Bibliometrically and systematically analyzing automated writing evaluation for English learning

Tiantian Wu[1], Zhonggen Yu[2,*]

[1] *Faculty of Foreign Studies, Beijing Language and Culture University, Beijing 100083, China*

[2] *Faculty of Foreign Studies, Academy of International Language Services, National Base for Language Service Export, Beijing Language and Culture University, Beijing 100083, China*

* Corresponding author: Zhonggen Yu, yuzhonggen@blcu.edu.cn

ABSTRACT: Automated writing evaluation is highly discussed in artificial intelligence for English learning. It is necessary to explore the effect of automated writing evaluation on learning English as a second language. This study combined bibliometric analysis and systematic review to explore the use of automated writing evaluation for learning English as a second language. VOSviewer was used to identify the highly discussed topics, the top ten cited authors, organizations, countries, references, and sources in the studies on automated writing evaluation. Fifty-six peer-reviewed articles were selected according to the Preferred Reporting Items for Systematic Review and Meta-analysis Protocols. The analysis revealed that automated writing evaluation is helpful, but its effectiveness varies according to the types of feedback, and it cannot replace the role of human feedback yet. Teachers' roles are significant in integrating automated writing evaluation into the classroom. Future research could focus on the specific ways to integrate automated writing evaluation into the classroom better.

KEYWORDS: automated writing evaluation (AWE); writing; English learning; feedback; bibliometric analysis

1. Introduction

The development of artificial intelligence (AI) has exerted a great influence on people's daily lives, including the learning of English as a foreign or second language. In a bibliometric analysis of studies about AI integrated into language education, it was found that AI has been frequently used to aid the language teaching and learning process, including vocabulary, writing, reading, speaking, and listening, with automated writing evaluation (AWE) presented as the most popular topic (Huang et al., 2023). AWE systems were developed based on interdisciplinary research and technological advances such as natural language processing, computer sciences, and latent semantic analysis (Shi and Vahid, 2022). Learners can use AWE in the self-learning process, and it has also been introduced into the classroom to reduce teachers' workload and improve students' writing skills by providing timely feedback (Feng and Chukharev-Hudilainen, 2022).

However, the practical use of AWE in the real learning process may come across many problems, partly as a result of the many different software and platforms providing AWE feedback for English learners. Several questions need to be asked: First of all, is the feedback beneficial to students' improvements in essay quality and writing ability? What are the attitudes and perceptions reflected in

different types of uses? When it is integrated into the classroom, can students and teachers accept it well and integrate it into the traditional teaching and learning process naturally? The development of AI tries to provide people with more convenient and efficient ways, but its implementation should always find a proper way so that it can play a positive rather than negative role. Related studies must explore and ensure its effectiveness.

Previous studies have mainly focused on the specific aspect of effectiveness or a survey about users' perceptions and attitudes. For example, an experiment was conducted to assess the reliability of AWE scoring (Bridgeman and Ramineni, 2017). Researchers have also investigated the effectiveness of the AWE feedback of one specific platform, like a comparison between Pigai and teacher feedback (Gao, 2021). The present study aims to use bibliometric analysis and systematic review to present an overall knowledge of present AWE platforms or software. Therefore, researchers could find the highly discussed topics in this field and the influential authors, organizations, references, and sources about the studies of AWE. Thus, researchers could contribute to a further discussion about AWE's impacts on the user, factors influencing the effectiveness of AWE feedback, and problems when integrated into the classroom.

2. Literature review

2.1. Users' engagement and perception of AWE feedback

Both teachers and students kept a positive attitude toward AWE feedback for its immediate and clear feedback, time-saving, and arousing interest in English writing (Lu, 2019). Students' perception of the usefulness of AWE feedback was highly related to their language proficiency (Xu and Zhang, 2022). Students' positive or negative attitude slightly influenced their decision whether or not to use the AWE feedback in a short time, but would highly affect their continuing use and recommendation of it to friends (Roscoe et al., 2017). The frequency of use affected their perceptions of the usefulness and drawbacks of AWE feedback (Miranty and Widiati, 2021). If students perceived more usefulness and fewer drawbacks in the high-frequency use of AWE feedback, this would lead to a higher acceptance of AWE. Because of students' different perceptions of AWE feedback, their perception of the feedback source would influence their writing performance (Reynolds et al., 2021).

Not all students used AWE feedback for revisions, but the adoption of AWE could significantly reduce the number of errors in their essays (Saricaoglu and Bilki, 2021). Their engagement was reflected in perceptions of AWE feedback and the use of revision operations (Z. Zhang, 2020). The positive effects of AWE feedback also was a reliable indicator of how students engaged with it behaviorally, emotionally, and cognitively (Z. Zhang, 2017). The feedback explicitness was the most critical factor influencing users' engagement with AWE feedback (Liu and Yu, 2022). Facing some generic rather than specific feedback, students would perceive a higher level of mental effort expenditure and lower levels of clarity and helpfulness (Ranalli, 2018). Previous studies mostly concluded that high engagement was due to learners' writing proficiency, while learners' trust was a determining factor (Ranalli, 2021).

2.2. The integration of AWE feedback in classroom

Teachers took different approaches to integrate AWE feedback into the classroom so that it could achieve the purpose of relieving the workload and improving students' writing performance. AWE feedback's implementation and efficacy were influenced by the instructional activity system (Wilson et al., 2021). Different approaches employed by the teacher could lead to a significant difference in students' essay submission frequencies, revision types, and changes in error rates (Z. Li, 2021). AWE feedback took characteristics consistent with a framework for deliberate practice when it was integrated into

different instructional contexts (Palermo and Wilson, 2020). In general, incorporating AWE into a program of writing instruction was beneficial to students' writing quality development (Palermo and Thomson, 2018).

The digital tools, including AWE feedback, could provide great benefits for both teachers and students when they are integrated into the classroom (Godwin-Jones, 2022). AWE feedback can be used to differentiate students' writing ability levels and predict the holistic scores of students' essays (Lim et al., 2022). It supported teachers' educational decisions, thus assisting teachers in identifying students' improvement of writing skills (Petchprasert, 2021). Many AWE feedback programs or applications were equipped with the scoring ability and were promised to be ideal helpers to teachers to evaluate students' writing performance (Z. Li, Link, Ma, et al., 2014).

2.3. Research questions

Based on the literature review, the researchers could find plenty of research about automated writing evaluation (AWE) feedback. However, most of the studies focused on a single platform and a single function of AWE feedback, and few researchers viewed the related studies in a more holistic way. In this research, through a bibliometric and systematic analysis of relevant studies, researchers hoped to have a comprehensive knowledge of AWE's research situation and find the common problems with AWE and its implementation. This research aimed to find the highly discussed topic in recent years and the influential researchers, references, organizations, and countries in related research fields. Based on that, the researchers aimed to find the impacts of AWE feedback, factors influencing its effectiveness, and problems in the implementation process in the classroom. The research questions are presented as follows:

RQ1: What are the highly discussed topics related to automated writing evaluation (AWE) in learning English as a second language?

