A Bibliometric Analysis of Automated Writing Evaluation in Education Using VOSviewer and CitNetExplorer from 2008 to 2022

Xinjie Deng, Beijing Language and Culture University, China*

iD https://orcid.org/0000-0002-8901-3080

ABSTRACT

As technology develops by leaps and bounds, automated writing evaluation (AWE) has caught increasing attention worldwide. This study aims to provide an overview of research literature focusing on AWE used in education through bibliometric analysis. The data of studies (N = 815) published from 2008 to 2022 were analyzed using the performance analysis and the science mapping analysis. VOSviewer and CitNetExplorer software conducted the mapping of science based on citation analysis, co-citation analysis, co-word analysis, and citation network analysis. The results indicated that the productive authors, institutions, and countries mainly came from North America, Asia, and Europe. The keywords in AWE studies ranged from technological terms to educational terms, with eight clusters. Three major groups in citation network analysis showed that different themes of AWE research were complementary to each other. This study provided references for entry into the AWE field and identifying future research directions.

KEYWORDS

Automated Writing Evaluation, Bibliometric Mapping Analysis, Bibliometric Review, Citation Analysis, Citation Network, CitNetExplorer, Educational Use, VOSviewer

INTRODUCTION

With the advancement of computer science, a variety of technologies are often integrated with education. Online automated writing evaluation (AWE) is one of the most popular topics in artificial intelligence-enhanced language learning (Huang et al., 2021). In the field of education, formative writing assessment plays a significant role in writing practice since it informs students of both achievement levels and specific weaknesses (Stevenson & Phakiti, 2014). Feedback, as an essential component of formative writing assessment, is usually provided by an agent, including teachers who could give corrective information (Hattie & Timperley, 2007). However, faced with many essays written by students, teachers may struggle to provide immediate feedback within a short time. In this case, AWE could serve as an assistant tool to lessen teachers' workload, contributing to the improvement of learners' writing performance (Parra & Calero, 2019).

Automated writing evaluation (AWE) is a program or software that provides immediate computer-generated feedback and scoring on written texts (Shermis et al., 2013; Wilson, Ahrendt, et al., 2021). The core element of AWE systems is a scoring engine supported by technologies such as

natural language processing and machine learning algorithms (Wilson & Roscoe, 2020). The natural language processing is responsible for linguistic, syntactic, semantic, and discourse features, while statistical algorithms are associated with generating holistic scores. Another central component of AWE technology is a feedback engine that provides detailed feedback to help learners revise their writing (Allen et al., 2016). Currently, widely used AWE platforms are Criterion, Write&Improve, My Access!, and WriteToLearn (Hockly, 2019).

There are various benefits of AWE implementation. The immediate feedback helped students develop their language and show confidence in submitting their essays (O'Neill & Russell, 2019). As scores would grow if students could revise their work based on the feedback, the iterative revision processes gave students opportunities to notice their progress, which promoted students' writing motivation (Wilson, Ahrendt, et al., 2021). In addition to psychological traits, AWE also exerted positive effects on writing-related outcomes. Students using AWE systems significantly improved their writing accuracy mainly because they noticed suggestions, explanations, and color-coded lines (Barrot, 2021). Moreover, the automated feedback was as effective as comments made by human teachers when the comments were pertinent to structure, organization, conclusion, coherence, and supporting ideas (Liu et al., 2017).

Nevertheless, disadvantages of AWE lie in formulaic writing, overcorrection, and perceived negative emotions. Scores induced students to attach importance to formulaic writing that values quantity and complexity (Perelman, 2014). Computer-generated comments were criticized to mislead students about the nature of writing. Students tended to meet the standards of AWE systems by developing test-taking strategies or tricks, for example, increasing the number of words (Wilson, Ahrendt, et al., 2021). Occasionally, overcorrection may discourage and frustrate students since the program still suggested revisions even if there were no errors (Barrot, 2021). More importantly, when receiving automatic feedback, students experienced anxiety, pressure, and control, influencing their identity representations (Zaini, 2018).

