of error on understanding words for e-rater. When I read feedback about the misspelled word, I did not why I need to change "medRxiv." I think the word is not in the dictionary so it identified the misspelled word. I rejected the feedback. (Reflection Survey 1)

Our analysis of Kim's reflection surveys showed that he shared some positive perceptions and positive affective engagement about the WCF:

> *Excerpt 7:* I still wonder about the passive/active voice sometimes such as "I'm bored" or "It's boring." So if e-rater catches them,

including usage of comma, it would be really helpful. (Reflection Survey 1)

We should note that Kim correctly rejected four inaccurate instances of WCF that he received, probably because of his extensive cognitive engagement with WCF. Overall, Kim was satisfied with the WCF he had received and reported that he would use it again.

**Fahad: Confused About and Dismissive of WCF**. Fahad did not understand some of the WCF. For instance, he was uncertain about adding an article before his name and about removing the article before "plastic pollution." His screencasts showed that he did not change the errors that e-rater accurately flagged, and he displayed somewhat negative cognitive engagement with the WCF. Also, Fahad drew upon his knowledge, rejected some accurate WCF, and he did not use external resources:

> *Excerpt 8:* The feedback says I used an incorrect word, but I think I used the correct one based on what we learned in class. I rejected the feedback we were taught in the class about the difference between "affect" and "effect." (Think-Aloud Protocol)

Excerpt 8 reveals that Fahad preferred his teacher's feedback when he mentioned the accuracy of "what we learned in class."

Fahad's affective engagement with the WCF was negative. As shown in his reflection surveys, Fahad felt surprised about the low number of e-rater comments:

> *Excerpt 9:* My overall impression of e-rater feedback is I did not like it a lot . . . I think I saw better websites on the internet that give better feedback for articles, so it will be like my second or third option if I want to get feedback. Computers and websites can not do everything perfectly. I prefer feedback from a teacher who reads and understands what I want to say or write. In addition to great feedback he/she can give you help, usually some tips and instructions to improve your weak points. (Reflection Survey 1)

Excerpt 9 includes two important points about Fahad's negative affective engagement with the WCF. First, he did not seem to trust e-rater as a credible source of grammar-centered feedback. Secondly, Fahad was generally dismissive of "computers and websites" because the feedback generated by these sources was not helpful in his writing. Therefore, he preferred feedback from his teachers.

**Eman: Highly Motivated and Engaged with WCF**. After analyzing the screencast data, we found that Eman was cognitively engaged with WCF. She, for example, used Google to ensure the accuracy of her substitutions. Eman's think-aloud protocol showed how she rejected the WCF and opted for a correct substitution instead:

> *Excerpt 10:* It wants possessive here. I don't think it understands me. I just need to delete an article here. (Think-Aloud Protocol)

Eman also carefully examined her sentence fragment errors and WCF. In Excerpt 11, she explained how she engaged with WCF cognitively:

> *Excerpt 11:* When I saw the feedback, some error I did not understand so I search to get it. It is good, I like where to use articles. It is very helpful to see how I did at my essay and how I can correct all mistakes. (Reflection Survey 1)

Eman was also affectively engaged with WCF. In fact, she reported trusting e-rater because of the plagiarism checkers embedded in Turnitin:

> *Excerpt 12:* I really like e-rater feedback because . . . it always tells you how to improve. As I said before it is really good to get feedback from professor or e-rater so then I know how to improve. I really like this feedback application/website. It helps me see if I used plagiarism (even if I did not want to) and then change my articles, or it helps me to correct small mistakes as misspelling or sentence correction. (Reflection Survey 1)

Excerpt 12 indicates that Eman preferred to receive WCF to notice her grammatical errors. Her reflection surveys also suggested that Eman positively evaluated the WCF from both her teachers and e-rater, as they helped her improve the accuracy of local aspects in her writing.