RQ2: What are the top ten cited authors, organizations, countries, references, and sources among the studies on AWE in learning English as a second language?

RQ3: What are the impacts of AWE feedback on users' writing development in learning English as a second language?

RQ4: What are the factors influencing the effectiveness of AWE feedback in learning English as a second language?

RQ5: What are the problems when AWE is integrated into the classroom in learning English as a second language?

3. Research method

The authors combined the systematic review and bibliometric analysis in this study to explore the use situation of AWE systems. For the bibliometric analysis, the authors mainly used the VOSviewer. We mainly applied the rapid evidence assessment review methods for the systematic review. Rapid evidence assessment is a process that uses a combination of key informant interviews and targeted literature searches to produce a report in a few days or a few weeks. This method provides a more structured and rigorous search and quality assessment of the evidence than a literature review but is not as exhaustive as a full systematic review. The whole research mainly included the following steps. Firstly, the first author searched the online database Web of Science and obtained extensive literature about AWE. Secondly, the VOSviewer was applied to visualize the literature to find the top ten cited authors, organizations, countries, references, and sources according to different criteria. Thirdly, we selected the

obtained literature according to the inclusion and exclusion criteria. Finally, authors reviewed the selected literature carefully and answered the research questions. The above content was a brief introduction to the research steps, and the following paragraph was detailed explanations for the concrete operation of each step.

Researchers collected the literature in the Web of Science on 27 July 2022. Researchers obtained 928 results by keying in "'automt*' (topic) AND 'writ*' (topic) AND 'evaluation' (topic)". The online database covered Science Citation Index Expanded (2013–2022), Social Science Citation Index (2008–2022), Arts & Humanities Citation Index (2008–2022), Emerging Sources Citation Index (2017–2022), Current Chemical Reactions (1985–2022), and Index Chemicus (1993–2022). To ensure the reliability and representativeness of the literature, the authors checked the top ten categories of searched literature through the Web of Science. The main categories of literature were Education Educational Research, Computer Science Information Systems, Linguistics, Computer Science Software Engineering, Computer Science Artificial Intelligence, Computer Science Interdisciplinary Applications, Engineering Electrical Electronic, Language Linguistics, Computer Science Theory Methods, and Medical Informatics (see **Figure 1**). All of these categories were highly related to the topic, and the types of categories were various. Therefore, we could say that the searched literature was reliable and representative for the analysis.
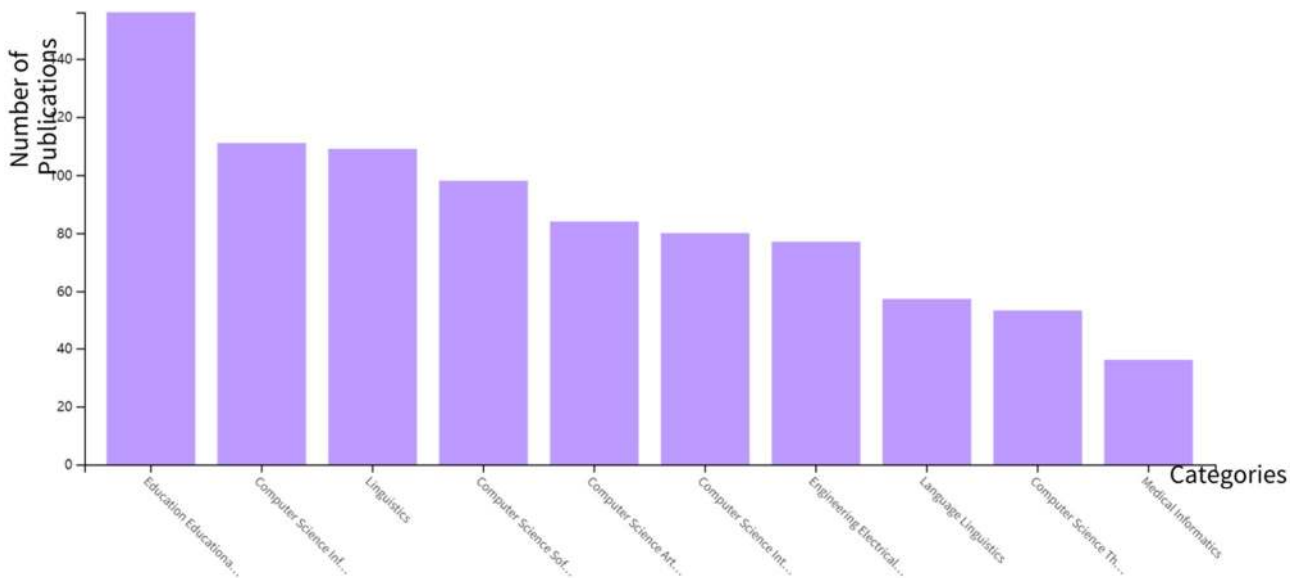

Figure 1. The top ten categories of literature.

The researchers applied the VOSviewer to find the highly discussed topics related to AWE and used it to visualize the top ten authors, organizations, and countries and cited references, sources, and authors among the studies on AWE. VOSviewer is a useful tool to visualize literature in the bibliometric analysis process (van Eck and Waltman, 2017). The clustering density could show readers a clear map of the most important issues and areas discussed by previous studies in the collected literature. For example, in the clustering map based on author keywords, the term in the circle of the biggest size indicates this term is the topic discussed most frequently in the searched literature. Furthermore, the line's thickness shows readers the relevancy between two terms connected by the line.

After the visualization of the literature, researchers had a careful selection of the obtained 928 articles based on the Preferred Reporting Items for Systematic Review and Meta-analysis Protocol (PRISMA-P) (see **Figure 2**) (Shamseer et al., 2015). PRISMA is an evidence-based minimum set of items for reporting

