Regular Paper

The JALT CALL Journa
ISSN 1832-4215 Vol. 20, No.1
https://doi.org/10.29140/jaltcall.v20n1.1200

Castledown

ChatGPT as an automated writing evaluation (AWE) tool: feedback literacy development and AWE tools' integration framework

Imelda Gozali
Universitas Negeri Malang
Universitas Katolik Widya Mandala Surabaya
imelda.gozali@ukwms.ac.id

Alberik Ryan Tendy Wijaya
Universitas Katolik Widya Mandala Surabaya
alberik@ukwms.ac.id

Anita Lie
Universitas Katolik Widya Mandala Surabaya
anita@ukwms.ac.id

Bambang Yudi Cahyono
Universitas Negeri Malang
bambang.yudi.fs@um.ac.id

Nunung Suryati
Universitas Negeri Malang
nunung.suryati.fs@um.ac.id

This study examines the use of ChatGPT, in conjunction with other applications, Grammarly and Quillbot, as an automated writing evaluation (AWE) tool in a Recount and Narrative essay course for 18 English as a Foreign Language (EFL) students in Indonesia. The primary objective was to investigate the impact of these tools on the development of learners' feedback literacy. A single, qualitative case study design was employed, gathering students' voices through a semi-structured interview, reflective journals, and artifacts comprising writing e-portfolios and the feedback evidence. The data analysis utilized an inductive-deductive approach to identify themes and patterns in the qualitative data. The findings revealed that the AWE tools complemented each other in supporting almost all aspects of students' feedback literacy, with "feedback processing"

being the aspect that ChatGPT could potentially enhance or diminish, contingent upon students' feedback-seeking behavior. Furthermore, as the result of the inductive coding of the qualitative data, the study offers an "AWE Tools Integration Framework," namely six elements that educators could consider when incorporating AWE tools, notably generative AI, in their writing classes. This study concluded with a call for greater support of students' digital literacy, equal access to technology, and ethical use of Artificial Intelligence in the classroom.

Keywords: artificial intelligence, assessment, automated writing evaluation (AWE) tools, feedback literacy

Introduction

Since its release in November 2022, ChatGPT (https://chat.openai.com/) has generated a plethora of research interest, not least in the field of language education (Barrot, 2023; Kostka & Toncelli, 2023; Su et al., 2023). Its generative ability and conversational format portend countless potentials in language teaching and learning (Kohnke et al., 2023), although concerns have been raised on its impact on students' academic integrity and over-reliance on its seemingly intelligent responses. Particularly, its role as writing assistants in various fields has been studied and debated with mixed results (Basic et al., 2023; Chen, 2023; Fang et al., 2023; Parker et al., 2023).

Nonetheless, the use of AI-powered tools in the field of English as a Foreign Language (EFL) is not entirely a novelty. EFL learners and non-native English writers have long used AI technology such as Grammarly, Quillbot, ProWritingAid, and the like. Known collectively as automated writing evaluation (AWE) tools, those applications improve writing performance by detecting the location of errors and suggesting the appropriate corrections. Of the myriads of grammar-checker and paraphraser tools available in the market, Grammarly is the most extensively researched (Rudolph et al., 2023). The advent of ChatGPT, which possesses similar grammar-checking capability depending on the prompts given, could potentially represent a transformation in the AWE field.

Another reason why AWE tools have been attracting the attention of assessment scholars and practitioners worldwide is their potential to develop students' feedback literacy (Carless & Boud, 2018). Situated within the shifting paradigm from feedback as the transmission of information to learners' sensemaking and partnership with the teachers (Carless, 2020), students' feedback literacy aims at developing students' disposition and capability to make the most of feedback for self-improvement. From this perspective, feedback is not merely teachers' responsibility but should actively be elicited by the learners (Molloy et al., 2020) from various sources (teachers, peers, technology). It is in this vein that the capabilities of AWE, in its role as automated writing evaluators in students' self-assessment endeavors, should be fully utilized,

particularly in the EFL context, where the fundamental structure of the English language still proves problematic for many.

Within the burgeoning literature on students' feedback literacy, several scholars have examined the mediating role of technology, such as the Canvas LMS or Google Docs, on feedback literacy or its enactment in online settings (Ducasse & Hill, 2019; Wood, 2021, 2022). AWE has had a long research history, from the seminal Project Essay Grade (Page, 2003) to a systematic literature review of AWE systems (Huawei & Aryadoust, 2023). Surprisingly, there is a lacuna of scholarly discussion on the relationship between student feedback literacy and AWE in the EFL settings. Such a study is urgently needed given the fast-growing development of AI-based AWE tools and the necessity to embrace such technology in supporting learner-centered assessment and feedback literacy. Besides, it might prove insightful to explore whether the novelty of ChatGPT as an AWE tool can upstage other more established AI-powered grammar checkers, such as Grammarly or Quillbot.

Given the preceding considerations, this study aims to respond to the conceptual gap in the relationship between AWE tools and students' feedback literacy, which in this study is operationalized as the set of disposition (appreciating feedback, commitment to feedback as improvement, managing emotion) and skills (eliciting, processing, and enacting feedback) needed to maximize the feedback's potentials (Carless & Boud, 2018; Malecka et al., 2022; Molloy et al., 2020). Set within the context of an EFL writing course in higher education, this study is guided by a single line of inquiry: To what extent do the AWE tools (ChatGPT, Grammarly, Quillbot) support the development of EFL students' feedback literacy?

Literature review

Automated writing evaluation (AWE) tools and their roles in feedback provision

While the craze of AI-related tools only rapidly heightened in late 2022 due to the birth of GPT-3 (ChatGPT), which was hailed as AI for the masses, AI has long been utilized in the scope of writing assessment as automated writing evaluation (AWE) tools. Dated back to the 1960s, Project Essay Grade (PEG) was conceived and became the pioneer of AWE tools (Page, 2003). Warschauer and Ware (2006) underscored the utility of AWE tools for teachers who need to manage many students through various drafting processes. Miranty and Widiati (2021) likewise noted the benefit of AWE tools, which provide rapid turnaround of revision in English writing practice.

In recent years, AWE tools have undergone accelerated development with a shift of focus towards formative assessment (Hockly, 2019). Several widely used AWE tools are those conveniently accessible by the masses through the Internet, like Grammarly and Quillbot (Ho, 2022; Ebadi et al., 2023; Miranty et al., 2021). A study done by Tambunan et al. (2022) found that Grammarly helped students who had the most difficulty in grammar to "revise better and proofread their

work" (p. 24). Students would minimize grammatical errors with Grammarly (Ebadi et al., 2023). Another study found QuillBot to effectively help students improve their academic writing (Kurniati & Fithriani, 2022). Furthermore, their participants saw QuillBot as a tool to reduce their writing anxiety and an opportunity to improve their grammatical knowledge.

