INTERNATIONAL CALL RESEARCH CONFERENCE

https://doi.org/10.29140/9780648184485-42

Exploring the Use of Automated Writing Evaluation in English Language Education: A Systematic Review

Andi Hidayat Thamrin*
Yogyakarta State University, Indonesia

Abstract

Automated writing evaluation (AWE) systems have emerged as indispensable tools in the realm of writing instruction, offering educators and students an efficient means to assess and enhance writing skills. In recent years, the proliferation of these systems has underscored the need for a thorough understanding of their functionalities, effectiveness, and implications for educational practice. Consequently, this study endeavors to address this imperative by conducting a systematic literature review spanning the years 2018 to 2023. My overarching objectives included (a) scrutinizing the landscape of AWE systems used during this timeframe, (b) dissecting the nature and scope of writing feedback provided by these systems, (c) delving into the perceptions and experiences of students and teachers regarding AWE systems, and (d) elucidating the contextual factors that shape the utilization and adoption of AWE systems across diverse educational settings.

Keywords: Automated writing evaluation, Automated writing feedback, Writing feedback

Introduction

Teachers spend much time evaluating their students' writing almost every day. It is a challenging task, especially if they have many students. Students, on the other hand, find it difficult to assess their own writing, since they might not know what to evaluate in the first place. Fortunately, technology has provided

*Corresponding Author E-mail: andi305fbs.2022@student.uny.ac.id

us with automated writing evaluation (AWE) systems that can help teachers evaluate their students' writing or allow students to evaluate their own. These systems work by comparing a written text to a substantial database of writing in the same genre, created in response to a specific prompt or rubric. To examine various measurable aspects of the text, such as syntax, complexity, total word count, and vocabulary range, these systems use statistical modeling and algorithms (Hockly, 2019).

The aims of this study were to provide an overview of AWE systems by identifying the specific tools used in evaluating writing from 2018 to 2023. This study detailed the types of feedback these systems provide and explored the perceptions of both students and teachers towards their use. Additionally, the study sought to understand the various educational and research contexts in which the AWE systems were implemented. This is important because it helps to elucidate the effectiveness and applicability of AWE systems in diverse settings.

For this purpose, the following research questions were used:

RQ 1: What AWE systems were used to evaluate students' writing in related research papers published between 2018 and 2023?
RQ 2: What kinds of writing feedback were given by AWE systems in related research papers published between 2018 to 2023?
RQ 3: In what contexts were the AWE systems used in related literature published between 2018 and 2023?

Methods

This study used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) rubric to achieve all of the aims. There are three steps in PRISMA, namely identification, screening, and inclusion (Page et al., 2021).

The process of selecting literature for this systematic literature review involved a comprehensive search conducted in the Scopus and Springer Link databases by using "automated writing evaluation" as the key phrase. These databases are two of the largest available and were chosen for this reason. Furthermore, to ensure relevance and accessibility, the selection criteria were limited to open-access research articles and conference papers. In addition, only papers published between 2018 and 2023 were considered, since AWE systems are a part of technology that keeps evolving. Limiting the publication years helped to broaden our insights on current AWE systems. This resulted in the identification of 39 and 386 articles in Scopus and Springer Link, respectively.

In total, 425 articles were found and subjected to further screening and analysis to identify the most relevant research for the review. To do this, the