in systematic reviews and meta-analyses. PRISMA primarily focuses on the reporting of reviews evaluating the effects of interventions, but can also be used as a basis for reporting systematic reviews with objectives other than evaluating interventions (Page et al., 2021). Two researchers participated in the main selection of literature with high inter-rater reliability ($k = 0.910$). If the two researchers could not come to an agreement, the third researcher would be invited to make the decision. Articles would be included if they were (1) highly related to the topic, (2) conducted based on rigorous research design, and (3) concluded convincingly with enough scientific evidence. The literature will be excluded if they (1) were weakly related to the topic, (2) had no profound research design, and (3) presented unconvincing conclusions. Detailed information about the fifty-six articles was provided in the Appendix (see **Table A1**), including the research object, research focus, and methods.
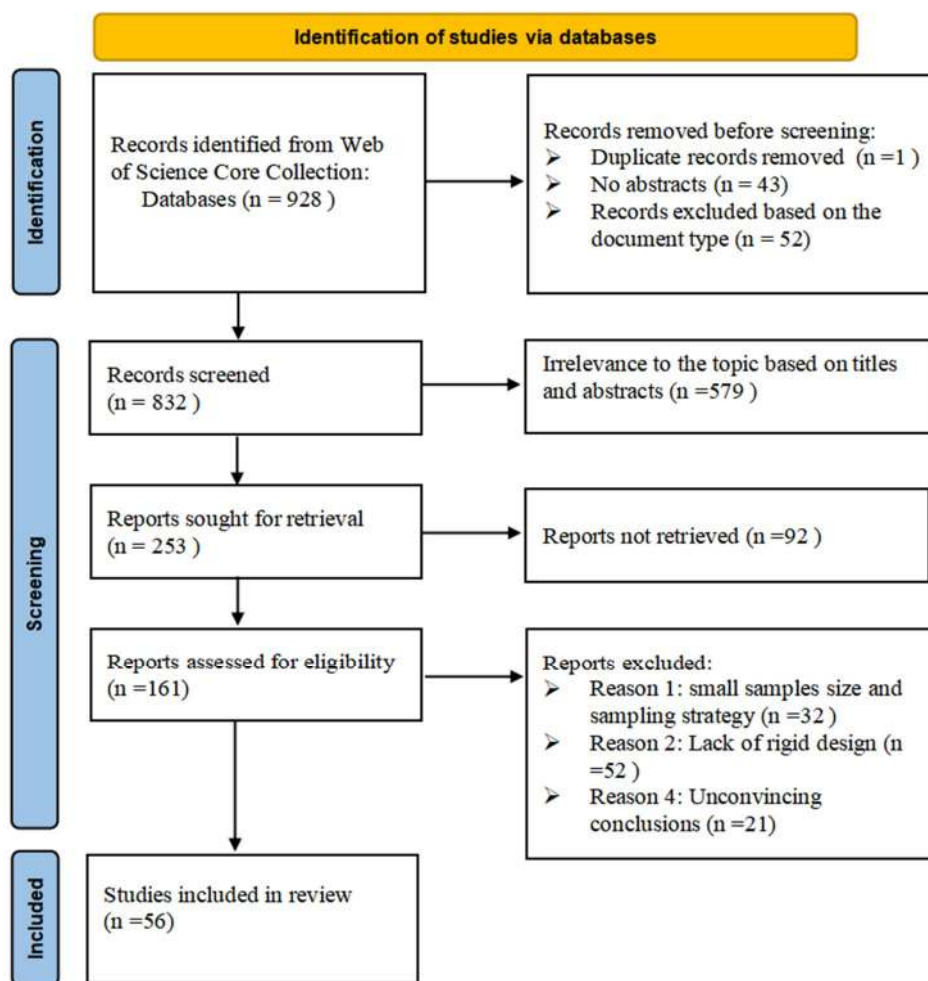
Identification of studies via databases

Identification

Records identified from Web of Science Core Collection: Databases (n = 928)

→ Records removed before screening:
➢ Duplicate records removed (n =1)
➢ No abstracts (n = 43)
➢ Records excluded based on the document type (n = 52)

Screening

Records screened (n = 832)

→ Irrelevance to the topic based on titles and abstracts (n =579)

Reports sought for retrieval (n = 253)

→ Reports not retrieved (n =92)

Reports assessed for eligibility (n =161)

→ Reports excluded:
➢ Reason 1: small samples size and sampling strategy (n =32)
➢ Reason 2: Lack of rigid design (n =52)
➢ Reason 4: Unconvincing conclusions (n =21)

Included

Studies included in review (n =56)

Figure 2. A flow chart of the literature inclusion based on PRISMA-P.

4. Results

In this part, researchers discussed and provided answers to the five research questions. By creating maps based on bibliographic data, researchers found important issues related to AWE. The map showed that machine learning, L2 writing, and feedback were discussed frequently in the field related to AWE in learning English as a second language. Researchers also used VOSviewer to find the top ten authors, organizations, and countries and cited references, sources, and authors. The results provided us with

influential references from different dimensions. Then the authors investigated the impact of AWE on its uses, factors influencing AWE's effectiveness, and the problems in integrating AWE into the classroom.

4.1. RQ1: What are the highly discussed topics related to automated writing evaluation (AWE) in learning English as a second language?

4.1.1. Cluster mapping based on author keywords

To visualize the highly discussed topics related to AWE, researchers created a map based on bibliographic data about the co-occurrence of author keywords by VOSviewer. The type of analysis was co-occurrence, and the unit of analysis was author keywords. Setting the minimum number of occurrences as five, 68 author keywords met the threshold of the total 3206 keywords (see **Figure 3**). All of their occurrences and total link strength were calculated. The researchers selected the top ten author keywords with the greatest occurrences and link strength: automated writing evaluation ($N = 52$, link strength = 74), natural language processing ($N = 58$, link strength = 66), writing ($N = 34$, link strength = 64), machine learning ($N = 51$, link strength = 56), artificial intelligence ($N = 27$, link strength = 30), automated essay scoring ($N = 18$, link strength = 27), feedback ($N = 15$, link strength = 27), feature extraction ($N = 10$, link strength = 26), L2 writing ($N = 13$, link strength = 26), and deep learning ($N = 29$, link strength = 23).
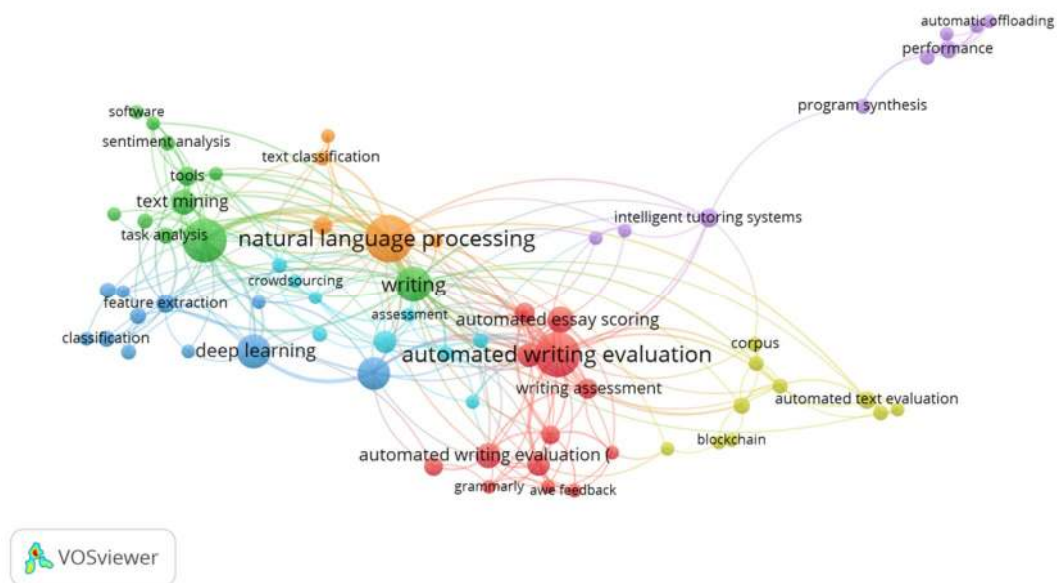

Figure 3. A clustering map of author keywords.