AWE tools have the potential to be on par with human raters regarding accuracy, validity, and reliability (Huawei & Aryadoust, 2023; Page, 2003). Nevertheless, as AWE tools focus more on the surface features of writing (e.g., grammar, spelling, punctuation), the availability of AWE tools is to complement the teachers, not replace them (Ebadi et al., 2023; Huawei & Aryadoust, 2023; Zhang, 2020). With the rapid turnaround provided by AWE tools in terms of surface-level editing, the teachers can increase the depth of the revision, focusing more on the idea organization and development and consequently enjoy workload relief.

In addition, it must be noted that AWE tools can only go as far as the users' technical know-how. This is especially true for generative AI like ChatGPT. One of its vast abilities is providing feedback for one's writing (Barrot, 2023; Fang et al., 2023; Parker et al., 2023; Schmidt-Fajlik, 2023; Wu et al., 2023). Schmidt-Fajlik (2023) pointed out that ChatGPT is user-friendly, provides detailed reviews, and gives comprehensible explanations due to its conversational nature. It is also capable of translating the feedback to the user's mother tongue. However, it will only do so when prompted by the user (Wu et al., 2023). As the key element of ChatGPT is chatting, there is a need to train the students to develop the most proper and effective prompts to gain their intended results. Thus, Zhang (2020) highlighted the importance for teachers to scaffold the students in using the AWE tools to react to the corrective feedback and go beyond surface-level revision. His findings further underscore the importance of increasing the students' feedback literacy.

Feedback literacy

Research on feedback literacy is gaining currency among feedback researchers (Author1 et al., 2023) due to its radical notion of conceiving feedback processes as a set of skills possessed by individuals and as a sociocultural partnership among teachers and students as agents (Nieminen & Carless, 2023). Initially coined by Sutton (2012) within the ambit of academic literacy, feedback literacy gained mounting interest upon the publication of the student feedback literacy framework by Carless and Boud (2018). This landmark article defines student feedback literacy as "the understandings, capacities, and dispositions needed to make sense of information and use it to enhance work or learning strategies" (Carless & Boud, 2018, p. 1316). Its constitutive elements are appreciating feedback, making judgments, managing affect, and taking action. Subsequent to this seminal work, several scholars built upon the framework through empirical research, validation, or reconceptualization from a fresh perspective. For example, Molloy et al. (2020) conducted a large-scale study involving 4,514 Australian students to determine what constitutes best feedback practices and,

in so doing, validated Carless and Boud's (2018) model and added more categories based on students' voices. The resulting learner-centered framework is represented in seven groupings, as seen in Figure 1. Thus, apart from the various attitudes needed to successfully benefit from feedback, student feedback literacy also comprises the acts of eliciting, processing, and enacting the feedback on the part of the students (Malecka et al., 2020).

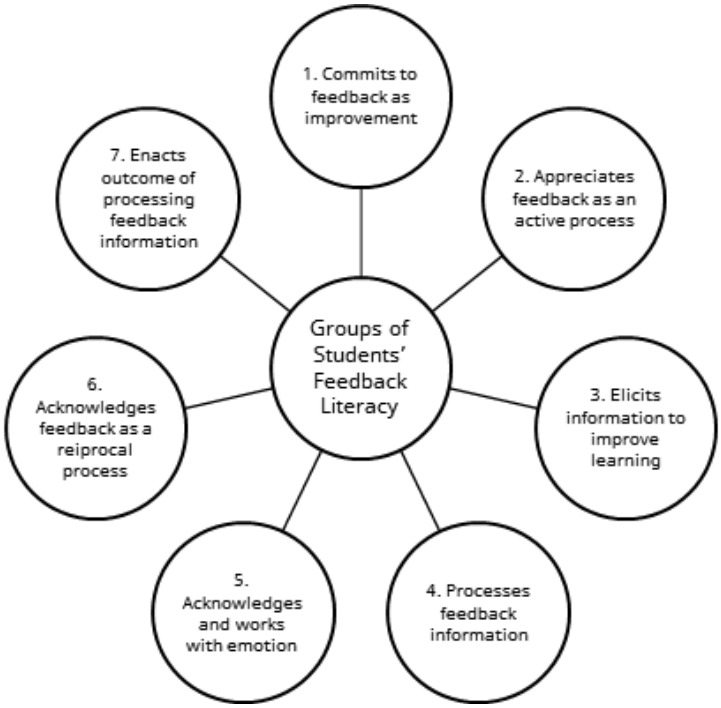


Figure 1. The seven groups of students' feedback literacy framework (Molloy et al., 2020)

In the original framework of Carless and Boud (2018), the role of technology in promoting students' feedback literacy has been clearly foregrounded. Within the component of "appreciating feedback," the use of technology to access, store, and revisit feedback is touted as one of the key indicators. Hence, the authors recommended the utilization of a Learning Management System (LMS) as a platform to store and annotate exemplars, as well as digital peer feedback through web-based applications and recorded video. Indeed, students have also voiced their preference for digital feedback mode as it is perceived to be more detailed, personalized, and useful (Ryan et al., 2019). Subsequent research works extending the feedback literacy framework discuss the role of technology in promoting learners' active role in the feedback process with encouraging outcomes (Ducasse & Hill, 2019; Wood, 2021, 2022).

Focusing on AI technology, two recent studies discussed the impact of AI tools on feedback literacy in the writing classroom. Tubino and Adachi (2022) made use of FeedbackFruits, an AWE tool and writing editor, in an academic writing course at an Australian university. The results revealed that, although only a small percentage of the students availed of the tool, those who used it commended it for its capability of rating and commenting on the usefulness of

the feedback. Next, Rad et al. (2023) investigated the deployment of Wordtune, an AI-based writing editor application, to enhance the feedback literacy, engagement, and writing outcome of 46 Iranian ESL (English as a Second Language) students. The quantitative data demonstrated the superiority of Wordtune in promoting the feedback literacy, engagement, and achievement of the students in the experimental group.

From what has been discussed in this section, it can be seen that research on the enactment of AI technology as an AWE tool in supporting feedback literacy of EFL students is still largely missing from the literature. This study then aims to narrow this gap.

