

## Trust in Automated Writing Evaluation (AWE): Multiple case study on feedback engagement and literacy

Nur Arifah Drajati <sup>1\*</sup>, Anis Handayani <sup>2</sup>

<sup>1</sup> Department of English Education, Universitas Sebelas Maret, Indonesia

[nurarifah\\_drajati@staff.uns.ac.id](mailto:nurarifah_drajati@staff.uns.ac.id)

\* Corresponding Author

<sup>2</sup> Department of English Education, Universitas Negeri Yogyakarta, Indonesia

[anishandayani@uny.ac.id](mailto:anishandayani@uny.ac.id)

DOI: <http://dx.doi.org/10.26418/jeltim.v6i2.83172>

### Abstract

Artificial Intelligence (AI) for educational purposes has recently been a research interest. Specifically for academic writing, AI-based feedback, often called automated writing evaluation (AWE), is an essential aspect to support students' writing. However, despite its popularity, the trust in AWE, which potentially contributes to the students' writing improvement, has been explored on a limited basis. Hence, this multiple case study attempts to examine the trust in AWE and how this trust influences feedback engagement and literacy. Of 42 students joining a writing class, a total of 4 undergraduate students were involved to explore this issue. Considering the Indonesian government's policy to conduct online learning amidst the COVID-19 pandemic era, classroom observation, stimulated recalls, and semi-structured interviews were conducted virtually via Zoom meetings to gather the data. Results show that the students' trust in AWE varied, indicated by their varied responses to AWE. Furthermore, their trust in AWE only affects the behavioral aspect, excluding the cognitive and affective aspects. Meanwhile, in feedback literacy, their trust only affected the aspect of appreciating the feedback. These findings imply that educators should pay interest to students' trust in AWE since it, in some ways, influences improvement of their feedback engagement and literacy.

**Keywords:** academic writing; automated writing evaluation (AWE); feedback engagement; feedback literacy; feedback trust

**How to cite this paper:** Drajati, N. A., & Handayani, A. (2024). Trust in Automated Writing Evaluation (AWE): Multiple case study on feedback engagement and literacy. *Journal of English Language Teaching Innovations and Materials (Jeltim)*, 6(2), 125-144. <http://dx.doi.org/10.26418/jeltim.v6i2.83172>

To link to this article: <http://dx.doi.org/10.26418/jeltim.v6i2.83172>



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

Technology development has influenced educational sectors, including students' learning process. As with the impact of prior COVID-19 pandemic, teaching and learning has adapted to technologically-based learning processes, as well as in writing class. Technologies supporting academic writing have been studied for years. Technological tools such as Grammarly (Ranalli, 2021), Criterion (Ranalli et al., 2016), and Pigai (Bai & Hu, 2016) have been proven to support academic writing. As academic writing is closely related to writing feedback, those tools offer an essential feature to support the writing process by providing automated feedback or automated writing evaluation (AWE) which is expected to ease the students' writing process and support writing improvement independently.

Prior studies demonstrated that AWE supports students' learning and writing improvement (Hojeij, 2022; Ranalli, 2021). This kind of feedback has been proven to have some potential for language learning such as noticing inaccurate writing as well as improving the rhetorical writing quality (Ranalli, 2021; Wilken, 2018). Furthermore, students' willingness to make use of the feedback improved when they received explicit feedback such as vocabulary enhancement feedback from Grammarly (Ranalli, 2021). However, despite their beneficial use of AWE, some students doubt the accuracy of this automated feedback (Koltovskaia, 2020; Zhang, 2020; Zhang & Hyland, 2018) which results in ignoring the AWE feedback (Ranalli, 2021). Li et al. (2014) revealed that students rely more on teachers' feedback than AWE feedback. It may happen since AWE is a tool that does not cover human-human trust. Also, high expectations of the tool are associated with high trust in the tool (Lyons & Guznov, 2018). In contrast, one may abandon using the tool when it does not meet their expectation (Ranalli, 2021). It seems that the tools' performance and the students' initial expectations have an impact on their trust and engagement with the tools.

Engagement in written feedback has been studied by researchers (Gravett & Winstone, 2019; Koltovskaia, 2020; Ranalli, 2021; Zhang, 2020; Zhang & Hyland, 2018). Feedback engagement, which can be interpreted as students' responses to such feedback (Ranalli, 2021), has become a critical key to achieving successful feedback. Furthermore, feedback engagement can be influenced by students' individual characteristics and contextual factors (Zhang, 2020). Liu & Yu (2022) reported that students are more engaged in indirect feedback than direct feedback, but the percentage of their revision on direct feedback is higher than in indirect feedback. Furthermore, they suggested that AWE needs the teacher's assistance to confirm the feedback accuracy, also found by Saricaoglu (2021). Jiang & Yu (2020) strengthen the claim that the impact of AWE on students' learning and writing quality is highly affected by the teacher's willingness to scaffold and provide the contextual correction.