The 68 items were classified into seven clusters. Cluster 1 included 13 items, such as automated writing evaluation, AWE feedback, second language learning, teacher feedback, and student engagement. Cluster 2 included 11 items, such as analytical model, machine learning, model checking, sentimental analysis, task learning, and text mining. Cluster 3 included 11 items, such as classification, deep learning, feature extraction, handwriting, natural language generation, and performance evaluation. Cluster 4 included 10 items, such as automated text evaluation, blockchain, curriculum-based measurement, reliability, and written expression. Cluster 5 included nine items. These items were mainly about computer programs, such as algorithms, automated offloading, intelligent tutoring system, interactive learning environments, and latent semantic analysis. Cluster 6 included nine items, such as academic writing, assessment, crowdsourcing, and formative assessment. Cluster 7 included five items: electronic

health records, information retrieval, natural language processing, systematic review, and text classification.

From the above clusters, we found that automated writing evaluation ($N = 50$, link strength = 65) was the author keyword with the highest occurrence and total link strength. Machine learning, artificial intelligence, writing, and feedback were important issues related to AWE. Therefore, researchers could get a great deal of information and references for these topics.

4.2. RQ2: What are the top ten cited authors, organizations, countries, references, and sources among the studies on automated writing evaluation (AWE) in learning English as a second language?

4.2.1. Clustering based on the authors, organizations, and countries

The researchers used the VOSviewer to visualize the top ten authors, organizations, and countries. In this way, researchers could find authors, organizations, and countries that contribute most to this field. After that, researchers could get abundant and valuable information from the literature of these authors, countries, and organizations. The type of analysis was co-authorship, and the counting method was full counting. Then researchers set authors, organizations, and countries as the unit of analysis respectively. In the bibliographic analysis of the three units of analysis, researchers set the minimum number of documents, organizations, or countries as three or five. The VOSviewer presents the citations and total link strength of each item that meets the threshold. Researchers selected the top ten items with total link strength as the main criterion and the citations as the second criterion. The top ten authors, organizations, and countries in terms of publication are presented in **Table 1**.

Table 1. Top ten authors, organizations, and countries.

N	Authors	Organizations	Countries
1	Wilson, Joshua	University of British Columbia	USA
2	Chao, Fei	University of Delaware	England
3	Lin, Chih-min	University of Houston	China
4	Yang, Longzhi	The Hong Kong Polytechnic University	Germany
5	Zhou, Changle	National Kaohsiung University of Applied Sciences	Switzerland
6	Cheng, Gary	National Taiwan Normal University	Canada
7	Myers, Matthew C.	CNRS	Spain
8	Potter, Andrew	Educational Testing Service	Australia
9	Xie, Haoran	INRIA	France
10	Zou, Di	Inserm	Netherlands

The author with the greatest total link strength was Joshua Wilson, with 15 documents. Its total link strength was 13, with a high citation of 252. The organization with the greatest total link strength was the University of British Columbia. Nine documents belonged to this organization. Its greatest total link strength was seven, and the citation was 34. The country with the greatest total link strength was the USA. Its total link strength was 122, much greater than the country with the second-greatest total link strength. The country with the second-greatest total link strength was England, with a 76 total link strength. Therefore, we can say that the USA has a dominant influence in this field. 239 documents were from this country, and the citation was 4567.

4.2.2. Clustering based on cited references, cited sources, and cited authors

The VOSviewer was also used to find the top ten cited references, sources, and authors. The highly cited number shows its influence and specialty in this field. Through the bibliographic analysis by VOSviewer, researchers could find the most influential literature, sources, and authors in this field. Thus, researchers could find the most valuable references about automated wiring evaluation. The type of analysis was co-citation, and the counting method was full counting. Then researchers set cited references, cited sources, and cited authors as the unit of analysis, respectively, and the minimum number of citations of a cited reference, source, or author was set as 20. We obtained all the items that met the threshold and their citations and total link strength. The top ten cited references, sources, and authors according to total link strength and citations are presented in **Table 2**.

Table 2. Top ten cited references, sources, and authors.

N	Cited references	Cited sources	Cited authors
1	(Stevenson and Phakiti, 2014)	Assessing Writing	Graham, S
2	(C. F. E. Chen and Cheng, 2008)	Lecture Notes in Computer Science	Wilson, J
3	(Wilson and Czik, 2016)	Computer Assisted Language Learning	Warschauer, M
4	(J. Li et al., 2015)	Journal of Second Language Writing	Shermis, M. D.
5	(Warschauer and Ware, 2006)	Computers and Education	Roscoe, R. D
6	(Y. J. Wang et al., 2013)	Journal of Educational Psychology	Stevenson, M.
7	(Warschauer and Grimes, 2008)	arXiv	Kellogg, R. T.
8	(Diklil and Bleyle, 2014)	Reading and Writing	Ranalli, J
9	(Grimes and Warschauer, 2010)	System	Crossley, S. A.
10	(Bai and Hu, 2017)	Language Learning and Technology	Attali, Y

In the bibliographic analysis of cited references, 38 references met the threshold of the 34,950 cited references. The cited reference with the greatest total link strength was the reference (Stevenson and Phakiti, 2014). Its citations were 53, and the total link strength was 461, showing its influence on research in this field. In the bibliographic analysis of cited sources, 264 cited sources met the threshold of 16,286 sources. The cited source with the greatest total link strength was Assessing Writing. Its citations were 408, and its total link strength was 12,782. Therefore, researchers can see that Assessing Writing was influential in studies about AWE. In the bibliographic analysis of cited authors, 81 authors met the threshold of a total of 24,497 authors. The author with the greatest total link strength was Graham. This author had a total link strength of 3691, and the citations were 154. The high total link strength and citation indicated that this author's documents were cited by other documents about AWE most frequently. Thus, researchers could find significant information in this author's documents.

4.3. RQ3: What are the impacts of AWE feedback on users' writing development in learning English as a second language?

The AWE overall positively affected students' writing quality (Zhai and Ma, 2023). AWE feedback was a good strategy to help students achieve writing goals and relieve their writing anxiety (Z. Chen et al., 2022). It can promote students' learning behavior and attitudinal technology acceptance (Nazari et al., 2021). Obvious enhancement could be observed in students' writing accuracy and learner autonomy awareness after using AWE feedback (Y. J. Wang et al., 2013). The use of AWE systems helped students focus and enhance lexical and grammatical use to achieve their writing goals (Feng and Chukharev-

Hudilainen, 2022). Students can be motivated by AWE feedback use, and they thought its use was useful, although students could perceive the shortcomings of the feedback (Fu et al., 2022).

Furthermore, the impact of AWE feedback varied across levels of students, genres of writing, and time of using AWE. For different levels of students, AWE feedback was more beneficial to higher-level learners' autonomy development (Koltovskaia, 2020). For lower-level students, AWE feedback was more effective in improving their single essay quality (Butterfuss et al., 2022). In improving writing fluency and accuracy, AWE feedback was more effective for beginners and intermediate learners than advanced learners (Mohsen, 2022). Considering the writing genre, AWE use showed more effectiveness in argumentative writing than in academic writing and mixed writing genres (Zhai and Ma, 2023). Regarding the long-term and short-term use of AWE, researchers could observe more significant improvement in essay quality in the short term rather than the long term (Z. Li, Feng, and Saricaoglu, 2017). Besides, the long-term impact, which means the impact on students' writing ability rather than the quality improvement of a single essay after receiving feedback, was found earlier among students with higher-level writing ability or willingness (Liao, 2016).