Methodology

Design

This study is framed as a qualitative case study (Creswell & Poth, 2016), chosen for its focus on examining a particular case through in-depth and detailed data collection and analysis, gathering various forms of qualitative data. In this research, the case is bounded by people, place, and time as EFL undergraduate students taking an English education major at a private university in Surabaya, Indonesia, in the even semester of the 2022/2023 academic year, spanning a period of approximately four months (February–April 2023). Hence, the specificity of the time is also signified as the time that ChatGPT is starting to be known among academics and educators. In line with the primary aim of this study, the design can also be classified as an instrumental case study (Creswell & Poth, 2016) owing to the intent of understanding a particular phenomenon, namely the use of AI-based technology, particularly the newly invented ChatGPT, as instruments for self-assessment. Lastly, this study is also conceived as a single case study of the entire class while at the same time attempting to be acquainted closely with each participant for richer, nuanced data.

Context and participants

The study is set within the context of a writing course at a private university in Surabaya, the capital of the East Java province, Indonesia. The participants were 18 undergraduate students (referred to as P1 to P18 in this study) in their second semester of an English Education study program, taking a course entitled Recount and Narrative Essays as a compulsory, 3-credit bearing subject. The 18 students comprise seven males and eleven females, constituting the entire cohort of the 2022 intake of the study program. In a TOEFL-benchmarked test conducted in their first semester, their scores ranged widely from 390 to 580, indicating a proficiency level between A2 and B2 on CEFR. The course ran for a total of 16 weeks, made up of 14 weeks of classes and two weeks of mid-term and final examinations. Throughout the course, the students were required to produce seven recount and narrative texts under a process writing approach with multiple drafting processes.

The first three authors were directly involved in the research data collection because they are faculty members in the English Language Education Study Program where the study was conducted. In addition, the first and second authors were the co-instructors in the writing course, thus allowing us to gain an emic, insider perspective of the participants. During the course, the first author led the feedback activities, including utilizing AWE tools on students' works. Hence, the participants were chosen through convenient and purposive sampling techniques: Convenient sampling because they were students of our class. Besides, laying the groundwork for successful feedback practice is deemed crucial at the early stages of students' undergraduate lives, hence the purposive sampling of students in the second semester.

Instruments

To harness rich data, several instruments were used to gather qualitative evidence. Firstly, a semi-structured interview question was drafted in general terms: "Which AWE tool did you use, and how did it help you in your self-evaluation?" Depending on the participants' responses, follow-up questions were posed for elaboration or clarification. As part of their assignments, the students were also asked to write three reflective journals, one of which pertains to this study as it asked students to relate their experience when doing the self-evaluation using the AWE tools in the preceding week. During the interview, a stimulated recall was also performed for data in the reflective journals that require further confirmation. To enhance the trustworthiness of the data (Lincoln & Guba, 1985), triangulation was used by examining the students' artifacts in the form of e-portfolios (in Google Docs format) containing their essay outlines and drafts, as well as peer feedback evidence, to corroborate the self-reported results of the interview and reflective journals.

Data collection

At the onset, care was taken to gain verbal approval from the head of the study program before proceeding with the research. In the first meeting, the students were informed that research was being conducted concurrently with the course but should in no way affect the integrity of the subject nor influence their grades. Overt data collection, such as questionnaires and interviews, counted on the participants' voluntary participation and informed consent.

As part of the first author's dissertation project regarding explicit instruction and interventions on feedback literacy, the students were introduced to the notion of self-evaluation using technology, which includes Grammarly, Quillbot, and ChatGPT. A quick class survey revealed that most students were familiar with Grammarly and Quillbot, but most had not heard of ChatGPT at that moment (February 2023). The first author then demonstrated how ChatGPT works, particularly how it can be utilized to proofread, provide feedback, and paraphrase one's work. As we wanted to make a continuous reference to the rubric as part of the feedback literacy program, the ChatGPT prompt was

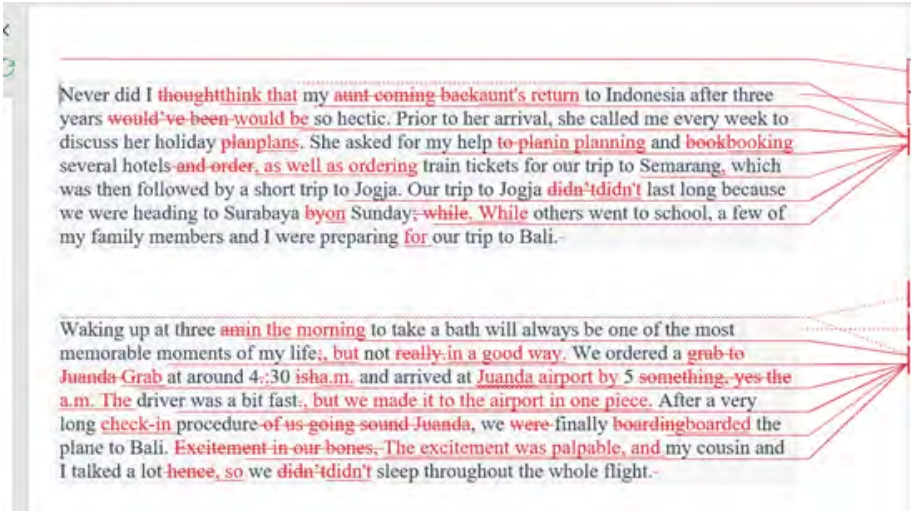
modeled as follows: “Revise my recount text and show the revision in terms of content, organization, grammar, vocabulary, and mechanics.” After that, a brief demonstration of Grammarly and Quillbot was also given. The students were then asked to make use of any AWE tool of their choice in the first draft of their recount essays, which were done online using Google Docs, thus constituting their e-portfolios. To provide an audit trail of their AWE tool usage, they were instructed to provide a screenshot of the self-evaluation process. For ChatGPT in particular, besides the screenshot, the students were encouraged to use the “compare” feature in the Microsoft Word document to highlight the changes made by ChatGPT if it is used as a paraphrasing tool. A sample of a student’s work asking ChatGPT to review her text is shown in Figure 2, and the comparison of her original writing and ChatGPT’s suggestion is shown in Figure 3. In the subsequent recount essay drafting, the students were encouraged to continue using their preferred AWE tools, although it was not compulsory for them to show the screenshots of their AWE tools’ usage.

Sure, here's my revision with some suggested edits:

Never did I think that my aunt's return to Indonesia after three years would be so hectic. Prior to her arrival, she called me every week to discuss her holiday plans. She asked for my help in planning and booking several hotels, as well as ordering train tickets for our trip to Semarang, which was then followed by a short trip to Jogja. Our trip to Jogja didn't last long because we were heading to Surabaya on Sunday. While others went to school, a few of my family members and I were preparing for our trip to Bali.