Students' trust and engagement in feedback is highly believed to affect feedback literacy, which is one of predictors in the success of feedback utilization. Carless & Boud (2018) defines feedback literacy as the need for "understanding, capacities, and dispositions" (p.1) to make sense of the feedback and use it effectively. Furthermore, they show that feedback literacy has four features namely (1) appreciating feedback–understanding and appreciating the feedback to improve writing quality, (2) making judgments–developing the capability to give academic judgments to their own and other's writings, (3) managing affect–maintaining positive imposition when receiving critical feedback, and (4) taking action–aware of the obligation to revise the writing after receiving the feedback. Students need to be involved in the frequent activity of generating and receiving feedback to develop their feedback literacy (Carless & Winstone, 2020). It is indicated that their activities in giving feedback to their peers and receiving it from other sources, such as AWE, help them foster their feedback literacy. Also, students' trust in AWE potentially affects their feedback literacy, as it determines how they perceive the feedback (Han, 2019; Yu & Liu, 2021). Zhang (2020) reported that students' engagement with AWE enhanced their feedback awareness, understanding, and acceptance. Students' feedback literacy can further be observed from their behaviors toward feedback given by writing assistant applications.

As a new platform in AI-based applications, Scribo takes a role in supporting academic writing with its features to facilitate the teaching and learning process. This website-based application developed by Literatu allows teachers to create classes that can be joined by their students, similar to the features of Google classroom. However, unlike Google classroom, Scribo also has a feature for checking students' writing accuracy by providing immediate and automated feedback covering many aspects of writing: usage, spelling, grammar, passive voice, cohesion, vocabulary, sentence, and readability. Furthermore, Scribo interestingly has more detailed feedback than those provided by other grammar-based applications. It also has features in which users may take a test and get a grade from the application which can be used to measure their writing improvement. Focusing on the feedback feature, it has the potential to support the academic writing process.

However, as many studies reported, AI-based applications, including Scribo, potentially deal with users' issues of trust that affect their feedback uptake. Chapelle et al. (2015) proved that distrust of AWE can arise due to inaccurate feedback given by the tool. It indicates that students' trust is one key factor in successful writing feedback. Besides, students' engagement with the feedback has been proven to influence students' success in using the feedback, specifically for those with high English proficiency (Zhang & Hyland, 2018). Furthermore, Zhang (2020) reported that individual differences and contextual factors influence students' feedback engagement. Feedback engagement is one of the factors influencing feedback literacy (Wood, 2021) –“an ability to read,

interpret, and use the written feedback” (Sutton, 2012, p.31). This somehow indicates that the more engaged the students are in feedback, the more gain they will receive in their feedback literacy.

Reviewing this prior literature, researchers studied the trust issues with AWE (Chapelle et al., 2015; Ranalli, 2021), feedback engagement (Zhang, 2020), and feedback literacy (Wood, 2021) concerning the success of writing feedback uptake. However, studies on how trust issues influence feedback engagement and literacy, which is closely related to successful feedback uptake and writing improvement, have remained unexplored, specifically in the EFL context. Responding to this gap, this study attempts to explore the trust issues with AWE provided by Scribo for EFL students and its influence on their feedback engagement and literacy. Our research explicitly seeks to discern the ways students respond to automated writing evaluation (AWE), in what ways issues of trust influence their feedback engagement, and in what ways issues of trust influence their feedback literacy.

## METHOD

### Research context and participants

We adopted a multiple case study as each participant constitutes the unit of analysis and each case was bounded by the individual and the duration of his or her participation in the research. This study was conducted in an Academic Writing class for English language education in a large university in Indonesia. Six students were selected based on their level of English proficiency, computer self-efficacy, and prior experience using the Scribo application. To represent the level of English proficiency, we selected two students with higher scores in their previous writing class (Mei and Aldi) and another two with lower scores (Dedi and Icha). This participant arrangement was then chosen to explore the relation of English proficiency, trust issues of feedback, and their response to the feedback as further analysis of Zhang & Hyland's (2018) finding that students with high English proficiency benefit more from AWE than those with low English proficiency. In obtaining students' computer self-efficacy, an online survey were given to them, adopting the ICT self-efficacy scale framework of Li et al. (2023).

**Table 1.** The Participants' Demography.

| Participants | Gender | Age | Academic writing score | ICT self-efficacy | Prior experience in using Scribo |
|--------------|--------|-----|------------------------|-------------------|----------------------------------|
| Mei          | F      | 19  | 92                     | High              | Yes                              |
| Aldi         | M      | 21  | 87                     | Intermediate      | No                               |
| Dedi         | M      | 19  | 75                     | High              | No                               |
| Icha         | F      | 21  | 65                     | Low               | Yes                              |

In approximately three months (eight meetings), the students were given a premium account for accessing Scribo. Their ages were in the range of 18-23 years old. The scores presented in Table 1 are from their previous writing classes in the third semester. Table 1 presents detailed information about the participants. All names are pseudonyms.

The students in this class were expected to learn how to write a summary of academic articles from reputable journals. This class consisted of eight meetings and was conducted virtually, supported by the Scribo and Zoom meeting applications. Table 2 shows the students' activities during those eight meetings.