**Ram: Dependent on His Own Knowledge.** Ram was cognitively engaged with WCF, but he tended to rely on his knowledge and did not utilize any outside sources in his revision operations. His screencasts showed that Ram checked the errors and WCF several times to ensure the accuracy of his revisions. Whereas he preferred meta-feedback on his errors, e-rater only provided direct WCF, which Ram disliked. We also analyzed Ram's reflection surveys and explored his affective engagement with WCF:

> *Excerpt 13:* When I saw it [preposition error] I knew it was a easy mistake and I did not check carefully before I submit the assignment. I accepted the e-rator feedback and changed the word because it's right. Sometimes we make mistakes in our writing without noticing and it really can help you correct them. (Reflection Survey 1)

Excerpt 13 shows that Ram was engaged with the WCF cognitively. He paid attention to his errors and the WCF, reviewed his errors, and double-checked his issues multiple times. Although his reflection surveys showed that Ram viewed WCF to draw his attention to errors, he preferred meta-feedback, which e-rater lacked. Therefore, his affective engagement with WCF was mixed.

### Changes in Learners' Engagement with WCF

With data collected over two summaries written two weeks apart, and the learners receiving two rounds of WCF on each of these summaries, we explored any developments in their engagement with the feedback. We first examined the changes in behavioral engagement by comparing the learners' revision strategies in response to WCF on

their two summaries. Table 3 presents the results for the first and second summaries:

**Table 3**

*The WCF on the First Drafts of Both Summaries*

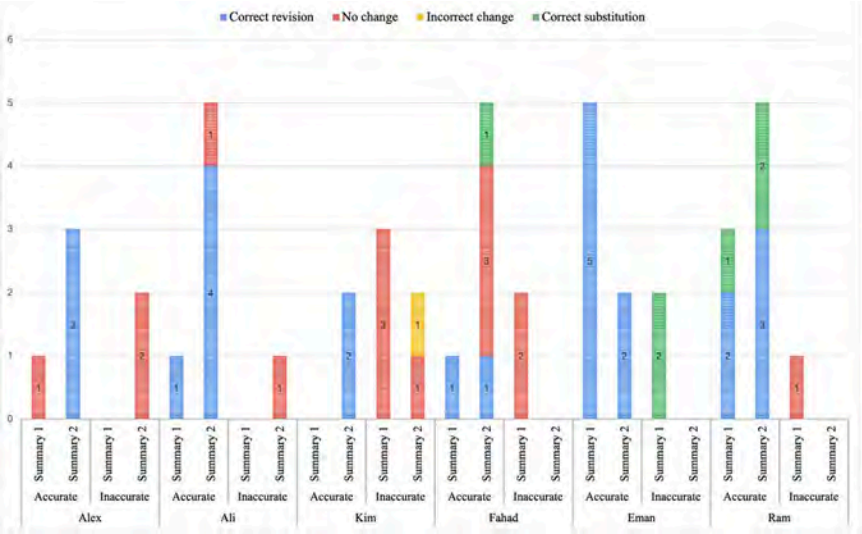
| | Alex | | Ali | | Kim | | Fahad | | Eman | | Ram | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | S1 | S2 | S1 | S2 | S1 | S2 | S1 | S2 | S1 | S2 | S1 | S2 | S1 | S2 |
| Grammar | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 2 | 6 | 5 |
| Mechanics | | | | 1 | 1 | 1 | | | | | 1 | | 2 | 2 |
| Usage | | 5 | | 4 | | 2 | 1 | 3 | 5 | | 2 | 3 | 8 | 12 |
| Spelling | | | | | 1 | 1 | 1 | 1 | 1 | 1 | | | 3 | 3 |
| Total | 1 | 5 | 1 | 6 | 3 | 4 | 3 | 5 | 7 | 2 | 4 | 5 | 19 | 22 |

*Note*: S1 = Summary 1; S2 = Summary 2

The analysis of the learners' behavioral engagement with the WCF on the first drafts of their two summaries yielded important results. Figure 5 includes our learners' revision operations in the first and second summaries:

**Figure 5**

*Revision Operations in the Second Drafts of the First and Second Summaries*




We first elaborate on the learners' engagement with WCF in their first summaries.