Waking up at three in the morning to take a bath will always be one of the most memorable moments of my life, but not in a good way. We ordered a Grab at around 4:30 a.m. and arrived at Juanda airport by 5 a.m. The driver was a bit fast, but we made it to the airport in one piece. After a very long check-in procedure, we finally boarded the plane to Bali. The excitement was palpable, and my cousin and I talked a lot, so we didn't sleep throughout the whole flight.

Figure 2. A screenshot of a student’s work: the ChatGPT output



Never did I ~~thought~~think that my ~~aunt coming back~~aunt’s return to Indonesia after three years ~~would've been~~ would be so hectic. Prior to her arrival, she called me every week to discuss her holiday ~~plan~~plans. She asked for my help ~~to plan~~in planning and ~~book~~booking several hotels ~~and order~~, as well as ordering train tickets for our trip to Semarang, which was then followed by a short trip to Jogja. Our trip to Jogja ~~didn't~~didn't last long because we were heading to Surabaya ~~by~~on Sunday~~, while~~. While others went to school, a few of my family members and I were preparing for our trip to Bali.

Waking up at three ~~a~~in the morning to take a bath will always be one of the most memorable moments of my life~~:~~, but not ~~really~~ in a good way. We ordered a ~~grab to Juanda~~ Grab at around 4~~.~~:30 ~~isha~~.m. and arrived at Juanda airport by 5 ~~something, yes the~~ a.m. The driver was a bit fast~~.~~, but we made it to the airport in one piece. After a very long check-in procedure ~~of us going sound Juanda~~, we ~~were~~ finally ~~boarding~~boarded the plane to Bali. ~~Excitement in our bones.~~ The excitement was palpable, and my cousin and I talked a lot ~~hence~~, so we ~~didn't~~didn't sleep throughout the whole flight.

Figure 3. A screenshot of a student’s work: the ChatGPT output compared with the original Text

In the last meeting before the mid-term examination, namely in Week 7, we distributed a Google Form to invite students for the interview and to indicate their contact details should they agree. Out of the 18 students, 12 expressed their willingness for the interview. All 12 students were subsequently interviewed individually in separate sessions out of class. The interview was conducted face-to-face and lasted between 15 and 30 minutes since this study forms part of a larger feedback project. A mix of English and the national language was used to facilitate the participants' ease of expression. All the interview sessions were recorded upon the participants' verbal consent for subsequent transcription, translation, and analysis. The transcription was carried out by a research assistant who is a recent graduate and is therefore familiar with the study context, and who has previous experience in transcription. Inaudible sounds were marked as such by the assistant and were checked by the authors. The translation into English and its proof-reading were done by the authors.

As for the artifacts, the students were assigned to write four recount texts and three reflective journals throughout the seven meetings. As mentioned earlier, the outlines and the drafts of the recount texts were written in their Google Docs e-portfolios, in which they showed evidence of self-assessment with AWE tools, received peer- and teacher feedback, and enacted the feedback uptake. Lastly, the students submitted their reflective journals through the university's Learning Management System.

Data analysis

The qualitative data sources (interview transcripts and reflective journals) were entered into NVivo12 for coding and analysis. In the program, both deductive and inductive analytical approaches (Azungah, 2018) were employed. For the former, a set of a priori codes based on the grouping in the framework of students' feedback literacy (Molloy et al., 2020), as shown in Figure 1, were created using the keywords for each element. The codes thus developed were "commit," "appreciate," "elicit," "process," "emotion," "reciprocal," and "enact." In the inductive approach, thematic analysis (Braun & Clarke, 2006) was adopted, in which open codes were created and grouped into themes, which were posteriorly reviewed and refined. The coding process thus followed the iterative procedure as postulated by Miles et al. (2014). Deductive coding was performed in the first cycle, and inductive coding in the second. The first author performed the coding, which was then reviewed by the third author until a satisfactory agreement was reached.

Findings

AWE tools in support of students' feedback literacy

We first illustrate the result of the data analysis through deductive coding based on the seven groups of Molloy et al.'s (2020) students' feedback literacy

framework (henceforth "Framework"). In presenting the result, we progress from the group with the least number of coded data to the most.

Firstly, there was hardly any empirical representation for Group 6 of the Framework, namely "Acknowledges feedback as a reciprocal process"; only one student was found to provide peer feedback suggesting that the friend use "Grammarly or whatever you are using right now" (P6, peer-feedback artifact) to improve upon lexical choice. While we did mention during class time that students can use AWE tools for peer review to help them provide peer feedback, it is possible that time constraints and inadequate motivation resulted in none of the students utilizing AWE tools to assist them in peer feedback provision.

Similarly, there was little evidence in Group 5 of the Framework, which is about acknowledging and working with emotion. Interestingly, the two quotes in this group reflect the opposing sentiments of two students about the use of AI in feedback elicitation. One student quipped that she prefers Grammarly to peer feedback because the tool is perceived to be more objective and will not comment on "the content, the theme, and the emotion they get from my work." (P1, interview). On the other hand, another student confessed that he found chatting with AI "weird" (P6, interview).

More students' expressions fit the category of Group 1, i.e., "Commits to feedback as improvement." In this category, students generally exhibited their appreciation in being introduced to AWE tools for obtaining feedback and the desire to learn more applications while at the same time expressing caution on being over-reliant on technology. Nevertheless, two students voiced their mistrust in the use of AWE tools and preferred to "use my own language." (P13, Journal 2). A representative quote from the more enthusiastic student is shown in the following excerpt:

> I think a lot of us want the teacher to keep continuing with the tools like how to use ChatGPT the other day. Maybe the teacher could teach us more about how to use another tool or maybe some recommended apps that could help us in improving our writing skills in this course. (P16, Journal 2)

Next, Group 7 of the Framework pertains to "Enacting outcome of processing of feedback information," which essentially describes the use of feedback for planning, goal-setting, and monitoring one's learning. Here, students indicated their plans to continue using ChatGPT in the future, mindful of its role in improving one's writing. Others stated that they have been using Grammarly for quite some time, indicative of a sort of monitoring of one's work quality. One student specified that she plans to use ChatGPT more for grammar-checking instead of paraphrasing, and another intends to keep using ChatGPT to learn how to incorporate advanced vocabulary.