**Table 2.** The students' activities in the Academic Writing class

| Meeting | Activity                                                                                                              |
|---------|-----------------------------------------------------------------------------------------------------------------------|
| 1       | Learning how to operate Scribo briefly and how to make a summary of academic articles from the lecturer's explanation |
| 2       | Searching for some articles from reputable journals based on the student's research interest                          |
| 3       | Writing, checking, and revising their summary of the first chosen research article in Scribo                          |
|         | <i>Outside class activities:</i> Exchanging their writings with peers and giving feedback on their peers' writings    |
| 4       | Presenting and discussing the results of the peer feedback in zoom meeting                                            |
| 5       | Lecturer giving feedback on the students' writings                                                                    |
| 6       | Writing, checking, and revising their summary of the second chosen article from a reputable journal in Scribo         |
|         | <i>Outside class activities:</i> Exchanging their writings with peers and giving feedback on their peers' writings    |
| 7       | Presenting and discussing the results of the peer feedback in zoom meeting                                            |
| 8       | Lecturer giving feedback on the students' writings                                                                    |

## Data collection

Amidst the Covid-19 pandemic, we collected the data virtually, facilitated by Zoom meetings. Before we began our data collection, we asked for participants' consent by sending emails informing them of our research aims, the kinds of data we needed, and how the data were collected, kept, and analyzed. After they gave their consent, we observed their classroom activities in the Academic Writing class for eight meetings. The observation focused on (1) time spent in checking texts in Scribo, (2) contents of the feedback they chose, and (3) their decision to accept or ignore the feedback given. As the observation finished, we then scheduled a Zoom meeting for six chosen participants according to their available time. The screen recording videos dealt with how they responded to



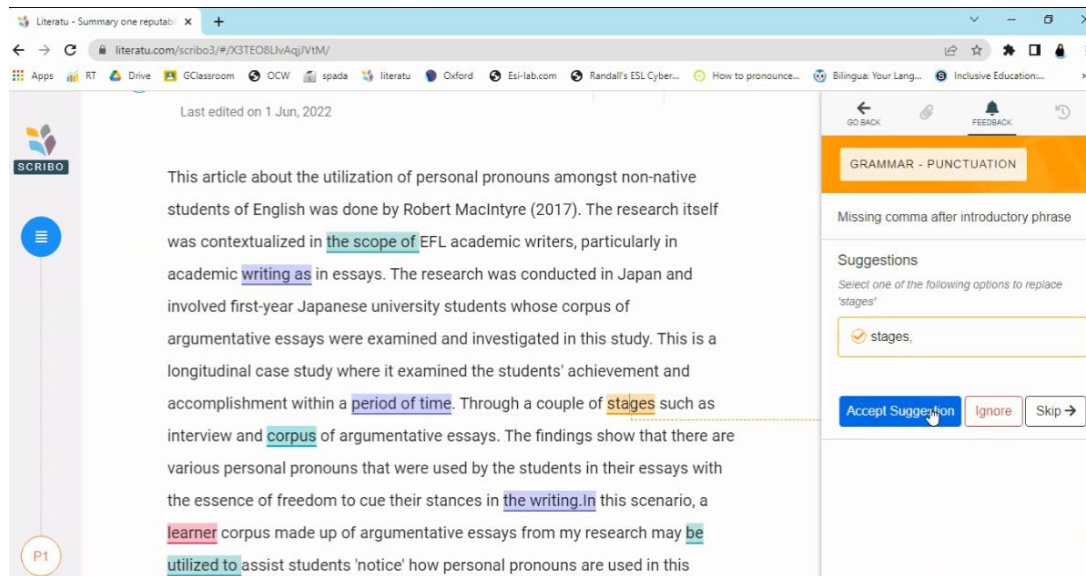
the automated feedback given by Scribo. The texts used were the summary of the first chosen article from a reputable journal they wrote in the third meeting, consisting of 100 – 200 words. Each participant was observed separately in different Zoom meetings and time allocations. While the participants were checking their texts in Scribo, we asked several questions about why they accepted, ignored, or skipped the feedback. This observation and stimulated recalls lasted for about 30-40 minutes for each participant.

After the observation and stimulated recalls finished, they were then, interviewed regarding their feedback engagement and feedback literacy. We adapted the concept of feedback engagement from Ellis (2010) and feedback literacy from Carless & Boud (2018) to construct the questions. These semi-structured interviews lasted for 20-30 minutes for each participant. We recorded the Zoom meetings to avoid data loss and ease the transcription and data analysis process.

### **Data coding and analysis**

The screen recording data were coded manually by two researchers to avoid research bias and analyzed using thematic analysis. For the data coding, we followed the feedback categories of Scribo, which consisted of grammar and spelling, vocabulary, sentences, topic coverage, and cohesiveness. However, since we did not control the participants in choosing which feedback they wanted to have, we only used the most chosen feedback categories, namely (1) grammar and spelling (henceforth GS), (2) vocabulary (V), and (3) cohesive (C) as our themes for analyzing the data. We two researchers, separately analyzed the screen recording videos by counting the number of feedback messages received by the participants and in which ones they accepted, ignored, or skipped the feedback, and analyzed their reasons for doing so. Since the responses of “ignore” and “skip” have the same action meaning that they did not employ the given feedback, we used the term “ignore” to simplify our interpretation. Thus, we used the term "accept" when the participants accepted and used the feedback and the term "ignore" when they ignored the feedback. We then, discussed the results and wrote the report in agreement. Figure 1 is an example of the GS category received by participants.

Audios of stimulated recalls and semi-structured interviews were transcribed using the Sonix application. To ensure the accuracy of the transcription, we reviewed the transcripts by comparing the transcripts and the recordings. As the accuracy was guaranteed, a thematic analysis was adopted to analyze the data. We first coded the interview transcript under themes of feedback trust (henceforth FT), feedback engagement (Ellis, 2010): cognitive (CE), affective (AE), and behavioral engagement (BE), and feedback literacy (Carless & Boud, 2018): appreciating the feedback (AF), making judgments (MJ), managing



**Figure 1.** The example of GS category feedback

affect (MA), and taking action (TA). We two researchers, separately coded and analyzed the data. We then discussed the results to assimilate the results. After an agreement was achieved, we concluded the results.