- As shown in Figure 5, Alex did not change one error that was accurately flagged, thus improving the accuracy of his second draft by 0%.
- Ali correctly revised one error based on one accurate instance of WCF, thus improving the accuracy of his second draft by 100%.
- Kim did not change three inaccurately flagged errors.
- Fahad correctly revised one error based on one accurate instance of WCF but did not change two inaccurately flagged errors, thus improving the accuracy of his second draft by 100%.
- Eman correctly revised five errors based on five accurate instances of WCF and correctly substituted two

inaccurately flagged errors, thus improving the accuracy of her second draft by 100%.

- Ram correctly revised two errors based on two accurate instances of WCF and correctly substituted one error based on one accurate instance of WCF. He did not change one error that was inaccurately flagged, thus improving the accuracy of his second draft by 100%.

We now discuss our learners' engagement with WCF provided on their second summaries. Figure 5 illustrates that Alex correctly revised three errors based on three accurate instances of WCF, but he did not change two inaccurately flagged errors, thus improving the accuracy of his second draft by 100%. Ali correctly revised four errors based on four accurate instances of WCF but did not change one error that was accurately flagged and one error that was inaccurately flagged, thus improving the accuracy of his second draft by 83%. Kim correctly revised two errors based on two accurate instances of WCF but did not change one error that was inaccurately flagged and incorrectly changed one error that was inaccurately flagged, thus improving the accuracy of his second draft by 75%. Fahad correctly revised one error based on one accurate instance of WCF and correctly substituted one error based on one accurate instance of WCF. However, he did not change three accurately flagged errors, thus improving the accuracy of his second draft by 40%. Eman correctly revised two errors based on two accurate instances of WCF, thus improving the accuracy of her second draft by 100%. Finally, Ram correctly revised three errors based on three accurate instances of WCF and correctly substituted two errors based on two accurate instances of WCF, thus improving the accuracy of his second draft by 100%.

We explored the differences in the learners' behavioral engagement with WCF in the first and second summaries. We noticed that Kim's engagement with WCF in the second summary showed more revision operations than in his first one. Fahad's second draft showed

his use of more incorrect revision strategies; he seemed more confused in the second round of revisions. Although Eman received more WCF on her first draft, she used accurate revision operations in both summaries. Similarly, Ram used accurate revision operations in both summaries.

We categorized the learners based on the changes in their patterns of affective engagement with WCF in the following groups. The first group comprised the learners with positive engagement with WCF after both summaries: Ali, Kim, and Eman. For example, Ali liked feedback because it helped increase his accuracy to "make [his] writing more professional." He also liked "different colors and underlines." Kim was surprised because he received very few comments and sometimes felt confused; however, he later noted that he would use the system again in future writing assignments because of the plagiarism-checker feature. In addition, Eman liked the WCF on her grammatical and spelling errors. She mentioned her article errors because they were her "weak" points. Like Kim, Eman valued the plagiarism-checker feature:

> **Excerpt 14:** I really like e-rater feedback because it shows me where my essay level is. If we talk about grammar or whole sentences it always tells you how to improve. (Reflection Survey 2)

The second group was composed of one learner with positive engagement with WCF after the first summary but negative after the second summary: Ram. Despite receiving only a few instances of WCF, Ram believed that the WCF drew his attention to his errors:

> **Excerpt 15:** Sometimes we make mistakes in our writing without noticing and it really can help you correct them and receive a better grade. (Reflection Survey 1)

Ram, however, noted that sometimes the system did not provide good explanations for his grammatical errors. This, therefore, led him to prefer Grammarly:

> **Excerpt 16:** I think Grammarly does a better job. Students should use [e-rater] if there are no other options to get feedback from. (Reflection Survey 2)

The third group included the learners with negative engagement with WCF after both the first and second summaries: Alex and Fahad. For example, Alex was surprised by a few flagged errors and found the WCF confusing, irrelevant, and unspecific. Although Alex thought feedback systems could be useful for improving writing, his overall opinion remained negative, thereby preferring teacher-generated WCF:

> **Excerpt 17:** While it is useful to catch some grammatical errors that I might not have seen while proofreading, it can also lead to confusion. Many mistakes that the program points out are not relevant. Also, the feedback is too general. (Reflection Survey 2)