Many researchers have investigated the use of AWE with different research methods. Wilson, Ahrendt, et al. (2021) adopted activity theory to qualitatively analyze elementary teachers' perceptions of AWE programs, students' writing motivation, and instructional challenges of AWE. Another study used a t-test to confirm the positive influence of AWE tools on undergraduate students' writing performance (Parra & Calero, 2019). It was also found that AWE tools provided feedback in terms of grammar, punctuation, style, and mechanics.

There also emerge some review studies on AWE research in education. Stevenson and Phakiti (2014) provided a critical review of 36 studies on the effect of AWE by coding research survey. Hibert (2019) selected 29 studies from 2007 to 2018 and explored theoretical foundations and methodological approaches of AWE use among university students. Nevertheless, both articles adopted content analysis and failed to give a whole picture of the AWE field. Likewise, another review of literature from 2000 to 2020 discussed AWE use in K12 education (Nunes et al., 2022). It followed PRISMA guidelines and identified only 8 studies to examine the impact of six AWE systems on text quality and learning outcomes. A more recent review provided a comprehensive analysis of 48 studies retrieved from SSCI journals (Fu et al., 2022). The analysis included methodology, types of learners, feedback and applications, learning outcomes, and implications. However, the most obvious weakness was the limited scope that all the finalized studies were SSCI articles. Table 1 summarizes previous review studies in terms of total number of included studies, research methods, and research foci.

Table 1. The comparison of review studies

No.	Authors	Number of included studies	Research method	Foci
1	Nunes et al. (2022)	8	PRISMA and coding method	Effectiveness and users' perceptions of AWE systems
2	Hibert (2019)	29	Coding method	Theoretical foundations and methodological approaches of AWE research
3	Stevenson & Phakiti (2014)	36	Coding method	Methodological features and the effect of AWE on writing quality
4	Fu et al. (2022)	48	Coding method	Methodology, types of learners, feedback and applications, learning outcomes, and implications
5	This study	815	PRISMA and bibliometric analysis	Productive authors, institutions, and countries; the structural networks among research constituents; and citation networks in AWE studies

While previous studies reviewed AWE research to some extent, few of them provided a general overview of the specific field. Therefore, none of them adopted the method of bibliometric analysis, let alone simultaneously using different analyzing software. The mapping technique could provide graphical representations of the relations between key terms (Heersmink et al., 2011; Yu, 2021). It could also reveal the development of research, indicating possible trends in a certain field (Vogel & Masal, 2015). Through mapping, the distribution of research information and the relations of concepts become obvious and clear (Yilmaz et al., 2019). Researchers used techniques to structure previous studies due to their objectivity (e.g., Yu et al., 2021; Yu et al., 2022). Yu (2020a) conducted a bibliometric study on the use of artificial intelligence in education through VOSviewer and CiteSpace. The study visualized previous literature through keywords, countries, co-citations, bursts, citation counts, betweenness centrality, and sigma.

It is suggested that VOSviewer and CitNetExplorer could be used together to cluster publications and analyze the resulting clusters (van Eck & Waltman, 2017). VOSviewer is a freely available tool for constructing and viewing the bibliometric data based on maps, while CitNetExplorer could provide timeline-based citation networks (van Eck & Waltman, 2014). Fellnhofer (2019) used both software tools to systematically cluster the research literature on entrepreneurship education. Therefore, given the importance of AWE and scant bibliometric literature reviews on AWE research, this study aims to unpack the prolific research constituents, reveal the bibliometric structures among research constituents, and explore the development of AWE research in education, paving the way for future research. The author thus proposed the following research questions:

RQ1. What are the structural networks among countries publishing AWE studies?
RQ2. What are the structural networks among institutions publishing AWE studies?
RQ3. What are the structural networks among authors publishing AWE studies?
RQ4. What is the distribution of the most used keywords in AWE research?
RQ5. What are the citation networks in AWE research?

The rest of this article is structured as follows. The Methods section explains the bibliometric method and the procedures this study followed. In Results and Discussion section, the article presents and analyzes the results of bibliometric analysis, including performance analysis and science mapping analysis. The last section shows the major contributions, implications, and limitations of this study.