## FINDINGS

### EFL students' response and trust in automated writing evaluation (AWE)

After the participants learned how to make a summary of academic articles from reputable journals, they wrote and checked their summaries in the AI-based application Scribo. In our study, we used their second summary in which they learned and became familiar with the application before analyzing their responses to this automated writing evaluation (AWE). In checking their texts in the application, they were given the freedom in choosing the feedback they wanted to receive. Due to various reasons, they mostly checked their texts regarding grammar and spelling, vocabulary, and cohesiveness. Table 3 summarizes the participants' frequency of the received feedback.

**Table 3.** The frequency of the feedback received

| Categories         | Low proficiency |      | High proficiency |      |
|--------------------|-----------------|------|------------------|------|
|                    | Dedi            | Icha | Mei              | Aldi |
| Grammar & Spelling | 55              | 10   | 13               | 14   |
| Vocabulary         | 5               | -    | 4                | -    |
| Cohesive           | 3               | 5    | 6                | -    |

In responding to this AWE, they have various responses and reasons behind their decision to accept or ignore the feedback. Since each student may have a

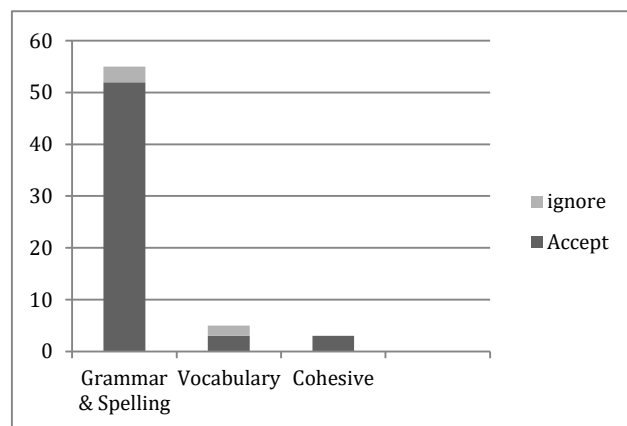
different trust level and various reasons, we decided to explore each participant's trust in AWE and his/her reasons behind his/her decision to accept or ignore the AWE.

### Dedi

In responding to AWE, Dedi only accepted some feedback that he thought was correct. He revealed that he always checked the context first before deciding to accept or ignore the feedback. Also, he felt that some feedback may be incorrect because AWE is automatically given by an AI which probably does not consider the context as seen in the excerpt Dedi-FT as follows:

*"I only accept the feedback when I think it is correct as I experienced that sometimes this application gives contextually incorrect feedback."* Dedi-FT

Despite his low trust in this AWE, he accepted numerous AWE feedback messages (92%) from this application. He reported that this application is mostly helpful in correcting his grammar and spelling, but not in other aspects. Figure 2 shows Dedi's response toward AWE from the application.

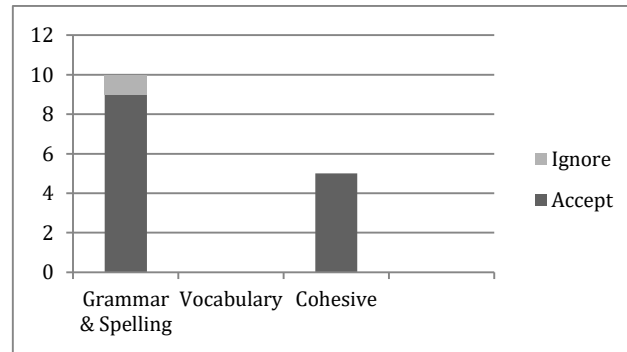


**Figure 2.** Dedi's response to AWE

### Icha

Icha only checked her text regarding grammar and spelling (GS) and cohesiveness (C). Furthermore, in responding to the feedback, she mostly accepted the suggestions (see Figure 3). Here, she revealed that the application helped her improve her writing quality by giving corrective feedback automatically. She realized that her writing skill is low, so she mostly accepted the feedback (93%) without further thinking, specifically in the cohesiveness category.





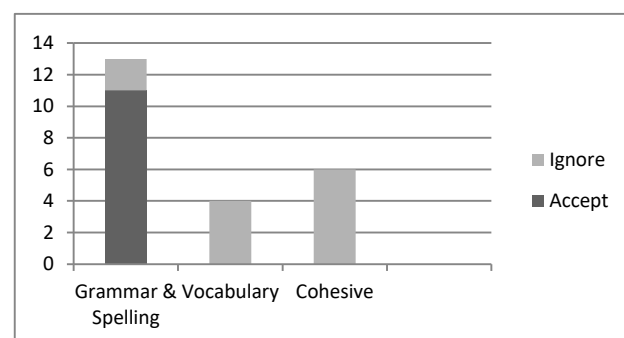
**Figure 3.** Icha's response to AWE

As a result of her rather strong trust in this AWE, she only ignored one feedback message, which was the suggestion to change the passive voice into an active voice. She said that changing the passive voice to active would change the meaning of the sentence. Thus, she decided to ignore the feedback. It was revealed in the excerpt Icha-FT.

*"The feedback is to change the sentence into I conclude or we conclude, but this article is not mine, so I can't use this feedback. It will make the sentence carry a different meaning."* Icha-FT

### Mei

As a student with high proficiency in English, she checked all chosen categories of the feedback given by the application. However, Mei tends to believe in her own ability rather than a mere website-based application. She revealed that this AI-based application can probably be wrong in giving feedback, so she only accepted a few feedback messages (47%) when she highly believed that they were correct, specifically in the context. Consequently, she ignored the feedback more frequently than she accepted it (see Figure 4).