Like Alex, Fahad was also surprised by a few flagged errors and preferred to receive teacher WCF. To him, e-rater utilized many "unknown" words, which made it difficult for him to incorporate the WCF:

> **Excerpt 18:** Computers and websites cannot do everything perfectly. I prefer feedback from a teacher who reads and understands what I want to say or what I want to write. In addition to great feedback he/she can give you help, usually some tips and instructions to improve your weak points. (Reflection Survey 2)

We identified some factors that shaped our participants' affective engagement with WCF. First, the learners who maintained positive engagement throughout the study generally sought accuracy in writing, so they viewed e-rater as a platform to improve grammatical accuracy. Second, Ram's turned-negative engagement with WCF after his second summary was partially because of the lack of meta-feedback, such as explanations about grammatical errors. In addition, the two learners who showed negative engagement with WCF felt confused and uncertain about the WCF, preferring teacher-generated WCF. Overall,

we noticed that dependence on teachers was an important factor that shaped the learners' engagement with e-rater WCF. Another factor that influenced this negative engagement was the monologic nature of receiving WCF from e-rater. Ram, Alex, and Fahad reported feelings of confusion when working with this feedback and mentioned that they preferred having face-to-face interactions with a teacher. Therefore, we observed a strong preference for teacher-generated WCF among these learners since this source of feedback can be more dialogic than AWE-generated WCF. In other words, these learners valued verbal communication with a teacher regarding their errors.

## Discussion

We investigated six ESL learners' behavioral, cognitive, and affective engagement with e-rater WCF. Our data were collected from these learners' writing, reflection surveys, screencasts, and think-aloud protocols. Because our data were collected from two summaries written over a two-week period, we could examine whether and how the learners' engagement with feedback changed. Here, we revisit our two research questions and discuss the results.

### Question 1: Learners' Engagement with e-rater WCF

The findings revealed that several factors played into the learners' engagement patterns with WCF, suggesting that learner engagement is a complex construct (Ellis, 2010; Han & Hyland, 2015; Saeli & Cheng, 2021). Behaviorally, all learners engaged with WCF, used various revision operations in response to the feedback, and improved their second drafts. Specifically, the participants only behaviorally engaged with 31 of the 46 WCF instances. A closer analysis of this pattern, however, paints a different picture. Out of the 15 flagged errors not changed in the second drafts, 10 were flagged by e-rater inaccurately. This means that, although our learners' behavioral engagement was negatively affected by the relatively high number of "no

change" instances, their cognitive engagement with WCF led them to identify the inaccurately flagged errors correctly. Our further analysis indicated that the learners exhibited several cognitive strategies when dealing with both accurate and inaccurate instances of WCF. First, we realized that the students who correctly revised their errors showed a good understanding of the WCF and their errors. At times, they utilized outside sources (e.g., Google) to verify the accuracy of the WCF. In addition, the learners who did not change the accurately flagged errors seemed to exhibit trust issues with e-rater. We discuss this in more depth when elaborating on the learners' effective engagement with WCF. Nevertheless, the learners who did not change the inaccurately flagged errors showed clear patterns of cognitive engagement with WCF. These learners sought reassurance from outside sources (e.g., Kim) and utilized their knowledge (e.g., Alex) to verify the accuracy of the provided WCF and decide whether they needed to change the flagged errors.

The learners' affective engagement also played into their behavioral engagement in that the learners (i.e., Fahad and Alex) who doubted the accuracy of the WCF did not extensively change the accurately flagged errors. The analysis of these learners' reflection surveys revealed their strong preference for teacher-generated WCF. We can conclude that lower-than-expected behavioral engagement with WCF may not necessarily be a concern when working with AWE systems. On the contrary, at least in our study, somewhat limited behavioral engagement with WCF was a sign of cognitive engagement and negative affective engagement. Our results show that the learners' engagement with automated feedback was dynamic and non-linear.