METHODS

The author followed the general guidelines proposed by Donthu et al. (2021) and conducted a bibliometric analysis study. The guidelines or procedures generally have four steps, i.e., defining research aims and scope, choosing bibliometric analysis techniques, collecting the bibliometric data, and running the analysis and reporting findings.

Defining The Research Aims And Scope

The aims of a bibliometric review should focus on the "retrospection of the performance" and "science of a research field" (Donthu et al., 2021, p. 291). Thus, in terms of performance, this study aims to reveal prolific research constituents in the AWE field, including authors, institutions, and countries. In terms of science, this study attempts to uncover the intellectual structures and relationships among research constituents. It is also designed to explore the development and general themes of AWE research. The scope of a bibliometric study, as the other aspect of the first step, should be adequately large. Specifically, the number of publications in a certain field is expected to exceed 500 (Donthu et al., 2021).

### Choosing Bibliometric Analysis Techniques

The author selected metrics such as total publications, total citations, average citations, and h-index to carry out the performance analysis. In the science mapping analysis, the author used VOSviewer version 1.6.17 and adopted techniques including citation analysis, co-citation analysis, and co-word analysis. Furthermore, in order to have a comprehensive understanding of AWE research, the author also performed network analysis and chose enrichment techniques. CitNetExplorer version 1.0.0 was employed to create a citation network visualization, which could reveal the evolution and development of a research field (van Eck & Waltman, 2014).

Collecting The Bibliometric Data

The author used Web of Science (WoS) Core Collection as the database for this review study. One reason is that WoS is a famous trusted multidisciplinary database of academic research, including high-quality and peer-reviewed journals around the globe (Yu, 2020a). WoS Core Collection includes databases of SCIE, SSCI, A&HCI, CPCI, ESCI, CCRE, and IC. Recent years have seen many bibliometric review studies using WoS articles (e.g., Zhang et al., 2022; Yu, 2020b). Another reason for choosing WoS is that data retrieved from this online database can be conveniently operated in most bibliometric analysis software tools, such as VOSviewer, CitNetExplorer, and CiteSpace.

The author applied a Boolean search on online databases. Informed by the search strings used in previous review studies and considering the objectives of this study, the author searched digital databases by keying in "automat*" OR "computer-generated" (topic) and "assess*" OR "evaluat*" OR "feedback" OR "scor*" (topic) and "writ*" OR "essay*" (topic) and "learn*" OR "teach*" OR "educat*" (topic) from January 2008 to 26 March 2022 due to the availability of library resources. The initial search resulted in 1779 documents.