**Figure 4.** Mei's response to AWE

As an example, she ignored the feedback to change the word "and" in the text to avoid repetition. She decided to ignore it since she believed that her use of

“and” was correct and did not need any change. Similarly, she also ignored the cohesiveness feedback since she did not trust AWE feedback after experiencing incorrect feedback from a similar application. The excerpt Mei-FT shows her change of attitude toward AWE due to her previous experiences.

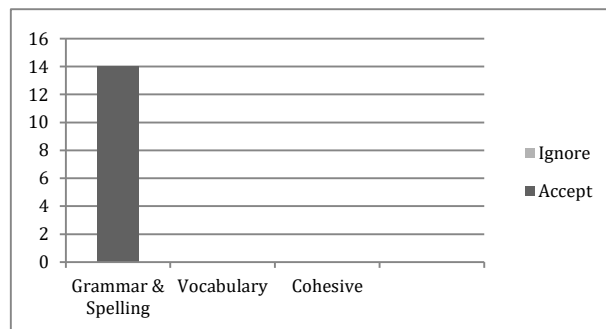
*“The use [of] and in my text is already contextually correct, why should I need to change it? Also, in cohesive, I often ignored the feedback because previously I received obviously incorrect feedback from a similar application. It seems like I lost my trust in AI-based feedback.” Mei-FT*

### Aldi

Aldi is typically a student who merges what he knows and the feedback received. In checking his text, he accepted the feedback to change some passive voices into active voices since he remembered his lecturer told him to use active sentences. Similar to other participants, he always tried to comprehend and match the context before accepting feedback. It was revealed in the excerpt Aldi-FT as follows:

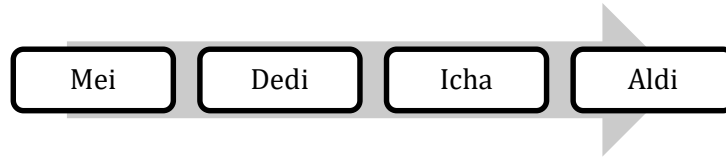
*“I accept it (to change some passive sentences into active ones) since my lecturer told me to use active sentences, specifically in writing the summary.” Aldi-FT*

However, he only checked his text regarding the grammar and spelling category. He accepted all the feedback received since he knew that the feedback was contextually correct and he made mistakes in grammar and spelling (see Figure 5).



**Figure 5.** Aldi’s response to AWE

Considering their results of the stimulated recalls indicating that they frequently check the accuracy of the feedback before accepting it, we may conclude the levels of their trust according to their acceptance frequency of the feedback given. Accordingly, Aldi is found to have the highest trust in AWE while Mei has the lowest trust. Figure 6 illustrates the participants’ trust level of AWE.



**Figure 6.** The level of participants' trust

### **Ways the trust level of AWE influences feedback engagement**

As we aimed to identify the time spent by students in engaging in the feedback, the students had flexible time to check their writing with the Scribo writing checker until they found their writing ‘appropriate’. Table 4 summarizes the time spent by students in checking their summaries consisting of 100-200 words.

**Table 4.** Time spent in engaging with the feedback

| Students | Time spent in the first checking |
|----------|----------------------------------|
| Aldi     | 11 minutes and 15 seconds        |
| Icha     | 10 minutes and 7 seconds         |
| Dedi     | 12 minutes and 21 seconds        |
| Mei      | 7 minutes and 18 seconds         |

As presented in Table 4, their trust in AWE did not actually determine the amount of time spent in checking the writing. For instance, Dedi spent more time than Icha, who has higher trust. Also, we can identify a notable impact of trust on Mei who had the lowest trust level and spent the least amount of time to check her writings.

Aside from the time spent, we can also break down feedback engagement into three types: cognitive, affective, and behavioral engagement. Regarding cognitive engagement, Icha who has a rather high trust in AWE among the participants, reported that she comprehended the feedback well since the application provided further explanation of the feedback in detail. In checking her text, she read the feedback more than once to comprehend the feedback completely (see the excerpt Icha-CA).

*“I did not find any difficulty in understanding the feedback. I mean all the explanation was there. We can just read it several times if necessary to understand the feedback.”* Icha-CA

Regarding affective engagement, despite their low or high trust in AWE, the participants were mostly glad in receiving critical or “bad” feedback. They revealed that receiving “bad” feedback was not that bad since they felt that they

were novice writers who needed corrective and helpful feedback. It was revealed by Dedi in the excerpt Dedi-AE.

*"I am glad (in receiving any kind of feedback including the bad ones). Since this feedback is from a technology (not human) and for my better improvement, I am not feeling down or sad in receiving the feedback."* Dedi-AE

Last, regarding behavioral engagement, low trust in AWE affected their motivation to revise their texts. This phenomenon happened to Mei, the one with the least trust of AWE among the participants. She reported that this feedback did not really motivate her to revise her text since she considered the feedback given by AI-based applications was often inaccurate (see excerpt Mei-BE).

*"Honestly, this AWE did not really motivate me to revise my text. This kind of AI-based application often gives me inaccurate feedback, so I prefer peer or teacher feedback over AWE."* Mei-BE

Accordingly, high trust in AWE did motivate the receivers to revise their text. Aldi, one having the highest trust of AWE, was motivated to revise his text after receiving AWE. Excerpt Aldi-BE revealed that this AI-based application helped him check his text automatically so that he could improve his text quality without help from other people.