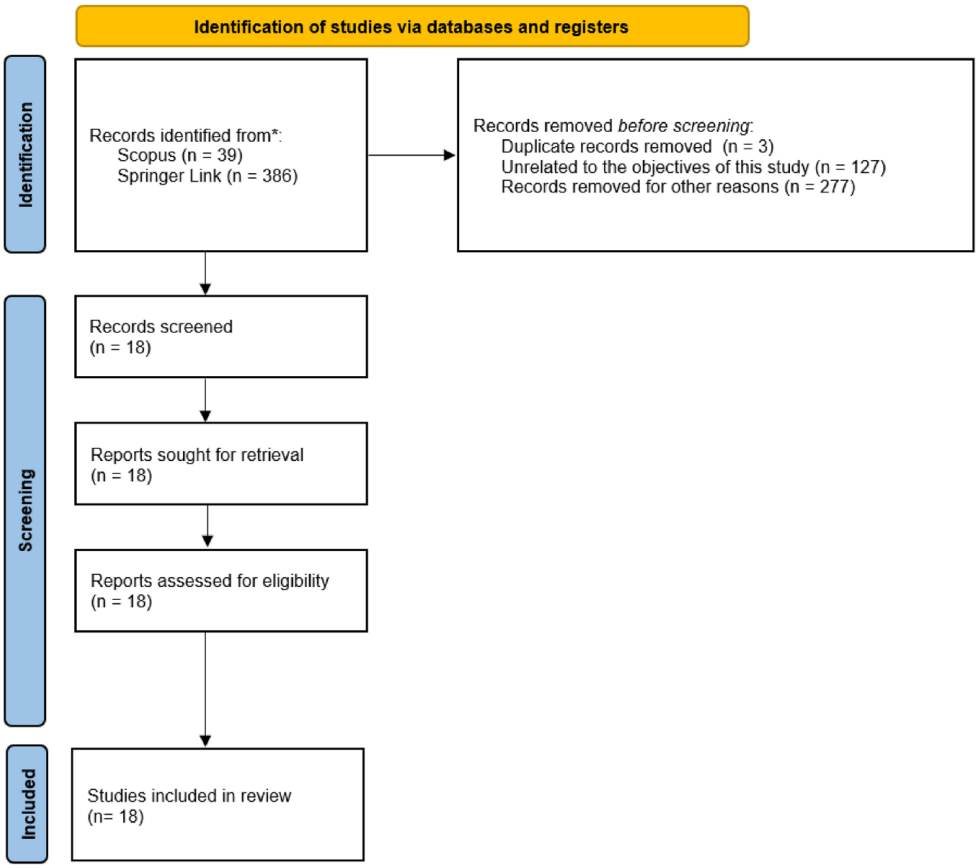

Figure 1. The process of searching for and selecting literature.

researcher read all the articles, including the title, abstract, and body. After the review, the researcher excluded three articles because of duplication in the databases, 127 articles that were unrelated to AWE, and 277 articles for other reasons. At the end of this phase, the researcher was left with 18 articles to be analyzed, as shown in Figure 1.

Results and Discussion

RQ 1

Between 2018 and 2023, various AWE systems were explored, revealing both strengths and limitations. Grammarly (www.grammarly.com) has been widely studied across multiple international settings (Miranty & Widianti, 2021; Tambunan et al., 2022; Dizon & Gayed, 2021; Parra et al., 2019; Thi & Nikolov, 2022; Yousofi, 2022), but has often been criticized for focusing too

heavily on grammar rather than broader writing skills. Pigai (www.en.pigai.org) has shown potential in Chinese contexts (Gao, 2021; Han et al., 2021; Yao, 2021), though it struggles with more complex writing issues. MI Write (www.miwrite.net) has been effective in US schools (Palermo & Wilson, 2020; Wilson et al., 2021) but has faced criticism for its generic feedback. WRITER has been used for error correction with English as a foreign language learners (Al-Inbari & Al-Wasy, 2022), yet its impact on broader writing skills is limited. Aim Writing and AI KAKU focus on grammatical competence and syntactic complexity, respectively, but have been critiqued for their narrow scopes (Chen & Pan, 2022; Gayed et al., 2022). AccurIT assists with constructing phrases (Grami, 2020), whereas Wordvice AI Proofreader (www.wordvice.ai) provides accuracy similar to that of human proofreaders (Heintz et al., 2022), though it lacks contextual feedback. Finally, Criterion offers detailed feedback but may not address individual learning needs fully (Saricaoglu & Bilki, 2021).