However, in some specific aspects, AWE feedback was less effective than the other two types of feedback compared with it, including peer feedback and teacher feedback. Peer feedback performed much better than AWE feedback in improving the essay's cohesion and coherence (M. Chen and Cui, 2022). Similarly, AWE could improve students' writing from a product-oriented aspect, but it was not as effective as human feedback (Fu et al., 2022). AWE feedback was also ineffective in improving essays' syntactic complexity and fluency (Xu and Zhang, 2022). Besides, AWE feedback was revealed to hinder the use of monitoring strategies, while teacher and peer feedback positively facilitated these strategies (J. Zhang and Zhang, 2022).

4.4. RQ4: What are the factors influencing the effectiveness of AWE feedback in learning English as a second language?

The types and categories of AWE feedback influenced its effectiveness. Generic feedback was less effective than specific feedback in students' successful corrections (Ranalli, 2018). AWE feedback appeared to be more efficient in subject-verb agreement correction than verb tense correction (Koltovskaia, 2020). The low-level feedback had little to no effect on the properties of students' essays (Allen et al., 2019). Besides, the accuracy and precision rate of AWE feedback varied across the feedback categories, and it is lower than that of peer and teacher feedback (Bai and Hu, 2017). However, it was better that students could be selective about the use of AWE feedback, assessing for themselves whether it was accurate and whether they would use it for their own corrections (Bai and Hu, 2017). Although students' assessment of feedback greatly depends on their personal language ability, this behaviour could improve the accuracy of feedback they receive.

AWE feedback's effectiveness varies across different software or platforms. Various kinds of software and platforms online provide AWE feedback, including Pigai in China, Grammarly, OpenEssayist, iWrite, and so on. Pigai cannot perform well in collocation and syntax but is competent in lexical error identification (Gao, 2021). Besides, the use of Pigai needs to improve students' lexical complexity in the writing process (Y. Han et al., 2021). Comparatively, Grammarly's error correction was more impressive because of the significant decrease in error scores after revision (Guo et al., 2021). Grammarly was demonstrated to be a potentially valuable AWE feedback platform, while Pigai did not (Ngo et al., 2022).

4.5. RQ5: What are the problems when AWE is integrated into the classroom in learning English as a second language?

When AWE is integrated into the classroom, AWE feedback can function as a good supplement instrument rather than a replacement for teacher or peer feedback. Complete AWE feedback without humans could result in frustration for students and limit their development of writing skills (C. F. E. Chen and Cheng, 2008). When it complemented teacher feedback, it was revealed to be positive (Parra and Calero, 2019). When it must be used with peer and teacher feedback, it inevitably influences the other two types of feedback (Jiang et al., 2020). Teachers provided the same amount of higher-level feedback but less lower-level feedback when using AWE feedback as a supplemental instrument compared to total teacher feedback (Link et al., 2022). Although the integration of AWE and teacher feedback did not significantly outperform teacher feedback only (Fan, 2023), the use of AWE could greatly aid teachers in saving time and energy.

Teachers' and students' attitudes and digital literacy were significant in the effectiveness of AWE feedback in the classroom (C. F. E. Chen and Cheng, 2008). AWE implementation in the classroom caused impacts on teacher feedback, and these impacts were mediated by teacher beliefs about AWE and students, teacher willingness to offer the scaffolding, and contextual factors (Jiang et al., 2020). For teachers, their willingness to apply AWE is heavily influenced by the application's effectiveness, efficiency, and complexity (Du and Gao, 2022). Students' reasons for not using AWE feedback include technical reasons and the need for more time to learn using the software when they can access AWE feedback (Foster, 2019). Some AWE platforms possess the peer review feature, which could allow teachers to create peer review groups of 2–5 students, and students could write comments and suggestions to improve the essay quality. However, study results showed that this supplemental pedagogical function was not utilized well (Potter and Wilson, 2021).

AWE score was not effective enough. One experiment suggested that AWE's use for predicting scores is less effective than traditional methods (Bridgeman and Ramineni, 2017). The use of sentences with syntactic complexity cannot be presented by scores generated by AWE (Qian et al., 2021). AWE platform and teachers gave completely inconsistent scores and grades to the same essay despite extensive detailed standards and procedures (Sari and Han, 2022). However, different platforms' abilities to predict scores were different. For example, InferSent was more effective in predicting human grades than others (Q. Wang, 2022). AWE's scoring ability was closely related to its effectiveness and scoring standards. AWE's scores could reflect users' performance on specific aspects, but may fail to evaluate users' whole writing performance. It could be a reference, but still could not be applied to the formal ability evaluation process as a critical indicator.

5. Discussion

5.1. Making use of the impacts of AWE in an appropriate way

AWE feedback impacted the user's writing ability positively and negatively in different aspects. Therefore, it is necessary to avoid the negative influence and use its positive aspects in learning English as a second language with AWE feedback (Yu, 2022). Students can use it to improve essay quality in the learning process to achieve a short-term goal. However, AWE feedback cannot help users make long-term progress in their writing ability. It is also ineffective for users to improve syntactic complexity and fluency (Xu and Zhang, 2022). Therefore, teachers and students cannot depend on using AWE feedback (Yu et al., 2019). User should possess the awareness that AWE feedback may be incorrect or

inappropriate for their own essays, and be careful when receiving the AWE feedback. Authors suggested that platform designers should pay enough attention to feedback not received and improve the feedback accuracy constantly. Designers also could provide a function where users could discuss feedback, they are uncertain about with accuracy, thus providing a communication platform for different learners.

5.2. Methods to improve the effectiveness of AWE feedback

Different aspects of AWE feedback need to be enhanced. First, the effectiveness of AWE feedback should be improved by elevating accuracy and explicitness. After all, improving essay quality and writing competence is the primary and ultimate goal of using AWE feedback. The feedback accuracy greatly influenced users' response accuracy directly or indirectly (Guo et al., 2021). Improving accuracy may be difficult because language usage must consider many factors, like meanings and different rules in different situations. Authors suggest that platform designers should pay close attention to feedback not received by users and try to develop good relationships with educators. AWE platform's development could obtain inspiration from the actual implementation and the users. Secondly, AWE software or platforms should be designed for convenient operation. Perceived ease of use was the most critical factor influencing students' intention to continue using AWE (R. Li, 2021). At the same time, teachers should be more active in developing their digital literacy for better adaptation to kinds of technology assisting the education process (Yu, 2020). Furthermore, AWE feedback could be added in students' first language, which can facilitate students having a positive attitude towards the feedback and increase the noticing of errors (Wilken, 2018).