*"I was quite helped in improving my text quality from receiving AWE from this application. When receiving the AWE, I understand which parts are wrong, so I can fix them and make my writing improved. This feedback gives me the motivation to revise and improve my writing quality."* Aldi-BE

In sum, the results of the interviews reveal that students' trust on AWE did not affect their cognitive and affective aspects of feedback engagement, as they mostly show good comprehension of the feedback and positive emotion in receiving the feedback (see Table 5). Furthermore, the only aspect affected by their trust in AWE is the behavioral aspect. This indicates that as the students have higher trust levels, they potentially acquire high motivation to revise their work after receiving the feedback.

**Table 5.** The summary of students' feedback engagement

| Students | Feedback engagement |           |            |
|----------|---------------------|-----------|------------|
|          | Cognitive           | Affective | Behavioral |
| Aldi     | Yes                 | Yes       | Yes        |
| Icha     | Yes                 | Yes       | Yes        |
| Dedi     | Yes                 | Yes       | No         |
| Mei      | Yes                 | Yes       | No         |

## Ways the trust in AWE influences feedback literacy

Adapting the theory of feedback literacy from Carless & Boud (2018), we see the participants' feedback literacy from the four aspects: appreciating the feedback, making judgments, managing affect, and taking action (see Table 6). The participants mostly have their feedback literacy improved, which can be seen from the observation of the sixth and seventh meetings of the class. After receiving AWE, all the participants, except Mei, mostly consider AWE as one of the efforts in improving their writing quality. They did not think that this AWE was there to judge their writings, but rather to help them check and improve the writings. Aldi, with high trust in AWE, has a very positive thought of AWE as a helpful tool (see excerpt Aldi-AF).

**Table 6.** The summary of students' feedback literacy

| Students | Feedback literacy         |                  |                 |               |
|----------|---------------------------|------------------|-----------------|---------------|
|          | Appreciating the feedback | Making judgments | Managing affect | Taking action |
| Aldi     | Yes                       | Yes              | Yes             | Yes           |
| Icha     | Yes                       | Yes              | Yes             | Yes           |
| Dedi     | Yes                       | Yes              | Yes             | Yes           |
| Mei      | No                        | Yes              | Yes             | Yes           |

*"Yes, this AWE is very helpful. It does help in checking and improving my writing quality."* Aldi-AF

As predicted, Mei, who has the lowest trust in AWE, had a completely different thought. In excerpt Mei-AF, she revealed that AWE did not really help her improve her writing since she had all the knowledge to improve her writing from her previous class, even before receiving this AWE. However, she agreed that this AWE helped her enrich her vocabulary.

*"This AWE did not really help me in improving my writing since I got the theory of improving my writing from my previous class (before checking with this AI-based application)."* Mei-AF

Surprisingly a different case happened from the making judgments aspect. Despite their trust in AWE, all participants felt that their confidence and insights into how to make judgments of texts improved after receiving AWE from this application. Furthermore, the AWE from this application covers many aspects of language such as grammar, spelling, vocabulary, sentences, cohesive, and even topic coverage. Excerpt Dedi-MJ revealed his confidence in making judgments after receiving AWE feedback.



*"Yes, since this application gives feedback just like a mentor showing our mistakes (in the writings). Here we can take an example from this feedback giving, for example, when we have to give a peer review. So, I think this AWE enriches my knowledge on how to give feedback to others."* Dedi-MJ

Their improved feedback judgment was demonstrated in the second peer feedback activity in the sixth and seventh meetings. While students only focused on the grammar aspect in the first peer feedback activity, they make a different judgment of their peers' writings in the second peer feedback activity. They pointed out the vocabulary, cohesiveness, and readability of their peers' writings.

Similarly, for managing affect, their trust in AWE did not influence their ability to manage their emotion in receiving feedback, even for "bad" feedback. After experiencing writing and checking their texts with this application, they considered "bad" feedback as an aid to help them improve their writing skills. They were not sad or "down" to receive such feedback since they knew it was for their writing improvement. Furthermore, they were rather motivated to improve their writing quality after receiving this kind of feedback. Excerpt Dedi-MA revealed that Dedi was motivated to revise and do better in his writing after receiving AWE feedback.

*"The feedback I got from Scribo (even the bad ones) motivates me to do better in my writings."* Dedi-MA

For taking action aspects, all participants agreed that AWE made them revise their text right after receiving the feedback. They said that since AWE was typically corrective feedback, they directly revised their text after receiving the feedback. For instance, it happens to Aldi in which he frequently revised his writing directly after receiving the AWE feedback (see excerpt Aldi-TA).

*"I directly revise my text, but I will recheck my text to make sure that the feedback is appropriate."* Aldi-TA

## DISCUSSION

To put it briefly, our results show that students have varied responses to AWE feedback. This result is probably expected since individuals have differences in mindset as well as the level of trust in the AI-based application. Furthermore, despite their English proficiency, the students were somehow selective in accepting or ignoring the feedback, supporting Bai & Hu's (2016) and Tian & Zhou's (2020) research. They comprehend the feedback first, suit it with the context, and accept it if they think the feedback is correct. This finding also strengthens the claims of Chapelle et al. (2015) that students do not take AWE

feedback blindly. Considering the students' responses to AWE, we can note that their proficiency in English does not really affect their trust in AWE. This result is somehow in contrast with Huang & Renandya's (2018) conclusion that low proficiency students think highly of AWE.