We will now present the last three groups of the Framework, which contain a greater number of expressions, which at times reveal to be rather contrasting in nature. Firstly, Group 3, "Elicits information to improve learning," refers to, among other things, students' capability of actively soliciting feedback from multiple resources. Here, we found statements from students indicating their use of AWE tools in combination with peer- and/or teacher feedback and the

course materials, arising from their doubts or perceived insufficiency in the AWE tools' responses. In addition, students also utilized a combination of AWE and other technological tools, for example, Quillbot or Grammarly for grammar-checking and ChatGPT for paraphrasing, or ChatGPT followed by Google to look up the meaning of words. The active disposition on the students' part was also demonstrated by the fact that a student used the AWE tool on her handwritten assignment on her own accord, and another had earlier used Grammarly on a friend's recommendation. However, not all students viewed AWE tools with enthusiasm; a student wrote the following:

> Personally, I prefer asking a friend who is an expert and comparing them and to see which one is correct because I think that using the AWE tools can be useful, but not all the time since some of them often make mistakes. I also think that it is best to rely on your gut more than AI because if you're confident you basically don't need AWE tools, maybe just to double-check that there are no errors, and that is all. (P18, Journal 2).

Secondly, there were numerous statements relating to Group 2 of the Framework: "Appreciates feedback as an active process." In this research, we interpret this as students' ability to identify the manner in which the AWE tools respond to their specific learning needs. When it comes to ChatGPT, students acknowledge the role it plays in exemplifying varied sentence structures, indicating proper punctuations, paraphrasing texts to suit a particular writing genre, aiding in lexical choice and avoiding repetition, and suggesting unique English idioms. Students also turned to ChatGPT for its generative capability in embellishing content with details, obtaining translation, and brainstorming ideas. Those who used Grammarly professed its profitability in improving one's grammar. Lastly, students also made use of the "spelling and grammar check" function in Microsoft Word and Google Docs for instantaneous corrections.

Finally, the group that garnered the greatest number and variety of evidence was group 4, "Processes feedback information," which essentially depicts students' ability to accept or reject feedback, as well as select key actionable information from the feedback. Here, we segregated the data into "positive," namely the expressions of the manner in which the AWE tools facilitated students' feedback processing, and "negative" for the contrary evidence. Within the negative representations, we were able to distinguish patterns among utterances made by proficient students and the less proficient ones.

In general, the positive comments were made by students we observed in class to be proficient or academically motivated. When using ChatGPT to paraphrase their texts, they exerted effort to identify the changes made manually or by using the "Compare" function of Microsoft Word. A student made sure he reviewed ChatGPT's product because "*it is a human-made program, so it may not always accurately interpret the meaning of our words.*" (P14, Journal 2). Another student realized that using ChatGPT involves the skill of prompt engineering and is still learning how to maximize its output using the most suitable prompt. One even made it a point to regenerate ChatGPT's responses several times until he obtained the best result. This is what he said:

> I didn't just try it once and then take it, no. I gave the prompt, and I regenerated the answer until about five to seven times. And I kept on editing the prompt until it was okay, and I regenerated it several times, and then I compared which response is more suitable for me. Sometimes, I did a mix and match; for example, for the first paragraph, I took the second answer; for the next paragraph, I took the 5th answer, etc. After that, I analyzed it again to see if it flows. I also checked the sentence, whether it has changed from past form to present form. (P11, interview).

As for the negative responses, denoting students' perception of how the AWE tools hinder their feedback processing, the more proficient but less motivated students lamented the fact that ChatGPT, in comparison with Grammarly or Quillbot, does not manifest the correction points instantly. A proficient student could detect that Grammarly sometimes wrongly suggested a correction on non-English words. Students disliked it when ChatGPT appeared to shorten their paraphrased texts or recreate their writing with different meanings. One student discerned that her friends used ChatGPT less judiciously. She remarked:

> Yeah, I think people, just when they think that ChatGPT can revise the whole thing, they were like, okay, I'm just going to create, that's it. They don't want to read it through. And then when others read it, they were like, what is this person talking about? So yeah, they need to not rely too much on ChatGPT. I don't know. It's too… too instant. (P1, interview).

On the other hand, the less proficient or less motivated students found the AWE tools unhelpful, in certain ways, for providing feedback. For example, a student found the AWE tools' corrections confusing, and another felt it made his writing even worse. Some students confessed that they felt they were cheating when they used ChatGPT, and another said she didn't like ChatGPT's output because she found the vocabulary too advanced. A student used a Grammarly add-on on her gadget's keyboard, and she was confused by the text-predictive feature of the application.

Considerations of AWE tools' usage in EFL writing class

We now present the result of the inductive coding, through which we identified several themes that we concluded to be the six aspects that educators need to be cognizant of when using AWE tools in support of students' feedback literacy in EFL writing classrooms, specifically in the era of generative AI. We propose to name it "AWE Tools Integration Framework," as shown in Figure 4.

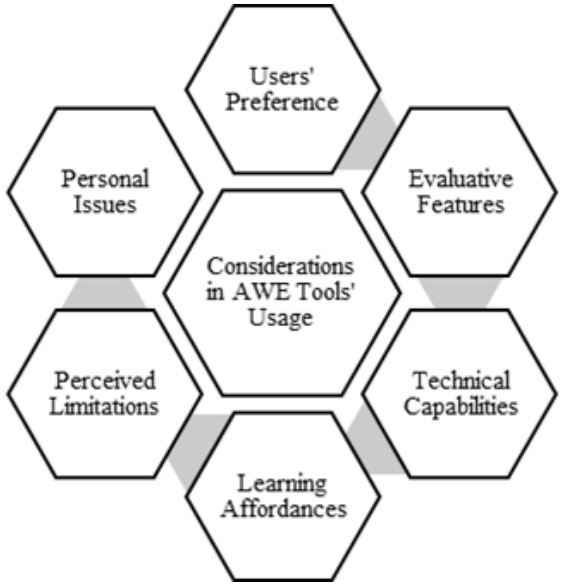


Figure 4. The AWE tools integration framework

Users' preference. From our data, several students clearly stated their preference for one AWE tool over another for various reasons. In terms of choice for ChatGPT, students cited its multi-purpose capabilities of grammar-checking, paraphrasing, lexical boosting, and instant revision as the reasons. Others favored Grammarly because of its ease of use and the annotation of the correction points. However, some prefer Grammarly over ChatGPT because they thought it was a paid application or because they claimed they did not know how to use it. Interestingly, one student stated that he thinks using Grammarly is good enough for him without having to recourse to other tools, and another chose Grammarly because using ChatGPT is perceived as cheating. Similarly, students expressed their preference for Quillbot over ChatGPT for almost identical reasons as Grammarly above, although one student clearly asserted that she opted for Quillbot over both ChatGPT and Grammarly. Hence, it is imperative that teachers provide options to students in the use of AWE tools to cater to differing preferences.

