

## **Automated Writing Evaluation Feedback in English Learning and Teaching**

Shenli Zhang, University of Alberta

Writing is a fundamental skill in an individual's life, serving as a medium to express opinions, record important information, and create imaginative worlds. For students, writing is more significant because it is a necessity to achieve academic success no matter what subjects they are studying (Graham, 2018). However, writing is a complex task because it requires students to master both basic cognitive processes such as handwriting or spelling, and complex cognitive processes such as idea generation, transformation of ideas, and writing modification (Graham & Perin, 2007). Moreover, they also need sufficient practice to gain proficiency (Burstein et al., 2020). These abilities do not develop naturally; thus, schools play a vital role in nurturing students with these abilities by offering adequate practice and instruction in writing (Graham, 2018, 2019). An important instructional practice in the classroom is providing writing feedback. By showing the gap between students' current capabilities and expected outcomes, feedback can help students improve their writing quality (Biber et al., 2011; Graham et al., 2015).

However, providing real-time, detailed, and cross-subject feedback at scale poses a challenging task for educators (Burstein et al., 2020). In this case, Automated Writing Evaluation (AWE), defined as the capability of a computer technology that employs artificial intelligence, to evaluate and score written text (Shermis & Hammer, 2013), can serve as a viable choice, given its ability to offer immediate feedback to students (Lee, 2020).

This paper illustrates the affordances and constraints of AWE by reviewing findings from research studies, and discusses its application in instructional settings. Furthermore, it presents a comparison between instructors' feedback and AWE feedback, before providing practical implications for teaching English for Academic Purposes (EAP).

### **Advantages of AWE**

Based on large database and statistical models, the AWE system can provide an overall score along with suggestions for writing improvement (Hockly, 2019; Wang et al., 2020). Various AWE platforms are used in various contexts. For example, tools like *MY Access!*, *Pigai* and *Criterion* are more frequently used for academic purposes, while others like *Grammarly* are popular for diverse evaluating purposes, including blog writing and email composition.

AWE systems bring numerous benefits to both teachers and learners regarding promptness, convenience, opportunities to redraft and learner autonomy (Godwin-Jones, 2022). For teachers, it can significantly save time and energy, especially when dealing with large-scale assessments (Weigle, 2013). Therefore, teachers can allocate more time and resources to other crucial aspects of teaching, such as preparing for classes, developing more interactive and engaging learning materials, and providing personalized attention to students who need extra help.

For students, AWE provides very immediate feedback at one time and tracks students' progress over time (Burstein et al., 2020). Since AWE can provide instant feedback, students can get access to the feedback after they finish their writing, without waiting for several days or weeks (Cotos, 2014). For novice writers, producing a high-quality piece on the first attempt is challenging. The automated feedback system can offer corrective feedback for each successive

draft, enabling editing and improving, leading those beginners to a final version. This process is essential in training their metacognitive knowledge about writing and language learning (Godwin-Jones, 2022). The immediate feedback and chance for multiple revisions at the learner's own pace, empowered by the AWE system, enhances the learner's awareness of their progress and supports their individual learning needs, thereby improving student self-regulation in learning (Burstein et al., 2020; Lu, 2019).

This autonomous activity also simplifies classroom management for teachers, as learners at different levels can make effective use of their time during writing sessions: faster writers can revise and resubmit their writing for higher scores, while slower writers can continue focusing on their first draft (Warschauer & Grimes, 2008). Moreover, since the computer-assisted system does not require face-to-face interactions with the feedback provider, students, especially those learning a foreign language, feel less anxious when receiving feedback and thus can focus on error correction and writing improvement (Weigle, 2013; Waer, 2021).

Furthermore, students are motivated to revise their work multiple times. This refining process not only gives them a sense of achievement but also deepens their engagement with the text, thereby significantly benefiting their writing skills (Grimes & Warschauer, 2010).

### Concerns of AWE

While AWE offers many benefits, it also has notable limitations. One primary concern is the effectiveness of AWE at different learning stages. Research indicates that AWE feedback is most beneficial in the early stages of language learning (Huang & Wilson, 2021). This is because effective revisions often involve grammar accuracy and vocabulary appropriateness (Burstein et al., 2016). The AWE system tends to overemphasize on language-related errors (Huang & Renandya, 2020) and provides less assistance in higher-level features of writing such as “argumentation strength, discourse coherence, or organization” (Godwin-Jones, 2022, p. 7). The AWE-driven system is based on statistical analysis, large database and machine learning to “determine the likelihood of text sequencing” (Godwin-Jones, 2022, p. 7). In other words, AI systems do not truly understand human language (Grimes & Warschauer, 2010). As a result, students may manage to deceive the system by using strategies such as enhancing syntactical complexity or even repeating the same paragraphs in order to achieve higher scores (Powers et al., 2002).