Furthermore, our results interestingly show that their previous experience with AWE influences their trust more than their English proficiency. This case happened to Mei who has previous experience with AWE from a similar application. After receiving several inaccurate feedback messages, she has low trust in AWE. She believes that automation cannot be trusted 100% since it is not superior to human intelligence. This finding somehow supports Cope et al.'s (2020) that AI cannot surpass human intelligence. Also, it strengthens Chapelle et al.'s (2015) conclusion that distrust may appear from inaccurate feedback given by an application, which indicates that negative experience from an application potentially decreases their trust in the application.

Regarding feedback engagement, students here have similar feedback engagement despite their English proficiency. This finding is somehow in contrast with the previous studies (Koltovskaia, 2020; Zhang, 2020), which reported that higher proficiency students have higher feedback engagement. Furthermore, rather than their English proficiency, their beliefs and trust in AWE contributed more to the feedback engagement. For instance, Icha, who has high trust in AWE, finds it easy to comprehend the feedback given since she decided to read the feedback and revise her text several times with AWE support, spending a longer time to engage with the feedback given. Furthermore, Ranalli (2018) revealed that students' trust in AWE may decrease due to the limited capacity of the application to provide expected feedback.

Our study also gets similar results that the students here do not trust the feedback since this AI-based application is restricted in giving contextually accurate feedback which may be provided by peer or teacher feedback. Consequently, the students sometimes get confused with the provided feedback. Furthermore, the fact that they cannot ask for any further information from the tools does not help them comprehend the feedback (Ranalli, 2018). However, still, at some points, this feedback helps them improve their writing quality. C. Lee (2019) also reported that AWE stimulated students to enrich their writing quality both in content and language.

A notable finding was spotted in relation to the amount of time spent in using Scribo. Mei with the lowest trust spent the least amount of time to check her writings. This means that students' trust indirectly affects students' feedback engagement, which supports Zhai & Ma's (2021) belief that students' trust influences their judgment on the usefulness of AWE feedback, potentially leading to shorter time spent in using the application. Application developers should put particular attention on the feedback experiences brought by the application as it highly affects the students' success in utilizing the given feedback, as reported by Zhang & Hyland (2018).

Our results also show that students' trust in AWE, in some ways, influenced their feedback literacy. However, this phenomenon only occurred in one aspect of feedback literacy: appreciating the feedback. This may imply that students with high trust in AWE appreciate the feedback more than the ones with low trust in AWE. As they appreciate the feedback, they will be more motivated to revise their writing, which may result in long-term accuracy of their writing (Link et al., 2020). Furthermore, Han & Xu (2021) reported that feedback literacy is multifaceted and that it is highly influenced by individual differences. It seems to be the most possible cause for the students to have varied feedback literacy after receiving AWE. As Jiang & Yu's (2020) research revealed the role of teachers' assistance in utilizing AWE, it suggests that the same teachers' assistance is needed here to optimize the impact of AWE, specifically in their feedback literacy.

Our finding also revealed that past experiences on feedback does not necessarily affect students' feedback literacy. All students in this study fulfilled their criteria of having good feedback literacy. This finding is somehow in contrast with Malecka et al. (2022) that past experiences give different impacts on students' behavior in utilizing feedback. These different results likely occurred due different kinds of past experiences between this study's participants and Malecka et al.'s (2022). While the participants have positive and supportive experiences with feedback in Malecka's study, the participants here have negative and unsatisfied experiences with feedback, so that it leads to more negative attitudes and trust during later feedback experiences. This particularly supports Walker et al.'s (2020) finding that negative past experiences demotivate students in revising their works in future events.

Viewed from our results, we can see that trust in AWE partly influences the students' feedback engagement and feedback literacy. This result somehow sheds light on ideas for educators to increase their students' trust in AWE since it influences students' engagement and literacy of feedback as well as their writing improvement.

## CONCLUSION AND SUGGESTIONS

We may summarize that students' trust in AWE may be varied depending on their experiences, personality, and beliefs. Also, it comes to light that trust in AWE did influence students' feedback engagement and feedback literacy, although only in some aspects such as behavioral engagement and appreciating the feedback aspects of feedback literacy. Furthermore, our finding also indicates that whereas AI gives feedback more quickly and automatically, teacher guidance and facilitation are still needed to make use of this AI more efficiently and accurately. However, since we analyze the data qualitatively, we cannot

prove this proposition statistically accurate. To have a more accurate result on this issue, further analysis is warranted.

This study only used qualitative data and small data sources. Thus, to achieve more accurate and generalizable results, further studies may add quantitative data and analysis to complete the results. Also, studying similar topics with different research methods may be worthwhile to compare the results. Furthermore, this study examining the students' trust in feedback may be further researched by studying their trust in other types of feedback such as peer or teacher feedback. Last, since the students' trust in AWE feedback partially affects feedback engagement and literacy, further research on the factors affecting their trust may be interestingly needed to complete the literature on feedback and automation.