5.3. Methods to facilitate the integration of AWE into the classroom

Teacher mediation played an important role. Their mediation helps students be informed consumers and gain meta-linguistic language (Godwin-Jones, 2022). With the teacher's support, it can also be considered an important sociocultural artifact mediating the integration of mediated learning experience theory (R. Li, 2021). Besides, Lai's research found that students generally opted for peer feedback compared with AWE feedback (Lai, 2010), and students tended to receive teacher feedback more frequently compared with AWE feedback (Link et al., 2022). Furthermore, students' perception of AWE feedback varied depending on their proficiency level and instructors' perception (J. Li, Link, and Hegelheimer, 2015). Therefore, teachers' mediation for students' better use of AWE was important.

Teachers and students should integrate the use of three types of feedback harmoniously. Although the effectiveness of AWE feedback alone was lower than peer and teacher feedback, the study showed that the combined automated-teacher feedback was equally competent to full teacher feedback in improving essay scores (T. Han and Sari, 2022). The accuracy of combined automated teacher feedback was 45.2% higher than AWE feedback alone (J. Li, 2022). Besides, the combined feedback had more strengths in reducing grammar and mechanics errors (T. Han and Sari, 2022). If the teacher can use AWE feedback as a supplement judiciously and effectively, it can support teachers' work well and enhance learners' writing motivation and development (Woodworth and Barkaoui, 2020). Therefore, human feedback was suggested to be combined with AWE feedback for better-contributing students' writing improvement (Lang et al., 2019). While in this process, teachers should pay attention to the coherence and cohesiveness of three types of feedback (Mehrabi-Yazdi, 2018).

6. Conclusion

6.1. Major findings

Implementing artificial intelligence has played an important role in teaching and learning English as a foreign or second language. AWE is a highly discussed topic for students' writing skill development. In this study, researchers used bibliometric analysis to find the most important issues related to AWE. The authors also used VOSviewer to visualize the top ten authors, organizations, and countries and cited references, sources, and authors among the studies on AWE. Through the systematic review, researchers got 56 peer-reviewed literature. These results aided researchers in getting the most valuable and significant references about AWE.

Through the discussion and analysis, researchers could find the impacts of AWE feedback on users' writing development, factors influencing the effectiveness of AWE feedback, and problems when AWE is integrated into the classroom. AWE feedback is helpful for users to achieve their writing goals and improve the quality of their essays. However, its effectiveness differs in every aspect of writing, and it is not always more helpful than the teacher or peer feedback. Different forms of feedback and different platforms possess different levels of effectiveness. When integrated into the classroom, it can function as a good supplemental instrument, but it cannot replace the teacher or peer feedback role. Teachers' digital literacy and attitudes can heavily influence the use of AWE feedback in the classroom.

6.2. Limitations

This study has a bibliometric analysis and systematic review of the previous studies about AWE feedback in learning English as a second language. Despite the authors' efforts, there were still some limitations to this research. The results were concluded based on the previous content without an empirical study. Thus, a well-designed empirical study may need to be employed to certify the results. Besides, in the selection of the top ten cited authors, organizations, countries, references, and sources, authors ranked them mainly according to the total link strength, and the citations were just the second criterion to be considered. That was not the one and only standard. We could also obtain more influential information when setting citations or documents as the main criterion, but that would be limited by the article's length. Furthermore, artificial intelligence is making progress quickly, and the platform may also have suffered some changes in the research progress. Therefore, AWE feedback may also have improved. Future research could focus on the changes different platforms have made and test the improvement of AWE feedback's effectiveness.

6.3. Recommendations for educators

Based on the results and discussion, the authors put forward suggestions to educators about AWE's use in the education process. Firstly, educators should embrace the implementation of AWE inside and outside the classroom. Although AWE still has weaknesses in many aspects, it could assist teachers' teaching process and release much pressure for teachers. Good implementation requires teachers' digital literacy. Secondly, educators should explore integrating AWE and traditional writing teaching strategies to achieve better teaching effectiveness. Finally, teachers should clearly know the advantages and shortcomings of AWE use, thus taking advantage of the positive effects and avoiding the adverse effects. Therefore, teachers should pay close attention to students' AWE use situations to prevent students' too much dependence on AWE use and the effects of incorrect and inappropriate AWE feedback.

6.4. Implications for future research

The present study has provided a map of highly discussed topics and top-cited information. Through a bibliometric analysis of AWE, this research demonstrated the top ten cited authors, organizations, countries, references, and sources among the studies on AWE in learning English as a second language among the studies on AWE. These results could help more people interested in AWE quickly get valuable and significant information. Different types of AWE users could find useful information about the platforms they are using. Future research could focus on the specific improvement of a single platform, thus aiding the platform designers to perfect their platforms.

This study has analyzed the effectiveness and impact of AWE feedback from different dimensions. Users can clearly know AWE feedback's effectiveness and impact from this research. Therefore, they can adopt different strategies in the future use process. They can be more selective in using different feedback provided by AWE. For the platform designers, they could find the shortcomings of their present platform and have better perfection. Future research could focus on the reasons leading to the different effects of AEW's different types of feedback and try to find solutions. Thus, AWE feedback could develop positively for users.

Furthermore, researchers have found problems can be overcome when AWE feedback is integrated into the classroom. AWE feedback can be a very helpful tool if it can be used in the classroom. It can relieve teachers' workload greatly. However, using AWE feedback in the classroom is much more complicated than using it for a single person. Different approaches applied by teachers could exert different influences on the effectiveness of AWE feedback in the use process. Therefore, in future studies, researchers could focus on which approach teachers apply to facilitate the integration of AWE into the classroom the most.

Author contributions

Funding

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

References

Allen LK, Likens AD, McNamara DS (2019). Writing flexibility in argumentative essays: A multidimensional analysis. *Reading and Writing: An Interdisciplinary Journal* 32(6): 1607–1634. doi: 10.1007/s11145-018-9921-y

Bai L, Hu G (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology* 37(1): 67–81. doi: 10.1080/01443410.2016.1223275

Bridgeman B, Ramineni C (2017). Design and evaluation of automated writing evaluation models: Relationships with writing in naturalistic settings. *Assessing Writing* 34: 62–71. doi: 10.1016/j.asw.2017.10.001

Butterfuss R, Roscoe RD, Allen LK, et al. (2022). Strategy uptake in writing pal: Adaptive feedback and instruction. *Journal of Educational Computing Research* 60(3): 696–721. doi: 10.1177/07356331211045304

Chen CFE, Cheng WYE (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology* 12(2): 94–112.