A study examining the use of the AWE program *Criterion* in an Egyptian English as a foreign language (EFL) academic writing class found that while students improved their revisions and subsequent drafts, this improvement was often due to avoidance strategies, which means students avoided potential language which would cause errors in usage as measured by *Criterion* (El Ebyary & Windeatt, 2010). Similarly, Warschauer and Grimes (2008) observed that students were more inclined to make changes in simple linguistic features like spelling, grammar, and vocabulary, rather than focusing on text organization or content logic, which required critical thinking. This tendency contradicts the primary goal of writing education, which is to foster students' ability of appropriate and strategic using of language with communicative potential, rather than teaching tricks that help them score higher (Hyland, 2003).

Another issue of using AWE to provide feedback is that they “privilege certain languages and cultural backgrounds” (Hockly, 2019, p. 84). For example, *Criterion* is based on standard American English, which may lead to ineffective recognition of rhetorical and stylistic

expressions from other linguistic traditions. This preference of linguistic homogeneity may result in biased evaluations for students from diverse regions (Herrington & Stanley, 2012).

### **Comparison of Feedback of AWE and Teachers**

Numerous studies have evidenced the efficacy of AWE in improving learners' writing quality. However, it is important to clarify whether this improvement is due to the specific nature of AWE feedback or simply the result of receiving any feedback. To address this, researchers have compared AWE feedback with other types.

In research investigating students' engagement with teacher and automated feedback researchers found that teacher feedback addressed a broader range of error categories (e.g., possessive, pronoun and capitalization) than automated evaluation feedback (Zhang & Hyland, 2018). Additionally, the AWE system sometimes failed to diagnose some error categories such as article and preposition use. This is consistent with the findings in research by Liu and Kunnan (2016), who observed that an AWE system, *WriteToLearn*, had difficulty in identifying mistakes in articles, prepositions, word choice, and expression made by Chinese undergraduate students.

In the research of Tian and Zhou (2020), automated feedback provided more surface-level feedback while teachers offered in-depth, paragraph-level, and meaning-related feedback. These comments in terms of writing content and organization are highly valued by learners since they know that this feedback comes from human beings instead of algorithms (Zhang & Hyland, 2018). Besides, teachers are more capable of offering comprehensive feedback addressing issues such as number and abbreviation using, which may not be covered by AWE adequately (Zhang & Hyland, 2018).

While traditional teacher feedback has its own strengths, AWE possesses distinct advantages. In addition to being time-efficient and energy-saving, AWE systems, unlike teachers who often correct errors directly, tend to highlight mistakes, aiding students in becoming more aware of their errors (Zhang & Hyland, 2018). Furthermore, the immediacy of AWE feedback and the opportunity for learners to redraft and resubmit their work contributes to higher engagement compared to traditional teacher feedback (Zhang & Hyland, 2018).

In sum, both AWE and teacher feedback bring benefits to students and it appears that whether provided by computers or instructors, it is the power of feedback that really matters (Graham et al., 2015). Therefore, combining the two forms of feedback could potentially yield desiring results.

### **Pedagogical Implications for Writing Practice**

As Attali (2013) stated, the AWE system exhibits greater value when employed as a supplement rather than a substitute for teachers because this approach leverages the complementary strengths of both automated evaluation systems and human raters (Cahill & Evanini, 2020).

Efficiently applying this technology in classroom settings and integrating it with teacher-led instruction involves several factors. Those include the role of the teacher in supporting AWE use, the time allocated for writing practice, and students' perceptions of the technology (Venkatesh, 2020). According to Venkatesh (2020), the way users perceived technological tools influenced their use and effectiveness of such tools. If students regard them as helpful, they are more likely to engage with them in their writing practice. Teachers, as instructors for the knowledge and assistants in learning, play an important role in encouraging students to accept

the use of technology as well as fostering the efficient use of technological tools in classroom context (Davidson et al., 2014). By integrating AWE tools into classroom activities and demonstrating their use, teachers can influence students' cognitive, behavioral, and emotional engagement with system (Fu et al., 2022). For example, teachers could offer guided demonstration or training activities, familiarizing students with an appropriate AWE tool, along with its features, how to interact with it, as well as how to interpret its feedback. Teachers also need to monitor students' learning progress with the AWE system and modify the curriculum as needed (Fu et al., 2022). For instance, encouraging students to keep writing logs about how they applied the feedback and what challenges they faced with AWE, could provide a record of student progress over time for teachers to assess their growth and areas needing improvement.

According to Charles and Pecorari (2015), EAP writing class should equip students with the fundamental linguistic abilities of writing error-free sentences, using punctuation appropriately, thinking critically to make comparisons, and discussing and analyzing ideas. Link et al. (2020) suggested an ideal hybrid situation where the AWE system offers linguistic level feedback and teachers solve higher level stylistic and content-related writing issues. Since the AWE system can process texts quickly and identify basic linguistic errors, it could be used for initial draft revision, but these systems are not very qualified to critique higher-level elements such as coherence, organization, and argument strength. Teachers should fill this gap by focusing on context-specific feedback. This approach not only maximizes the effective use of teachers' time and inspires greater student engagement, but also closely mirrors the professional writing processes in academia, where authors use software for basic grammar and spelling checks before seeking more nuanced feedback from editors or peers (Morss & Murray, 2001). Thus, this blended approach prepares students for their future professional endeavors.