RQ 2

Grammarly excels in detecting a wide range of grammar issues, including prepositions, determiners, and subject–verb agreement, while also enhancing contextual spelling and lexical richness (Tambunan et al., 2022; Thi & Nikolov, 2022; Dizon & Gayed, 2021). It supports teachers by addressing error correction beyond its scope, though it sometimes falls short in areas like content and organization (Dizon & Gayed, 2021; Tambunan et al., 2022; Thi & Nikolov, 2022; Yousofi, 2022). Pigai offers diagnostic feedback on grammar, vocabulary, and sentence structure, promoting self-regulated learning through identifying errors (Yao, 2021; Gao, 2021; Han et al., 2021). MI Write focuses on feedback regarding the organization and style of writing, supporting peer evaluations and the quality of revisions but often requires better integration and the teachers' perceptions vary (Palermo & Wilson, 2020; Wilson et al., 2021). WRITER provides error correction and enhances clarity with collaborative features, and is noted for its ease of use and effective explanations of errors (Al-Inbari & Al-Wasy, 2022). Aim Writing supports English proficiency across educational levels but struggles with user engagement and the quality of feedback (Chen & Pan, 2022). AI KAKU improves writing complexity and provides suggestions, though it needs more user engagement to be effective (Gayed et al., 2022). AccurIT enhances vocabulary and corrects errors, and has received positive feedback for its opportunities for interaction (Grami, 2020). Wordvice AI Proofreader identifies errors well but needs improvement in handling stylistic issues (Heintz et al., 2022). Criterion delivers detailed feedback and supports various essay types but faces challenges with student engagement (Saricaoglu & Bilki, 2021).

RQ 3

Different AWE systems exhibit varying levels of effectiveness, depending on their context of application and the users' proficiency levels. For instance, Grammarly, which has been explored in diverse settings such as Indonesian universities (Miranty & Widianti, 2021; Tambunan et al., 2022), a Japanese university (Dizon & Gayed, 2021), an Ecuadorian university (Parra et al., 2019), Burmese universities (Thi & Nikolov, 2022), and Afghan private language schools (Yousofi, 2022), has proven useful across various educational contexts. However, its effectiveness can be influenced by the students' proficiency levels; Yousofi (2022) suggested that beginner students may find the excessive feedback overwhelming. Meanwhile, Gao (2021) observed that students with limited proficiency struggle to utilize the feedback effectively. Therefore, while AWE tools such as Grammarly can be used by students, they might still need assistance from their teachers.

Additionally, Pigai was used in research involving Chinese teachers and students (Gao, 2021; Han et al., 2021; Yao, 2021). Additionally, MI Write was studied in North Carolina and mid-Atlantic elementary schools (Palermo & Wilson, 2020). These studies offer insights into different educational levels and settings. The effectiveness of AWE systems like WRITER (Al-Inbari & Al-Wasy, 2022), Aim Writing (Chen & Pan, 2022), and AI KAKU (Gayed et al., 2022) further underscores the variability in AWE systems' utility depending on the specific educational and geographical contexts.

Furthermore, the application of these systems is often limited by factors such as internet access and the availability of electricity, which can affect their effectiveness in regions with infrastructural constraints (Gao, 2021). As a teacher, it is crucial to recognize that the effectiveness of educational technologies can be significantly impacted by factors like limited internet access and unreliable electricity. In areas with inadequate infrastructure, these systems might not function optimally, affecting both teaching and learning experiences. This limitation can hinder the ability to fully integrate technology into the classroom, highlighting the need for practical solutions and support to address these infrastructural challenges and ensure that all students have equitable access to educational resources.

Conclusion

This systematic review of AWE systems from 2018 to 2023 has revealed a diverse landscape of tools, each with its strengths and limitations. Grammarly stood out as one of the most widely studied across international contexts.

This proves its effectiveness in identifying grammatical errors, although it often overlooks broader writing skills. Other systems like Pigai, MI Write, and WRITER have shown promise in specific educational settings.