Evaluative features. For AWE tools in general, a student commented that using AWE tools is preferable for her because of the technical and objective nature of the suggested correction, as compared to peer feedback, which she regards as subjective. ChatGPT, as pointed out previously, has been singled out by several students for its generative capabilities, which facilitate grammatical, lexical, mechanical, and style improvement, as well as the inspiration for ideas, content, and genre. On the other hand, Grammarly and Quillbot seem to complement ChatGPT through their focus on grammatical editing, which can be detected instantly and shows the specificity of the syntactical issues. In addition, the grammatical corrective capability of Grammarly has apparently been established among students such that it has become the tool that they immediately resort to.

Technical capability. Some students were able to independently explore the technical features of ChatGPT and Grammarly to their advantage. For ChatGPT, students realized the need for prompt engineering to obtain the optimal output. As previously mentioned, a student took the pain to regenerate the ChatGPT's output several times and even pick and choose the results. In the case of Grammarly, students commended it for its ease of access and convenient use. The more seasoned users have installed Grammarly Keyboard on their mobile phones, and a student integrated Grammarly into her word processor.

Learning affordances. Our data unveiled differing types of learning opportunities afforded by Grammarly and ChatGPT. A few students commended Grammarly for helping them with verb tenses, a seemingly typical, persistent problem among EFL students. ChatGPT, on the other hand, stimulated students' curiosity to explore various writing genres or to look up the meaning of words. Even if ChatGPT's paraphrased output does not indicate the changes from the original, students were motivated to compare and contrast the differences or patiently work paragraph by paragraph to detect the changes in stages.

Perceived limitations. Students also gave voice to the perceived limitations of the AWE tools, although this perception might have arisen due to the incomplete knowledge they have of the applications or, in the case of Grammarly and Quillbot, not experiencing the full features of the free version of the software. Therefore, a student claimed that Grammarly did not provide sufficient corrections or "cannot get the writing context" (P3, interview). Besides, Grammarly could not recognize non-English words and thus flag them as errors. In the case of ChatGPT, students lamented that it does not provide an instant error recognition feature, is perceived as over-paraphrasing and shortening the text, provides inaccurate data (something that ChatGPT itself cautions the users), and makes incorrect interpretations of the writing context.

Personal issues. Last but not least, excerpts identified as students' personal issues, namely subjective, negative experiences of the students, emerged from our data. For AWE tools in general, a student found the corrections confusing, and another mistrusted the tools altogether. Regarding ChatGPT, students admitted that they submitted ChatGPT output without reviewing it and felt uneasy about it. Another was worried that she had become too dependent on technology, and another felt strange chatting with a robot. A Grammarly user felt overwhelmed by the word-by-word corrections, and one disliked the text-predictive function of the Grammarly Keyboard. Lastly, a Quillbot user did not fully comprehend the suggested corrections and decided to ignore them altogether.

We provide below an extract from a participant's journal, edited slightly for clarity, about her view on ChatGPT:

> Luckily, my lecturers introduced me to ChatGPT, which made it easier [for me] to do my assignment. At first, I did not want to use it because I thought

it was kind of cheating. But once you get used to it, you will open your mind that technology makes your life easier. So, I used ChatGPT. Despite that, ChatGPT also could make me lazy. So I had to convince myself that I needed to learn something from ChatGPT, not just take the answer blindly without thinking where my mistakes were. So, until now, I am still learning how to be a good writer. I hope that in the next assignment, I can use less of ChatGPT as my feedback; it means that I hope I will get less corrections from ChatGPT and improve my writing skills. (P5, Journal 2).

Discussion

The results of the deductive coding have shed insightful light on the manner in which the different AWE tools supported the development of students' feedback literacy. Overall, our results complemented the quantitative work of Rad et al. (2023), who found a statistically significant increase in the feedback literacy of a class using an AI-powered application as their AWE tool since our qualitative data are able to provide a further description of how it comes about. Firstly, under "Commits to feedback as improvement," ChatGPT as an AWE tool seems to constitute a discovery for most students and awakens a sense of adventure in both present and future use. This might be due to the revolutionary potential of ChatGPT in providing dialogue-like feedback in multiple contextualized conversations that sets it apart from other rule-based AWE tools (Parker et al., 2023). Secondly, AWE tools aided students in identifying and addressing gaps in their English writing skills, which falls under the feedback literacy skill of "Appreciates feedback as an active process." In this case, ChatGPT and Grammarly/Quillbot seem to play a complementary role, with the former for enhanced paraphrasing and the latter utilized as explicit grammar corrections. While the utility of Grammarly and Quillbot in grammatical improvement has been relatively established (Ebadi et al., 2023; Kurniati & Fithriani, 2022), students can consider generative AI for enriching lexical repertoire and enhancing writing style. Similarly, the AWE tools also supported the feedback literacy skill of "Elicits information to improve learning." Some of the participants in our study were aware of the need to use a combination of AWE tools or have recourse to friends and teachers for a holistic feedback experience. For ChatGPT, in particular, our participants felt the need to consult multiple sources due to the inherent limitations of the generative chatbot (O'Neill & Russell, 2019; Rudolph et al., 2023). We concur with Hockly (2019) in her call for a more integrative view of assessment encompassing all aspects of teacher, peer, and technological feedback.

The most controversial aspect of the impact of AWE tools on students' feedback literacy seems to be the "Processes feedback information," with students being divided into promoters and detractors. A group of students was able to critically engage with ChatGPT (Carless, 2023), mastering the art of prompt engineering and response regeneration in the process. In this sense, ChatGPT might have played a role in developing the students' evaluative judgment (Tai et al., 2018) by exemplifying standards of good work. On the other hand,

other students still preferred the tacit grammatical expositions of Grammarly/Quillbot and found ChatGPT's instantaneous response unprofitable for learning. In addition, a few students were not able to benefit from the AWE tools since they could not perceive the improvement made. Overall, this highlights the need for instructors to scaffold students in processing automated feedback (Zhang, 2020) and leveraging ChatGPT's prompt creation for optimal results (Wu et al., 2023). Next, in terms of "Acknowledges and works with emotion," AWE tools appear to support students' feedback literacy through its capability of providing supposedly objective and purely technical feedback, thus benefiting students who were disinclined to giving or receiving human feedback (Ryan et al., 2019).

As for "Acknowledges feedback as a reciprocal process," we did not find much evidence in our data due perhaps to the time constraint for enacting AWE-mediated peer feedback in this writing course. Nonetheless, the quantitative data indicated a positive disposition in the students for such exercise, and this area may be a fertile ground for further research. Lastly, AWE tools supported students' feedback literacy in the "Enacts outcomes of processing feedback information" component through their technical capabilities, enabling the learners' sustainable, long-term use. Our students' statement on their having used Grammarly since high school is perhaps indicative of Grammarly's continuous improvement in its algorithm and user interface. When used in tandem with tutors' guidance, feedback from Grammarly was shown to be more comprehensible and actionable (O'Neill & Russell, 2019).