Chen M, Cui Y (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing* 57: 100915. doi: 10.1016/j.jslw.2022.100915

Chen Z, Chen W, Jia J, Le H (2022). Exploring AWE-supported writing process: An activity theory perspective. *Language Learning & Technology* 26(2): 129–148. doi: 10125/73482

Diklil S, Bleyle S (2014). Automated Essay Scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing* 22: 1–17. doi: 10.1016/j.asw.2014.03.006

Du Y, Gao H (2022). Determinants affecting teachers' adoption of AI-based applications in EFL context: An analysis of analytic hierarchy process. *Education and Information Technologies* 27(7): 9357–9384. doi: 10.1007/s10639-022-11001-y

Fan N (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *Sage Open* 13(2). doi: 10.1177/21582440231181296

Feng HH, Chukharev-Hudilainen E (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology* 26(2): 58–77. doi: 10125/73479

Foster S (2019). What barriers do students perceive to engagement with automated immediate formative feedback. *Journal of Interactive Media in Education* 2019(1): 15. doi: 10.5334/jime.516

Fu QK, Zou D, Xie H, Cheng G (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*. doi: 10.1080/09588221.2022.2033787

Gao J (2021). Exploring the feedback quality of an automated writing evaluation system Pigai. *International Journal of Emerging Technologies in Learning (iJET)* 16(11): 322–330. doi: 10.3991/ijet.v16i11.19657

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

Grimes D, Warschauer M (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning, and Assessment* 8(6).

Guo Q, Feng R, Hua Y (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning* 35(9): 2312–2331. doi: 10.1080/09588221.2021.1879161

Han T, Sari E (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*. doi: 10.1080/09588221.2022.2067179

Han Y, Zhao S, Ng LL (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. doi: 10.3389/fpsyg.2021.752793

Huang X, Zou D, Cheng G, et al. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society* 26(1): 112–131. doi: 10.30191/ETS.202301_26(1).0009

Jiang L, Yu S, Wang C (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System* 93: 102302. doi: 10.1016/j.system.2020.102302

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

Lai YH (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology* 41(3): 432–454. doi: 10.1111/j.1467-8535.2009.00959.x

Lang F, Li S, Zhang S (2019). Research on reliability and validity of mobile networks-based automated writing evaluation. *International Journal of Mobile Computing and Multimedia Communications* 10(1): 18–31. doi: 10.4018/IJMCMC.2019010102

Liao HC (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process

approach. *System* 62: 77–92. doi: 10.1016/j.system.2016.02.007

Li J (2022). English writing feedback based on online automatic evaluation in the era of big data. *Mobile Information Systems* 2022: 9884273. doi: 10.1155/2022/9884273

Li J, Link S, Hegelheimer V (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing* 27: 1–18. doi: 10.1016/j.jslw.2014.10.004

Lim K, Song J, Park J (2022). Neural automated writing evaluation for Korean L2 writing. *Natural Language Engineering* 29(5): 1341–1363. doi: 10.1017/S1351324922000298

Link S, Mehrzad M, Rahimi M (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning* 35(4): 605–634. doi: 10.1080/09588221.2020.1743323

Li R (2021). Modeling the continuance intention to use automated writing evaluation among Chinese EFL learners. *Sage Open* 11(4). doi: 10.1177/21582440211060782

Liu S, Yu G (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology* 26(2): 78–105. doi: 10125/73480

Li Z, Feng HH, Saricaoglu A (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *Calico Journal* 34(3): 355–375. doi: 10.1558/cj.26382

Li Z (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System* 99: 102505. doi: 10.1016/j.system.2021.102505

Li Z, Link S, Ma H, et al. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System* 44: 66–78. doi: 10.1016/j.system.2014.02.007

Lu X (2019). An empirical study on the artificial intelligence writing evaluation system in China CET. *Big Data* 7(2): 121–129. doi: 10.1089/big.2018.0151

Mehrabi-Yazdi O (2018). Short communication on the missing dialogic aspect of an automated writing evaluation system in written feedback research. *Journal of Second Language Writing* 41: 92–97. doi: 10.1016/j.jslw.2018.05.004

Miranty D, Widiati 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. doi: 10.47750/pegegog.11.04.12

Mohsen MA (2022). Computer-mediated corrective feedback to improve L2 writing skills: A meta-analysis. *Journal of Educational Computing Research* 60(5): 1253–1276. doi: 10.1177/07356331211064066

Nazari N, Shabbir MS, Setiawan R (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon* 7(5): E07014. doi: 10.1016/j.heliyon.2021.e07014

Ngo TTN, Chen HHJ, Lai KKW (2022). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive Learning Environments*. doi: 10.1080/10494820.2022.2096642

Page MJ, McKenzie JE, Bossuyt PM, et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372: n71. doi: 10.1136/bmj.n71

Palermo C, Thomson MM (2018). Teacher implementation of Self-Regulated Strategy Development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology* 54: 255–270. doi: 10.1016/j.cedpsych.2018.07.002

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. doi: 10.17239/jowr-2020.12.01.04

Parra GL, Calero SX (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction* 12(2): 209–226. doi: 10.29333/iji.2019.12214a

Petchprasert A (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education* 6: 1. doi: 10.1186/s40862-020-00107-w

Potter A, Wilson J (2021). Statewide implementation of automated writing evaluation: Analyzing usage and associations with state test performance in grades 4–11. *Educational Technology Research and Development* 69(3): 1557–1578. doi: 10.1007/s11423-021-10004-9

Qian L, Yang Y, Zhao Y (2021). Syntactic complexity revisited: Sensitivity of China's AES-generated scores to syntactic measures, effects of discourse-mode and topic. *Reading and Writing* 34: 681–704. doi: 10.1007/s11145-020-10087-5

Ranalli J (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning* 31(7): 653–674. doi: 10.1080/09588221.2018.1428994

Ranalli J (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing* 52: 100816. doi: 10.1016/j.jslw.2021.100816

Reynolds BL, Kao CW, Huang Y (2021). Investigating the effects of perceived feedback source on second language writing performance: A quasi-experimental study. *Asia-Pacific Education Researcher* 30(6): 585–595. doi: 10.1007/s40299-021-00597-3

Roscoe RD, Wilson J, Johnson AC, Mayra CR (2017). Presentation, expectations, and experience: Sources of

student perceptions of automated writing evaluation. *Computers in Human Behavior* 70: 207–221. doi: 10.1016/j.chb.2016.12.076

Sari E, Han T (2022). Using generalizability theory to investigate the variability and reliability of EFL composition scores by human raters and e-rater. *Porta Linguarum: An International Journal of Foreign Language and Learning* 38: 27–45. doi: 10.30827/portalin.vi38.18056

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

Shamseer L, Moher D, Clarke M, et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *British Medical Journal* 349: g7647. doi: 10.1136/bmj.g7647

Shi H, Vahid A (2022). A systematic review of automated writing evaluation systems. *Education and Information Technologies* 28: 771–795. doi: 10.1007/s10639-022-11200-7

Stevenson M, Phakiti A (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing* 19: 51–65. doi: 10.1016/j.asw.2013.11.007

van Eck NJ, Waltman L (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* 111(2): 1053–1070. doi: 10.1007/s11192-017-2300-7

Wang Q (2022). The use of semantic similarity tools in automated content scoring of fact-based essays written by EFL learners. *Education and Information Technologies* 27(9): 13021–13049. doi: 10.1007/s10639-022-11179-1

Wang YJ, Shang HF, Briody P (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning* 26(3): 234–257. doi: 10.1080/09588221.2012.655300

Warschauer M, Grimes D (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal* 3(1): 22–36. doi: 10.1080/15544800701771580