### **Ethical Considerations**

Even though the development of technology is inevitable in improving students' academic performance in various ways, it is important to ensure that the decision to utilize these tools should be guided not by their ease of use, but by their alignment with educational and curriculum goals (Li, 2021). Teaching objectives and expected learning outcomes should drive the implementation of technology, rather than being directed by the technology (Godwin-Jones, 2022).

### **Conclusion**

In summary, this paper explored the multifaceted role of AWE systems in English learning and teaching. Tracing their evolution from historical development to contemporary application, AWE systems have shown considerable promise in enhancing the writing skills of learners. While the benefits, such as immediate feedback and the ability to deal with large-scale student work, are significant, the limitations of AWE, particularly in addressing higher-order writing skills and its potential biases should not be ignored. The comparison between AWE and teacher feedback highlights the benefits of a hybrid approach in classrooms to maximize the effectiveness of writing instruction. Thus, the integration of AWE systems should be guided not only by technological capabilities but also by pedagogical considerations.

## References

- Attali, Y. (2013). Validity and reliability of automated essay scoring. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 181–198). Routledge/Taylor & Francis Group.
- Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series, 2011*(1), i–99. <https://doi.org/10.1002/j.2333-8504.2011.tb02241.x>
- Burstein, J., Riordan, B., & McCaffrey, D. (2020). Expanding automated writing evaluation. In *Handbook of automated scoring* (pp. 329–346). Chapman and Hall/CRC.
- Cahill, A., & Evanini, K. (2020). Natural language processing for writing and speaking. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring: Theory into practice* (pp. 69–92). CRC Press/Taylor & Francis Group.
- Charles, M., & Pecorari, D. (2015). *Introducing English for academic purposes*. Routledge.
- Cotos, E. (2014). Automated writing evaluation. In E. Cotos (Ed.), *Genre-based automated writing evaluation for L2 research writing* (pp. 1–64). Palgrave Macmillan.
- Davidson, L. Y. J., Richardson, M., & Jones, D. (2014). Teachers' perspective on using technology as an instructional tool. *Research in Higher Education, 24*, 1–25. <https://eric.ed.gov/?id=EJ1064110>
- El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies, 10*(2), 121–142. <https://doi.org/10.6018/ijes/2010/2/119231>
- Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning, 1-43*. <https://doi.org/10.1080/09588221.2020.1743323>
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology, 26*(2), 5–24. <https://hdl.handle.net/10125/73474>
- Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist, 53*(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277–303. <https://doi.org/10.3102/0091732x18821125>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(5), 523–547. <https://doi.org/10.1086/681947>
- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1–44. <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- Herrington, A., & Stanley, S. (2012). Criterion: promoting the standard. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 47–61). Peter Lang.
- Hockly, N. (2019). Automated writing evaluation. *ELT Journal, 73*(1), 82–88. <https://doi.org/10.1093/elt/ccy044>

- Huang, S., & Renandya, W. A. (2020). 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>
- Huang, Y., & Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*, 62, 102675. <https://doi.org/10.1016/j.compcom.2021.102675>
- Hyland, K. (2003). *Second language writing*. Ernst Klett Sprachen.
- Khoshnevisan, B. (2019). The affordances and constraints of automatic writing evaluation (AWE) tools: A case for Grammarly. *ARTESOL EFL Journal*, 2(2), 12–25.
- Lee, Y. J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *English Teaching*, 75(1), 67–92.
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, S., & Kunnan, A. J. (2016). Investigating the application of automated writing evaluation to Chinese undergraduate English majors: A case study of “WriteToLearn”. *CALICO Journal*, 33(1), 71–91.
- Lu, X. (2019). An empirical study on the artificial intelligence writing evaluation system in China CET. *Big Data*, 7(2), 121–129. <https://doi.org/10.1089/big.2018.0151>
- Morss, K., & Murray, R. (2001). Researching academic writing within a structured programme: Insights and outcomes. *Studies in Higher Education*, 26(1), 35–52.
- Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 38(2), 599–620. <https://doi.org/10.1111/jcal.12635>
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior*, 18, 103–134. [https://doi.org/10.1016/S0747-5632\(01\)00052-8](https://doi.org/10.1016/S0747-5632(01)00052-8)
- Shermis, M. D., & Hamner, B. (2013). Contrasting state-of-the-art automated scoring of essays. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 313–346). Routledge/Taylor & Francis Group.
- Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing* (pp. 125–142). Cambridge University Press. <https://doi.org/10.1017/9781108635547.009>
- Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, 91, 102247.
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11, 342–365. <https://doi.org/10.1287/isre.11.4.342.11872>
- Waer, H. (2021). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 1–25. <https://doi.org/10.1080/17501229.2021.1914062>

## Automated Writing Evaluation Feedback

- Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, 44, 100449.  
<https://doi.org/10.1016/j.asw.2020.100449>
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36.
- Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 36–54). Routledge/Taylor & Francis Group.
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102.