The outcomes of the inductive coding process unveil six themes, which were aggregated into the AWE Tools Integration Framework. The aspects related to users' preferences, evaluative features, technical capability, and learning affordances emphasize the need for educators to drive learner's agency in selecting the AWE application of their choice. In fact, teachers can leverage the capability of the different tools and encourage their complementary use in conjunction with peer and teacher feedback (Huawei & Aryadoust, 2023). The findings in this area also imply that it is perhaps too early to claim that ChatGPT could supplant Grammarly or Quillbot in their AWE roles. The long-standing presence of Grammarly, for example, as well as its real-time, comprehensive, and overt language editing features (Barrot, 2020), might still prove helpful for EFL students, particularly low-proficiency students (Tambunan et al., 2022). Besides, the frenetic pace of research and deployment of a greater number of AI-based applications necessitate continuous exploration by both teachers and students to discover other AWE tools to support learning and assessment.

Two aspects of the AWE Tools Integration Framework – perceived limitations and personal issues – provide insights into the need to enhance digital literacy among EFL students (Godwin-Jones, 2022), as well as bridge the inequality in technological access. The perceived limitations of the AWE tools may have been due to the possibility that some students have not used the complete features of the paid versions and that, despite their age grouping, not all of them are digital natives who tend to be skilled users of technological tools (Bašić et al., 2023) and familiar navigators of the Internet. As a matter of fact, there is

still a disparity among students in accessing technology in terms of their gadget sophistication and internet access beyond campus hours. Therefore, to optimize the benefits of AWE tools, teachers ought to first be conversant with the AWE tools themselves, utilizing them for their research works, for example, so as to be able to support students who face technical challenges in utilizing the tools. In this respect, teachers' self-efficacy in technological integration comes into play (Lailiyah & Cahyono, 2017; Muslimin et al., 2023). Educational institutions may need to evaluate the means so that, through budget allocation or extraneous financial support, students and faculty can gain access to the premium version of technological applications if they so wish.

Furthermore, education institutions are known to be oftentimes lagging behind technological advances. Yet, it's never too late to address the ethical issues associated with the use of AI in education, such as data privacy, plagiarism issues, biases, and discrimination (Alexander et al., 2023). Schools and colleges need to agree on and issue clear guidelines so students and teachers can use AI productively to enhance student learning and reach the common good (Su et al., 2023). The discomfort felt by the students in this study when using ChatGPT for writing has been shared by others (Fyfe, 2022), while AI text detectors could give false positives (Bašić et al., 2023). Thus, efforts should be spent on guiding students to use AI profitably rather than apprehending them for academic misconduct.

Particularly, the guidance should be on how to critically use and evaluate AI and, in a broader spectrum, technology. As AI's development is an uncharted territory where engineers are tweaking on a daily basis, AI can become very different in just a short time. For example, a study by Chen et al. (2023) found a wide variation of quality and accuracy in ChatGPT's answers in just three months. If the students are not supported by the educators in using AI, most of them would take AI's answers for granted, as shown in this research by the lowest score on students' disposition to look up other resources. Thus, there is an urgent need to 1) introduce AI and how to use it properly and, especially, 2) develop critical thinking and critically evaluate students' use of AI.

The results of this study add to the rapidly growing studies on the use of generative AI, ChatGPT in particular, in the English Language Teaching milieu. On the theoretical front, our findings may have provided more support for the body of students' feedback literacy research and its application in self-assessment using technology. In addition, our seminal AWE Tool Integration Blueprint may prove to be a modest guideline for teacher-practitioners in the use of AI-powered AWE tools in formative assessment, as well as pave the way for further research to validate the framework.

We acknowledged that our roles as insider participants, being researchers and instructors in the classroom simultaneously, might have resulted in potential bias in the students' behavior and responses. The coding process could have been made more rigorous through multiple coders with measured inter-rater agreement. Notwithstanding these limitations, we hope that the findings of this study have advanced the understanding of the use of AWE tools in support of students' feedback literacy in the era of generative AI.

Conclusion

This study set out to explore the utilization of various Automated Writing Evaluation (AWE) tools, including ChatGPT, in an EFL writing course at a higher education level while at the same time examining the manner in which the tools supported students' feedback literacy. Eighteen second-semester students participated in the study, with rich qualitative data gathered from a semi-structured interview with stimulated recall, reflective journals, and students' artifacts. Data analysis comprising both inductive and deductive coding revealed that ChatGPT, Grammarly, and Quillbot, alone or in combination, supported the development of almost all aspects of students' feedback literacy. In particular, the AWE tools were perceived as either augmenting or obstructing the "processing of feedback information" aspect, with the proficient and motivated students endorsing ChatGPT for enhancing their writing skills. The inductive coding resulted in the proposal for an "AWE Tool Integration Framework," which is hoped to guide instructors in the utilization of the various AWE tools in the paradigm of generative AI for their EFL writing courses. Nonetheless, several issues of digital literacy, technological access inequality, and ethical use of generative AI remain unresolved, thus becoming an exhortation for action on the part of educators and educational institutions.

Contributions

Imelda Gozali conceptualized the idea, performed the data collection with Alberik Ryan Tendy Wijaya, conducted the data analysis, and wrote the manuscript with Alberik Ryan Tendy Wijaya and Anita Lie. Bambang Yudi Cahyono and Nunung Suryati supervised the work and reviewed the manuscript.

References

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology, 23*(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Azungah, T. (2018). Qualitative research: deductive and inductive approaches to data analysis. *Qualitative Research Journal, 18*(4), 383–400. https://doi.org/10.1108/qrj-d-18-00035

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bašić, Z., Banovac, A., Kruzic, I., & Jerkovic, I. (2023). Better by you, better than me, ChatGPT3 as writing assistance in students' essays. *arXiv preprint arXiv:2302.04536*. https://doi.org/10.48550/arXiv.2302.04536

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Carless, D. (2020). Longitudinal perspectives on students' experiences of feedback: A need for teacher–student partnerships. *Higher Education Research & Development, 39*(3), 425–438. https://doi.org/10.1080/07294360.2019.1684455

Carless, D. (2023, August 22). *Assessment re-designs for the generative-AI era* [Video]. https://polyu.hk/dUBbv

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT's behavior changing over time? arXiv, 1(1), 1–23. https://doi.org/10.48550/arXiv.2307.09009

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches.* Sage publications.