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

Wilken JL (2018). Perceptions of L1 glossed feedback in automated writing evaluation: A case study. *Calico Journal* 35(1): 30–48. doi: 10.1558/cj.26383

Wilson J, Ahrendt C, Fudge EA, et al. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education* 168: 104208. doi: 10.1016/j.compedu.2021.104208

Wilson J, Czik A (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education* 100: 94–109. doi: 10.1016/j.compedu.2016.05.004

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

Xu J, Zhang S (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher* 31(4): 357–367. doi: 10.1007/s40299-021-00577-7

Yu Z (2020). Extending the learning technology acceptance model of WeChat by adding new psychological constructs. *Journal of Educational Computing Research* 58(6): 1121–1143. doi: 10.1177/0735633120923772

Yu Z (2022). A meta-analysis and bibliographic review of the effect of nine factors on online learning outcomes across the world. *Education and Information Technologies* 27(2): 2457–2482. doi: 10.1007/s10639-021-10720-y

Yu Z, Zhu Y, Yang Z, Chen W (2019). Student satisfaction, learning outcomes, and cognitive loads with a mobile learning platform. *Computer Assisted Language Learning* 32(4): 323–341. doi: 10.1080/09588221.2018.1517093

Zhai N, Ma X (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research* 61(4): 875–900. doi: 10.1177/07356331221127300

Zhang J, Zhang LJ (2022). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning*. doi: 10.1080/09588221.2022.2069822

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

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

Zhang Z, Hyland K (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing* 36: 90–102. doi: 10.1016/j.asw.2018.02.004

Appendix

Table A1. A list of reviewed studies.

N	Literature	Research object	Research focus	Method	N	Literature	Research object	Research focus	Method
1	(Allen et al., 2019)	An AWE system	Effectiveness of feedback	Experiment	29	(Lu, 2019)	Juku AWE	Perception	Experiment
2	(Bai and Hu, 2017)	Pigai	precision	Corpus analyses	30	(Mehrabi-Yazdi, 2018)	Criterion	Participation through dialogue	Communication
3	(Bridgeman and Ramineni, 2017)	Experimental model	Effectiveness of predicting scores	Corpus analyses	31	(Miranty and Widiati, 2021)	Grammarly	Perception	Experiment and questionnaire
4	(Butterfuss et al., 2022)	Writing Pal	Students' perception of effectiveness	Experiment	32	(Mohsen, 2022)	AWE feedback	Writing skills	Meta-analysis
5	(C. F. E. Chen and Cheng, 2008)	MY Access!	Improvement in writing	Observation	33	(Nazari et al., 2021)	AI-powered writing tools	Efficacy	Experiment
6	(M. Chen and Cui, 2022)	iWrite system	Cohesion and coherence	Experiment and questionnaire	34	(Palermo and Thomson, 2018)	NC Write	Integration with writing instruction	Experiment
7	(Z. Chen et al., 2022)		Incorporation	Interview	35	(Palermo and Wilson, 2020)	MI Write	Integration with writing instruction	Mixed-methods
8	(Diklil and Bleyle, 2014)	Criterion	Automated essay scoring	Experiment	36	(Parra and Calero, 2019)	Grammarly	Writing skills	Experiment
9	(Du and Gao, 2022)		Teachers' adoption	Corpus analyses	37	(Petchprasert, 2021)	Coh-Metrix	Writing performances	Corpus analyses
10	(Fan, 2023)	Grammarly	Writing quality	Quasi-experimental design and questionnaire	38	(Potter and Wilson, 2021)	An AWE system	Use situation	Questionnaire
11	(Feng and Chukharev-Hudilainen, 2022)	Genre-based AWE system	Development and evaluation	Experiment	39	(Qian et al., 2021)	Systems-pigai and iWrite	Syntactic complexity and scoring	Corpus analyses
12	(Foster, 2019)	OpenEssayist	Students' adoption	Questionnaire	40	(Ranalli, 2018)	An AWE system	Ability to use AWE	Experiment
13	(Gao, 2021)	Pigai	Quality of feedback	Experiment	41	(Ranalli, 2021)	Grammarly	Engagement	Experiment and interview
14	(Y. Han et al., 2021)	Grammarly	Error-correction effect	Experiment	42	(Reynolds et al., 2021)	An AWE system	Perceptions of the feedback source	Experiment
15	(Huang et al., 2021)		AI in education	Bibliometric analysis	43	(Roscoe et al., 2017)	An AWE system	Perceptions of the feedback source	Experiment
16	(Jiang et al., 2020)	An AWE system	Impact on teacher's feedback	Interview, Questionnaire and Observations	44	(Sari and Han, 2022)	An AWE system	Reliability of scoring	Experiment
17	(Koltovskaia, 2020)	An online writing system	Impact on self-correction	Interview, Questionnaire and Observations	45	(Saricaoglu and Bilki, 2021)	Criterion	Use situation	Corpus analyses
18	(Lai, 2010)	MY Access	Students' perception	Experiment	46	(Stevenson and Phakiti, 2014)	Six selected tools	Scoring of fact-based essays	Experiment

Table A1. (*Continued*).

N	Literature	Research object	Research focus	Method	N	Literature	Research object	Research focus	Method
19	(Lang et al., 2019)	Pigai	Reliability and validity	Experiment	47	(Y. J. Wang et al., 2013)	An AWE system	Writing performance	Experiment
20	(J. Li, 2022)	An AWE system	Combination of teacher and automatic feedback	Experiment	48	(Wilken, 2018)	An AWE system	L1 Glossed Feedback	Survey and interview
21	(J. Li, Link, and Hegelheimer, 2015)	Criterion (R)	Perception	Mixed-methods	49	(Wilson et al., 2021)	MI Write	Teachers' perceptions	Inductive coding
22	(R. Li, 2021)	An AWE system	Perception	Questionnaire	50	(Wilson and Czik, 2016)	PEG Writings	Impact on teacher's feedback	Experiment
23	(Z. Li, 2021)	Criterion (R)	Teachers' role	Experiment	51	(Woodworth and Barkaoui, 2020)	AWE systems	AWE in classroom	Questionnaire ans interview
24	(Z. Li, Feng, and Saricaoglu, 2017)	Criterion	Grammatical accuracy	Mixed-methods	52	(Xu and Zhang, 2022)	An AWE system	Learners with different proficiency	Experiment
25	(Z. Li, Link, Ma, et al., 2014)	Criterion	Scoring	Survey and interview	53	(Zhai and Ma, 2023)	AWE feedback	Writing quality	Meta-analysis
26	(Liao, 2016)	Criterion	Grammatical accuracy	Observation and interview	54	(Z. Zhang, 2017)	An AWE system	Engagement	Case study
27	(Link et al., 2022)	An AWE system	Impact on teacher's feedback	Experiment	55	(Z. Zhang, 2020)	An AWE system	Engagement	Corpus analyses
28	(Liu and Yu, 2022)	Write and improve with Cambridge	Engagement	Experiment	56	(Z. Zhang and Hyland, 2018)	An AWE system	Engagement	Case study