Figure 1. A flow diagram of the study selection based on PRISMA-P
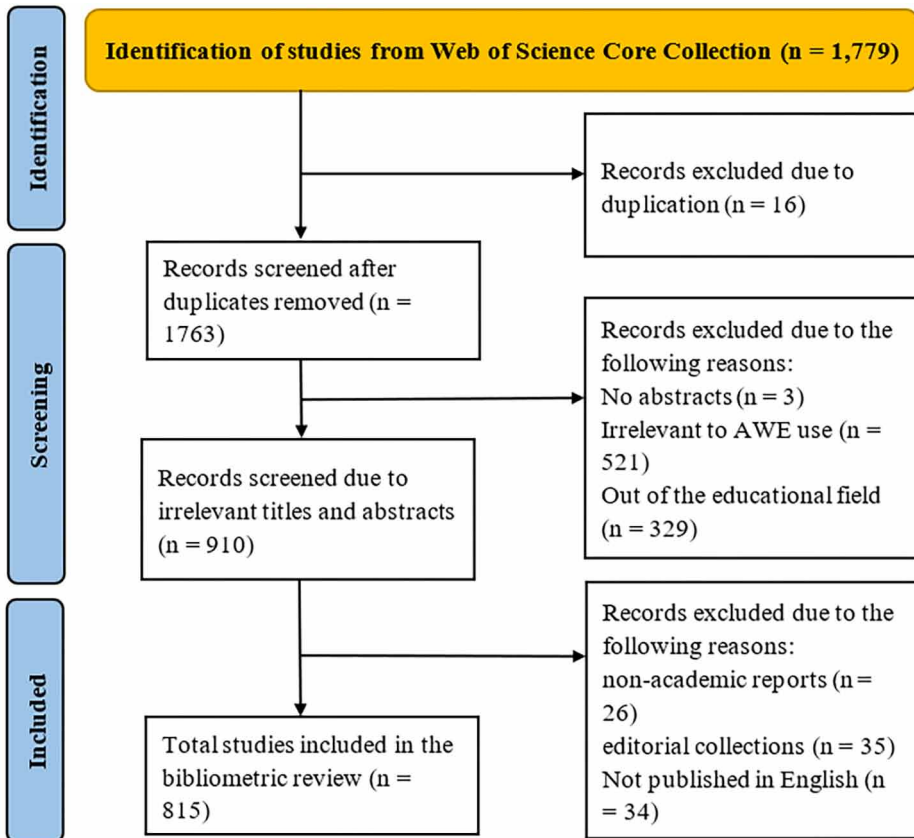
To ensure the relatedness and quality of included studies, two researchers examined the results. Based on the Preferred Reporting Items for Systematic Review and Meta-analysis Protocol (PRISMA-P) (Moher et al., 2015), they included studies if the studies (1) focused on the use of AWE in educational field, (2) were published after stringent peer-reviewed process, (3) were rigidly designed, and (4) were published in English. They excluded the studies if they (1) were irrelevant to the use of AWE, (2) were out of the educational scope, (3) were editorial collections, retracted articles, and non-academic reports, and (4) had no abstracts, and (5) were not published in English. Finally, the selection process yielded 815 results, including 443 articles, 364 proceeding papers, and 14 reviews. Figure 1 illustrates the process of study screening and selection. The author downloaded full record contents of the results (N = 815) in the form of plain texts.

RESULTS AND DISCUSSION

Figure 2. Number of AWE studies per year
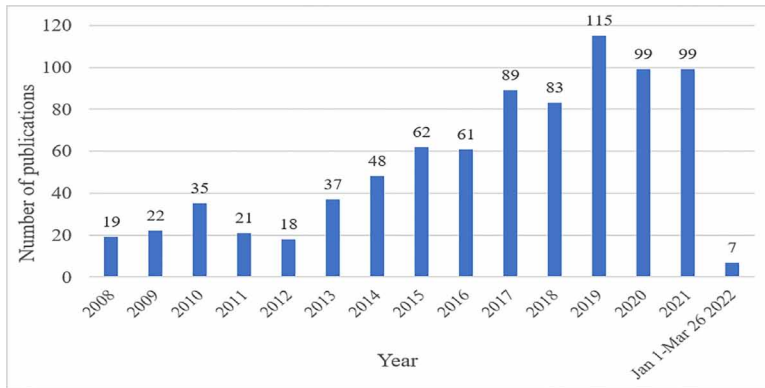
Performance analysis of AWE studies

Figure 2 shows the number of publications on AWE per year since 2008. In total, 815 documents were included in this bibliometric review. It can be observed that the number of articles remained at a low level until 2013 when AWE studies started to increase, reaching a peak in 2019. The year 2019 was thus determined an important year for AWE studies, with more than 100 publications. Moreover, the relevant studies in 2022 were not thoroughly revealed since the data was obtained on March 26, 2022.

Table 2 reveals that the most productive author in AWE studies is Wilson, J., with 14 articles, followed by Crossley, S. A. and McNamara, D. S., with 13 articles each. Wilson, as an associate professor at University of Delaware, devotes himself to the teaching and learning of language writing with AWE systems. Crossley who is a professor at Georgia State University and McNamara, a professor at Arizona State University, are known for the research on enterprise computing and natural language processing tools. Interestingly, Crossly and McNamara as co-authors have published 6 articles together. They are also the most cited authors in AWE research.