Our results partially contradict the results of previous research that points to low behavioral engagement with automated feedback (Attali, 2004; Chapelle et al., 2015; Koltovskaia, 2020). Specifically, two learners (Eman and Ram) exhibited extensive behavioral engagement, successfully improved their drafts per e-rater suggestions, and

substituted inaccurate e-rater WCF with correct revisions. However, two other learners (Alex and Fahad) doubted the accuracy of e-rater WCF. This preference for teacher feedback has been reported in the literature (e.g., Saeli & Cheng, 2021), and these teacher-dominant preferences contributed to some of our learners' negative affective engagement with WCF. These patterns of negative affective engagement also fed into insufficient cognitive engagement with WCF. For example, Fahad's negative affective engagement with e-rater WCF led him to dismiss several errors that had been flagged accurately, affecting his cognitive engagement with WCF.

The learners also had varying levels of cognitive engagement with WCF. While all learners noticed the highlighted errors and WCF, some did not understand the corrective intention of the WCF. Most of these learners were critical of e-rater WCF and referred to outside sources (e.g., Google) to verify its accuracy. Still, some others relied on their intuition to engage with the WCF, which did not always result in the successful use of revision operations. Affectively, the participants showed various emotional reactions to e-rater WCF. While some found this WCF useful as it helped them improve their drafts, others considered it confusing and irrelevant. An important factor that contributed to some learners' negative affective engagement with WCF was their cognitive engagement with WCF. For instance, although Kim initially reported being confused about the WCF, he verified the WCF and used correct revision operations. This result has also been reported by Chapelle et al. (2015). The second factor was some learners' preference for teacher-generated WCF, which led them to place less value on AWE-generated WCF.

### Question 2: Changes in Learners' Engagement with e-rater WCF

We noticed interesting patterns in the changes in our learners' engagement with WCF during a two-week interval and over two summaries. For example, Eman and Ram showed positive engagement with

WCF in both summaries. We think positive perceptions about AWE systems possibly affected this positive engagement. Although Kim's behavioral engagement with WCF was rather negative, his cognitive and affective engagement led him to reject inaccurately flagged errors. Ram was the learner whose affective engagement with WCF grew more negative after the second summary. Interestingly, he correctly substituted three accurately flagged errors, indicating his cognitive engagement with WCF, but distrusted e-rater's confusing feedback, showing negative affective engagement with WCF.

Alex and Fahad were the learners whose affective engagement with e-rater WCF remained negative over the two summaries. Although Alex was more cognitively engaged with WCF than Fahad, Alex reported his doubts about the accuracy of e-rater WCF. Fahad, similarly, dismissed more than half of the flagged errors, which points to his strong cognitive disengagement with WCF. This lack of engagement was, at least partially, caused by his negative affective engagement with WCF. We conclude that affective engagement with e-rater WCF may have been the most critical factor in shaping our learners' initial engagement with e-rater WCF and any changes in their engagement with WCF.

Although our data were collected during only two weeks for just two summaries, our results suggest that learners may benefit from teacher intervention and teacher-generated verification, at least when they begin using AWE feedback. This way, their negative affective engagement might turn positive over two drafts written over two weeks. This could be especially true for those students who report patterns of distrust in AWE systems as a method of receiving WCF. Overall, our results suggest that, without effective instruction on the benefits of AWE systems as beneficial providers of WCF, students with initial negative affective engagement with this type of feedback will likely remain disengaged with AWE WCF, primarily due to these trust issues. Again, our results point to the importance of teacher intervention in

such instances because, in our study, the students who showed negative affective engagement with AWE WCF were those who appeared to rely on their course instructor for "accurate" WCF. Therefore, our data point to the importance of timely intervention that is aimed at altering students' negative (mis)perceptions about the use, effectiveness, and benefits of AWE WCF.