The review highlights the importance of considering the context of application when selecting AWE systems, as factors such as the educational level, the users' proficiency, and geographical constraints significantly impact their utility. Therefore, it is important to note that the effectiveness of these tools varies depending on the context and users' proficiency levels. Although these systems can provide valuable support in writing instruction, their effectiveness is often moderated by the students' ability to engage with the feedback. In addition, the availability of infrastructure, such as reliable internet and electricity, may also affect how effective the tools can be.

Ultimately, this review underscores the need for educators to critically evaluate AWE tools, accounting for the specific needs of their students and the constraints of their educational environment. While AWE systems offer valuable assistance in enhancing writing skills, they should be used in conjunction with teacher guidance to address their limitations and maximize their potential in diverse educational contexts.

Funding

This work was funded by the Indonesian Endowment Fund for Education under Grant number 202207111210965.

Acknowledgments

The researcher expresses his gratitude to the Indonesian Endowment Fund for Education for funding his study, and especially for the opportunity to participate in this international conference.

References

Al-Inbari, F. A. Y., & Al-Wasy, B. Q. M. (2022). The impact of automated writing evaluation (AWE) on EFL learners' peer and self-editing. *Education and Information Technologies, 28*, 6645–6665 https://doi.org/10.1007/s10639-022-11458-x

Chen, H., & Pan, J. (2022). Computer or human: A comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing. *Asian-Pacific Journal of Second*

and Foreign Language Education, 7(1), 34. https://doi.org/10.1186/s40862-022-00171-4

Dizon, G., & Gayed, J. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, 17(2), 74–92. https://doi.org/10.29140/jaltcall.v17n2.336

Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system Pigai. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(11), 322. https://doi.org/10.3991/ijet.v16i11.19657

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Grami, G. M. A. (2020). An evaluation of online and automated English writing assistants: Collocations and idioms checkers. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(4), 218. https://doi.org/10.3991/ijet.v15i04.11782

Han, Y., Zhao, S., & Ng, L.-L. (2021). How technology tools impact writing performance, lexical complexity, and perceived self-regulated learning strategies in EFL academic writing: A comparative study. *Frontiers in Psychology, 12*, 752793. https://doi.org/10.3389/fpsyg.2021.752793

Heintz, K., Roh, Y., & Lee, J. (2022). Comparing the accuracy and effectiveness of Wordvice AI Proofreader to two automated editing tools and human editors. *Science Editing, 9*(1), 37–45. https://doi.org/10.6087/kcse.261

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

Miranty, D., & Widianti, U. (2021). An automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts. *Pegem Journal of Education and Instruction, 11*(4), 126–137. https://doi.org/10.47750/pegegog.11.04.12

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, 71. https://doi.org/10.1136/bmj.n71

Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, *12*(1), 63–108. https://doi.org/10.17239/jowr-2020.12.01.04

Parra, L. G., & Calero, X. S. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction, 12*(2), 209–226. https://doi.org/10.29333/iji.2019.12214a

Saricaoglu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. ReCALL*: Journal of Eurocall, 33*(3), 265–277. https://doi.org/10.1017/S0958344021000021

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

Thi, N. K., & Nikolov, M. (2022). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher, 31*(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Wilson, J., Huang, Y., Palermo, C., Beard, G., & MacArthur, C. A. (2021). Automated feedback and automated scoring in the elementary grades: Usage, attitudes, and associations with writing outcomes in a districtwide implementation of MI Write. *International Journal of Artificial Intelligence in Education, 31*(2), 234–276. https://doi.org/10.1007/s40593-020-00236-w

Yao, D. (2021). Automated writing evaluation for ESL learners: A case study of Pigai system. *The Journal of Asia TEFL*, *18*(3), 949–958. https://doi.org/10.18823/asiatefl.2021.18.3.14.949

Yousofi, R. (2022). Grammarly deployment (in)efficacy within EFL academic writing classrooms: An attitudinal report from Afghanistan. *Cogent Education, 9*(1), 2142446. https://doi.org/10.1080/2331186X.2022.2142446