Table 2. The most productive authors in AWE studies

N.	Authors	Organization	Number of publications	Total citations	Average citations	h-index
1	Wilson, J	University of Delaware	14	189	13.50	8
2	Crossley, S. A	Georgia State University	13	619	47.62	6
3	McNamara, D. S	Arizona State University	13	501	38.54	7
4	Dascalu, M	Polytechnic University of Bucharest	8	18	2.25	3
5	Trausan-matu, S	Polytechnic University of Bucharest	8	27	3.38	3

Table 2 continued

N.	Authors	Organization	Number of publications	Total citations	Average citations	h-index
6	Liu, M	Hebei University	7	55	7.86	4
7	Calvo, R. A	Imperial College London	7	175	25.00	6
8	Kyle, K	University of Oregon	6	248	41.33	5
9	Shum, S. B	University of Technology Sydney	6	78	13.00	5
10	Mulholland, M	Educational Testing Service	6	46	7.67	3

Table 3 indicates institutions with the most publications and related indicators. The most productive institutions are Educational Testing Service, with 26 publications and Georgia State University, with 20 articles. Both are also the most cited institutions, with 409 citations and 568 citations respectively. It is believed that these results benefit from employees in the institutions and the networks they generate (Gaviria-Marin et al., 2018). The publications from Wilson, Crossley, and McNamara contributed to their universities ranking the forefront of the AWE field. In terms of h-index, Educational Testing Service (https://www.ets.org/), the largest organization for educational testing and assessment in the world, occupies the first position.

Table 3. The most productive institutions in AWE studies

N.	Institutions	Country	Number of publications	Total citations	Average citations	h-index
1	Educational Testing Service	USA	26	409	15.73	11
2	Georgia State University	USA	20	568	28.4	9
3	Iowa State University	USA	18	294	16.33	10
4	University of Delaware	USA	14	189	13.50	8
5	Arizona State University	USA	14	275	19.64	7
6	University of Alberta	Canada	11	40	3.64	3
7	University of Technology Sydney	Australia	9	81	9.00	5
8	University of Sydney	Australia	8	252	31.50	7
9	Michigan State University	USA	8	24	3.00	4
10	Education University of Hong Kong	China	8	63	7.88	4

Table 4 presents the most productive countries and relevant metrics. The findings suggested that productive countries were mainly in North America, Asia, Europe, and Oceania. The USA stands first among the most productive countries, with 241 publications, accounting for 29.6% of the included AWE studies. It may be attributed to several productive institutions in USA. Following USA, China ranks the second with 149 publications. In addition, it should be noted that although England and Australia did not publish many articles, they still exerted great influence on AWE research with the *h*-index of 12 and 13 respectively.

Table 4. The most productive countries in AWE studies

| N.	Countries	Number of publications	Total citations	Average citations	*h*-index
1	USA	241	3574	14.83	32
2	China	149	550	3.69	12
3	England	44	592	13.45	12
4	Australia	43	493	11.47	13
5	Japan	40	65	1.63	4
6	Canada	37	220	5.95	7
7	Spain	31	199	6.42	9
8	India	30	58	1.93	5
9	Germany	28	173	6.18	8
10	France	18	179	9.94	6

Science Mapping Analysis of AWE Studies

This section performs a science mapping analysis with bibliographic software. The analysis tries to delve into the interactions and connections among scientific actors of the AWE field. Therefore, this study used VOSviewer and CitNetExplorer software to visualize the bibliometric material.

The author analyzed the highly cited countries through VOSviewer by selecting citation as the analysis type and countries as the analysis unit. The minimum number of documents of a country was set at 5, and minimum number of citations of a country was adjusted at 1. Of the 68 countries, 34 met the thresholds (see Figure 3). It shows that these countries are countries that researchers focus on when they conduct AWE studies. They prefer publications from these countries to further improve their research works.