## Conclusion and Implications

This study focused on ESL university students' engagement with WCF generated by e-rater. Our results point to several factors mediating learners' engagement with WCF. While behavioral, cognitive, and affective engagement are interrelated and dynamic, we think affective engagement is the most important factor in shaping learners' initial engagement with AWE feedback. To maximize learners' engagement with automated feedback, teacher intervention may be necessary to change learners' negative affective engagement. Teacher intervention can also help inform students of the inaccuracies in AWE feedback to avoid confusion about AWE feedback and to examine and respond to AWE feedback. We also noticed that not every instance of AWE feedback was accurate and should not be implemented without a critical eye. This means that teachers should emphasize the importance of critical evaluation of AWE feedback among their students. Teachers should, therefore, encourage students to use other sources to verify the accuracy of e-rater feedback. Overall, e-rater can be used as an additional source of feedback, as some students may prefer other sources of WCF. Ultimately, teacher feedback may be necessary as some students find it more authoritative (Ferris, 2006), more motivating (Woodworth & Barkaoui, 2022), and more helpful (Zhang & Hyland, 2018) than AWE feedback. We agree with Woodworth and Barkaoui (2022) that AWE-generated feedback should be combined with teacher feedback. Also, as Saeli and Cheng (2021) conclude, teacher WCF may be most effective when provided after AWE WCF,

mainly because some learners may still seek their teachers' confirmation on the accuracy of AWE WCF.

Saeli, H., Rahmati, P., & Koltovskaia, S. (2023). Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation. *Journal of Response to Writing*, *9*(2), 1–39.

## References

Attali, Y. (2004 , April 13–15). *Exploring the feedback and revision features of Criterion* [Paper presentation]. National Council on Measurement in Education, San Diego, CA.

Bitchener, J., Young, S., & Cameron, D. (2005). The effectiveness of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

Burstein, J., & Chodorow, M. (1999). Automated essay scoring for nonnative English speakers. *Proceedings of the ACL99 Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing*. ACL Anthology. https://aclanthology.org/W99-0411/

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*(3), 267–296. https://doi.org/10.1016/S1060-3743(03)00038-9

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing, 32*(3), 385–405. https://doi.org/10.1177/0265532214565386

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. http://dx.doi.org/10125/44145

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Sage Publications. https://doi.org/10.4135/9781452230153

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

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, *10*(2), 121–142. https://doi.org/10.6018/ijes/2010/2/119231

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). 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. https://doi.org/10.1017/S0272263109990544

Ellis, R., Sheen, Y., Murakami, M., & Takashima. H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, *36*(3), 353–371. https://doi.org/10.1016/j.system.2008.02.001

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, *32*(2), 181–201. https://doi.org/10.1017/S0272263109990490

Ferris, D. R., & Roberts, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*(2), 161–184. https://doi.org/10.1016/S1060-3743(01)00039-X

Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 1–43. https://doi.org/10.1080/09588221.2022.2033787

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. https://ejournals.bc.edu/index.php/jtla/article/view/1625

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. https://doi.org/10.1016/j.

jslw.2015.08.002

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

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

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.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, *10*, 277–303. https://doi.org/10.1016/S1060-3743(01)00046-7

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. https://doi.org/10.1080/01443410.2015.1136407

Saeli, H., & Cheng, A. (2021). Peer feedback, learners' engagement, and L2 writing development: The case of a test-preparation class. *TESL-EJ, 25*(2), 1–18. https://files.eric.ed.gov/fulltext/EJ1314054.pdf

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, *42*, 1–16. https://doi.org/10.1016/j.

compcom.2016.05.001

Stevenson, M., & Phakiti. A. (2019). Electronic resources for feedback. In K. Hyland, *Feedback in second language writing* (pp. 125–142). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, *32*, 303–334. https://doi.org/10.1017/S0272263109990532

Vojak, C., Kline, S., Cope, B., McCarthey, S., & Kalantzis, M. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition*, *28*, 97–111. https://doi.org/10.1016/j.compcom.2011.04.004

Woodworth, J., & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom. *TESL Canada Journal*, *37*(2), 234–247. https://doi.org/10.18806/tesl.v37i2.1340

Zhang, Z. V. (2017). Student Engagement with computer-generated feedback: A case study. *ELT Journal, 71*(3), 317–328. https://doi.org/10.1093/elt/ccw089

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z. V., & 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

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. https://doi.org/10.1016/j.asw.2018.03.001