Ducasse, A. M., & Hill, K. (2019). Developing student feedback literacy using educational technology and the reflective feedback conversation. *Practitioner Research in Higher Education, 12*(1), 24–37. https://ojs.cumbria.ac.uk/index.php/prhe/article/view/513

Ebadi, S., Gholami, M., & Vakili, S. (2023). Investigating the effects of using Grammarly in EFL Writing: The case of articles. *Computers in the Schools, 40*(1), 85–105. https://doi.org/10.1080/07380569.2022.2150067

Fang, T., Yang, S., Lan, K., Wong, D. F., Hu, J., Chao, L. S., & Zhang, Y. (2023). Is ChatGPT a highly fluent grammatical error correction system? A comprehensive evaluation. *arXiv preprint arXiv:2304.01746.* https://doi.org/10.48550/arXiv.2304.01746

Fyfe, P. (2022). How to cheat on your final paper: Assigning AI for student writing. AI & SOCIETY, 1–11. https://doi.org/10.1007/s00146-022-01397-z

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology, 26*(2), 5–24. http://doi.org/10125/73474

Gozali, I., Syahid, A., & Suryati, N. (2023). Ten years after Sutton (2012): Quo vadis feedback literacy? (A Bibliometric study). *Register Journal, 16*(1), 139–167. https://doi.org/10.18326/register.v16i1.139-167

Ho, C. C. (2022). The QuillBot grammar checker: friend or foe of ESL student writers? *Journal of Creative Practices in Language Learning and Teaching (CPLT), 10*(1), 10–31. https://ir.uitm.edu.my/id/eprint/66534

Hockly, N., (2019). Automated writing evaluation, *ELT Journal, 73*(1), 82–88. https://doi.org/10.1093/elt/ccy044

Huawei, S., & Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies, 28*(1), 771–795. https://doi.org/10.1007/s10639-022-11200-7

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal,* 00336882231162868. https://doi.org/10.1177/00336882231162868

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

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *The Electronic Journal for English as a Second Language, 27*(3), 1-19. https://www.tesl-ej.org/pdf/ej107/int.pdf

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics, 7*(3), 437–451. https://doi.org/10.21462/jeltl.v7i3.852

Lailiyah, M., & Cahyono, B. Y. (2017). Indonesian EFL teachers' self-efficacy towards technology integration (SETI) and their use of technology in EFL teaching. *Studies in English Language Teaching, 5*(2), 344–357. https://doi.org/10.22158/selt.v5n2p344

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage.

Malecka, B., Boud, D., Tai, J., & Ajjawi, R. (2022). Navigating feedback practices across learning contexts: implications for feedback literacy. *Assessment & Evaluation in Higher Education, 47*(8), 1330–1344. https://doi.org/10.1080/02602938.2022.2041544

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). Qualitative data analysis: A methods sourcebook. Sage.

Miranty, D., & Widiati, U. (2021). An automated writing evaluation (AWE) in higher education. *Pegem Journal of Education and Instruction, 11*(4), 126–137. https://doi.org/10.47750/pegegog.11.04.12

Miranty, D., Widiati, U., Cahyono, B. Y., & Sharif, T. I. S. T. (2021). The effectiveness of using Grammarly in teaching writing among Indonesian undergraduate EFL students. In *International seminar on language, education, and culture (ISoLEC 2021)* (pp. 41–45). Atlantis Press. https://doi.org/10.2991/assehr.k.211212.008

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education, 45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research, 8*(1), 149–175. https://doi.org/10.17239/jowr-2016.08.01.05

Muslimin, A. I., Mukminatien, N., & Ivone, F. M. (2023). TPACK-SAMR digital literacy competence, technostress, and teaching performance: Correlational study among EFL lecturers. *Contemporary Educational Technology, 15*(2), ep409. https://doi.org/10.30935/cedtech/12921

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education, 85*(6), 1381–1400. https://doi.org/10.1007/s10734-022-00895-9

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1). https://doi.org/10.14742/ajet.3795

Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410606860-12

Parker, J. L., Becker, K., & Catherine Carroca, D. H. S. (2023). ChatGPT for Automated Writing Evaluation in scholarly writing instruction. [preprint] https://www.researchgate.net/profile/Kimberly-Becker/publication/372158169_ChatGPT_for_Automated_Writing_Evaluation_in_Scholarly_Writing_Instruction/links/64c6bb42539d123b18d6c72f/ChatGPT-for-Automated-Writing-Evaluation-in-Scholarly-Writing-Instruction.pdf

Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: a case of Wordtune application. *Interactive Learning Environments,* 1–21. https://doi.org/10.1080/10494820.2023.2208170

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education*? Journal of Applied Learning and Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Ryan, T., Henderson, M., & Phillips, M. (2019). Feedback modes matter: Comparing student perceptions of digital and non-digital feedback modes in higher education. *British Journal of Educational Technology, 50*(3), 1507–1523. https://doi.org/10.1111/bjet.12749

Schindler, L. A., Burkholder, G. J., Morad, O. A., & Marsh, C. (2017). Computer-based technology and student engagement: a critical review of the literature. *International Journal of Educational Technology in Higher Education, 14*(1), 1-28. https://doi.org/10.1186/s41239-017-0063-0

Schmidt-Fajlik, R. (2023). ChatGPT as a grammar checker for Japanese English language learners: A comparison with Grammarly and ProWritingAid. *AsiaCALL Online Journal, 14*(1), 105–119. https://doi.org/10.54855/acoj.231417

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57,* 100752. https://doi.org/10.1016/j.asw.2023.100752

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

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgment: enabling students to make decisions about the quality of work. *Higher Education, 76,* 467–481. https://doi.org/10.1007/s10734-017-0220-3

Tambunan, A. R. S., Andayani, W., Sari, W. S., & Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback. *Indonesian Journal of Applied Linguistics, 12*(1), 16–27. https://doi.org/10.17509/ijal.v12i1.46428

Tubino, L., & Adachi, C. (2022). Developing feedback literacy capabilities through an AI automated feedback tool. *ASCILITE Publications,* e22039–e22039. https://doi.org/10.14742/apubs.2022.39

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Wood, J. (2021). A dialogic technology-mediated model of feedback uptake and literacy. *Assessment & Evaluation in Higher Education, 46*(8), 1173–1190. https://doi.org/10.1080/02602938.2020.1852174

Wood, J. (2022). Making peer feedback work: the contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education, 47*(3), 327–346. https://doi.org/10.1080/02602938.2021.1914544

Wu, H., Wang, W., Wan, Y., Jiao, W., & Lyu, M. (2023). ChatGPT or Grammarly? evaluating ChatGPT on grammatical error correction benchmark. *arXiv preprint arXiv:2303.13648.* https://doi.org/10.48550/arXiv.2303.13648

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