Figure 3. Cluster mapping based on countries
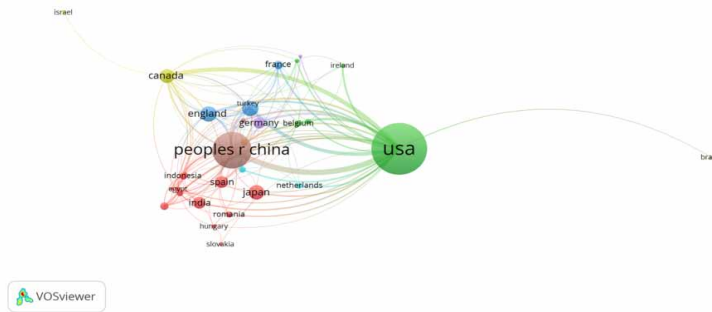
The top 10 countries with the greatest total link strength are as follows: USA (Citations = 3574, Documents = 241, Total link strength = 578), The People's Republic of China (Citations = 550, Documents = 149, Total link strength = 333), Canada (Citations = 220, Documents = 37, Total link strength = 149), Australia (Citations = 493, Documents = 43, Total link strength = 148), England (Citations = 592, Documents = 44, Total link strength = 70), Germany (Citations = 173, Documents = 28, Total link strength = 44), Turkey (Citations = 58, Documents = 6, Total link strength = 44), Singapore (Citations = 91, Documents = 9, Total link strength = 43), Belgium (Citations = 149, Documents = 12, Total link strength = 42), and Malaysia (Citations = 26, Documents = 13, Total link strength = 38).

The author also analyzed the frequently cited organizations by selecting citation and organizations as the analysis type and the unit of analysis respectively. The minimum number of documents of an organization was set at 5. Of the 789 organizations, 42 met the threshold (see Figure 4). It is indicated that these organizations have the highest total link strength of co-citation links with others. Researchers would like to pay additional attention to the studies conducted by these organizations or the authors affiliated with these institutions when they conduct relevant studies.

**Figure 4. Cluster mapping based on organizations**
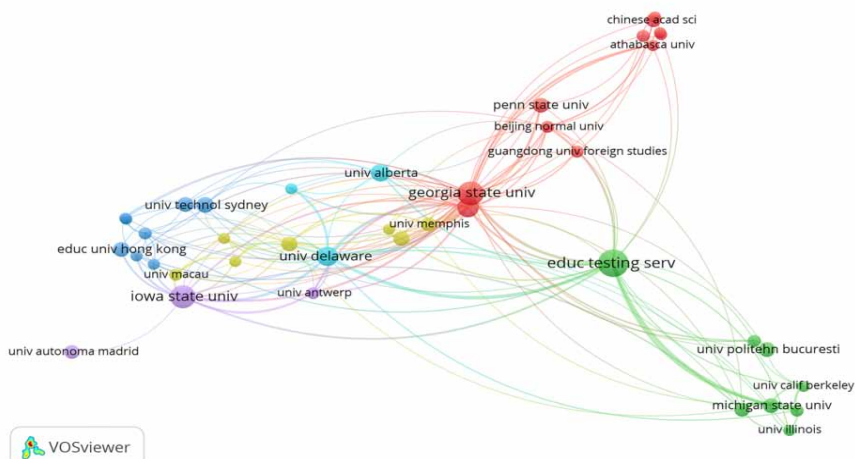
The top 10 organizations with the greatest total link strength were Georgia State University (Citations = 568, Documents = 20, Total link strength = 108), University of Delaware (Citations = 189, Documents = 14, Total link strength = 92), Educational Testing Service (Citations = 409, Documents = 26, Total link strength = 84), Arizona State University (Citations = 275, Documents = 14, Total link strength = 83), Iowa State University (Citations = 294, Documents = 18, Total link strength = 65), University of Georgia (Citations = 25, Documents = 6, Total link strength = 37), University of Sydney (Citations = 248, Documents = 8, Total link strength = 35), Michigan State University (Citations = 21, Documents = 8, Total link strength = 29), Carnegie Mellon University (Citations = 79, Documents = 7, Total link strength = 26), and Beijing Normal University (Citations = 37, Documents = 6, Total link strength = 23).