Despite our research limitations, our study implies that trust issues in AWE somehow contribute to the students' feedback engagement and feedback literacy. It somehow shed light for teachers on one factor influencing the students' feedback engagement and feedback literacy regarding AWE. It also indicates that students' trust in such feedback is essential in improving their writing quality. Thus, teachers and educational practitioners should then find such a way to improve their students' trust. Furthermore, since students' trust is influenced by individual factors such as students' beliefs and previous experiences, teachers should consider these factors before implementing such feedback for their academic writing class. Last, we expected our study to be initial research leading to other studies related to AWE for other applications, concerning how to make use of this AWE effectively, specifically in Indonesia.

## REFERENCES

- Bai, L., & Hu, G. (2016). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150–163. <https://doi.org/10.1080/13562517.2020.1782372>
- 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>

- Cope, B., Kalantzis, M., Sears, D., Cope, B., Kalantzis, M., & Sears, D. (2020). Artificial intelligence for education : Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229–1245. <https://doi.org/10.1080/00131857.2020.1728732>
- Gravett, K., & Winstone, N. E. (2019). 'Feedback interpreters': The role of learning development professionals in facilitating university students' engagement with feedback. *Teaching in Higher Education*, 24(6), 723–738. <https://doi.org/10.1080/13562517.2018.1498076>
- Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System*, 80, 288–303. <https://doi.org/10.1016/j.system.2018.12.009>
- Han, Y., & Xu, Y. (2021). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*, 26(2), 181–196. <https://doi.org/10.1080/13562517.2019.1648410>
- Hojeij, Z. (2022). Effectiveness of using digital feedback on EFL student writing skills. *International Journal of Computer-Assisted Language Learning and Teaching*, 12(1), 1-18. <https://doi.org/10.4018/IJCALLT.291111>
- Huang, S., & Renandya, W. A. (2018). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching*, 14(1), 15–26. <https://doi.org/10.1080/17501229.2018.1471083>
- Jiang, L., & Yu, S. (2020). Appropriating automated feedback in L2 writing : experiences of Chinese EFL student writers. *Computer Assisted Language Learning*, 35(7), 1329–1353. <https://doi.org/10.1080/09588221.2020.1799824>
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44(April 2020), 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Lee, C. (2019). A study of adolescent English learners ' cognitive engagement in writing while using an automated content feedback system. *Computer Assisted Language Learning*, 33(1–2), 26–57. <https://doi.org/10.1080/09588221.2018.1544152>
- Li, Z., Link, S., Ma, H., Yang, H., & Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System*, 44, 66–78. <https://doi.org/10.1016/j.system.2014.02.007>
- Li, Z., Zuo, T., Wei, X., & Ding, N. (2023). ICT self-efficacy scale: The correlations with the age of first access to the internet, the age at first ownership of a personal computer (PC), and a smartphone. *Medical Education Online*,



- 28(1), 1-8. <https://doi.org/10.1080/10872981.2022.2151068>
- Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback : An eye-tracking study. *Language Learning & Technology*, 26(2), 78–105. <https://doi.org/https://doi.org/10125/73480>
- Lyons, J. B., & Guznov, S. Y. (2018). Theoretical Issues in Ergonomics Science Individual differences in human – machine trust : A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4), 440–458. <https://doi.org/10.1080/1463922X.2018.1491071>
- Malecka, B., Boud, D., Tai, J., & Ajjawi, R. (2022). Navigating feedback practices across learning contexts: Implications for feedback literacy. *Assessment and Evaluation in Higher Education*, 47(8), 1330–1344. <https://doi.org/10.1080/02602938.2022.2041544>
- Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674. <https://doi.org/10.1080/09588221.2018.1428994>
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52(June 2021), 1–16. <https://doi.org/10.1016/j.jslw.2021.100816>
- Ranalli, J., Link, S., & Chukharev-hudilainen, E. (2016). 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>
- Saricaoglu, A. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL*, 33(3), 265–277. <https://doi.org/10.1017/S0958344021000021>
- Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, 49(1), 31–40. <https://doi.org/10.1080/14703297.2012.647781>
- Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback , peer feedback and teacher feedback in an online EFL writing context. *System*, 91(July 2020), 102247. <https://doi.org/10.1016/j.system.2020.102247>
- Walker, R., Oliver, R., & Mackenzie, R. (2020). Interviews with secondary school students: Perceptions of feedback. *Issues in Educational Research*, 30(4), 1576–1595. <http://www.iier.org.au/iier30/walker.pdf>



- Wilken, J. L. (2018). Perceptions of L1 glossed feedback in automated writing evaluation: A case study. *Computer Assisted Language Instruction Consortium*, 35, 30–48. <https://doi.org/https://doi.org/10.1558/cj.26383>
- Wood, J. (2021). Making peer feedback work: the contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment and Evaluation in Higher Education*, 47(3), 327–346. <https://doi.org/10.1080/02602938.2021.1914544>
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48(April 2021), 100525. <https://doi.org/10.1016/j.asw.2021.100525>
- Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback : a systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, 35(9), 2817–2842. <https://doi.org/10.1080/09588221.2021.1897019>
- Zhang, Z. (Victor). (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, 43(January 2020), 100439. <https://doi.org/10.1016/j.asw.2019.100439>
- Zhang, Z. (Victor), & 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>

### Authors' Brief CV

**Nur Arifah Drajadi** is a lecturer of the English Language Education Department of Universitas Sebelas Maret (UNS). She contributed to several reputable journals. Her research interests lie in technology in language learning, TPACK, Multimodal, and IDLE.

**Anis Handayani** is a lecturer of the English Language Education Department of Universitas Negeri Yogyakarta. She contributed to several reputable journals. Her research interests lie in academic writing, educational technology, and teacher professional development.