The findings showed that USA was the most cited country and that Georgia State University was the most cited organization in AWE studies. One possible reason is that highly cited countries and organizations were productive. Previous review articles (Stevenson & Phakiti, 2014; Nunes et al., 2022) also evidenced that USA was the most productive country since more than half of their included studies came from USA. Thus, it was not surprising that the majority of the top ten most cited institutions were situated in USA. Another reason lies in the content of publications. As one of the leaders in AWE research, Georgia State University introduced either new approaches to essay scoring or tools for automatic analysis, contributing to high citations.

The author obtained the top 10 co-cited authors by selecting co-citation as the analysis type and cited authors as the unit of analysis. The minimum number of citations of an author was set at 20. Of the 12575 authors, 123 met the threshold (see Figure 5). The node size and the links to other authors showed the author's contribution and importance in academia (Brika et al., 2021). Since these authors were co-cited frequently, they were among the most influential ones. Other researchers mostly relied on their ideas.

Figure 5. Cluster mapping based on authors
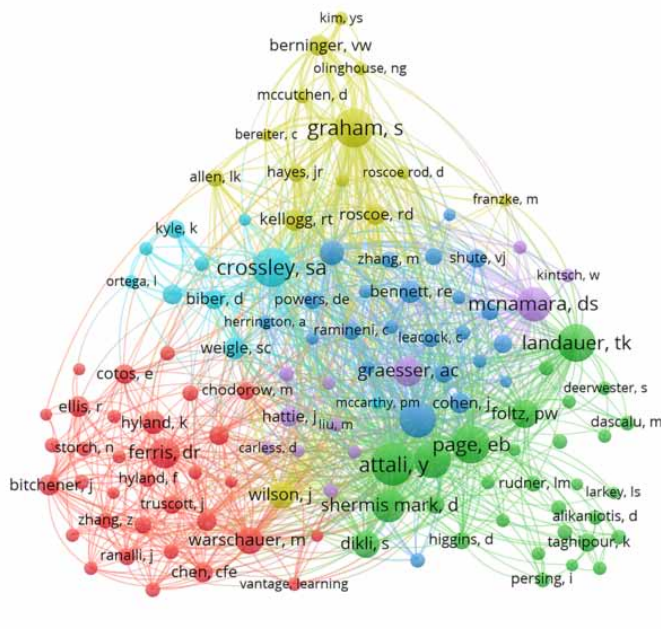
The top 10 co-cited authors with the greatest total link strength were Graham, S. (Citations = 193, Total link strength = 4259), Crossley, S. A. (Citations = 195, Total link strength = 3332), Attali, Y. (Citations = 229, Total link strength = 3139), Shermis, M. D. (Citations = 173, Total link strength = 3133), Page, E. B. (Citations = 182, Total link strength = 2611), Burstein, J. (Citations = 175, Total link strength = 2442), McNamara, D. S. (Citations = 152, Total link strength = 2403), Wilson, J. (Citations = 97, Total link strength = 2324), Landauer, T. K. (Citations = 185, Total link strength = 2132), and Warschauer, M. (Citations = 97, Total link strength = 1694).

Steve Graham is Professor at Arizona State University, the fourth most cited organization in AWE research, specializing in research on essay writing, teacher education, and special education. It was noticed that several highly cited meta-analyses made Graham the most co-cited author in AWE studies. Scott Andrew Crossley, the second most productive and the most cited author in AWE research, is also the second most co-cited author. Yigal Attali is a group leader at Educational Testing Service and an expert in educational measurement, automated scoring, and psychometrics.

VOSviewer software carried out the keyword co-occurrence analysis. The author created a map by selecting all keywords as the unit of analysis and full counting as the counting method. The minimum number of occurrences of a keyword was set at 4. Of the 2496 keywords, 214 met the threshold. As visually presented in Figure 6, these keywords were classified into eight clusters with different colors. The larger the circle in the network was, the more frequently a term occurred in all documents (van Eck & Waltman, 2021).

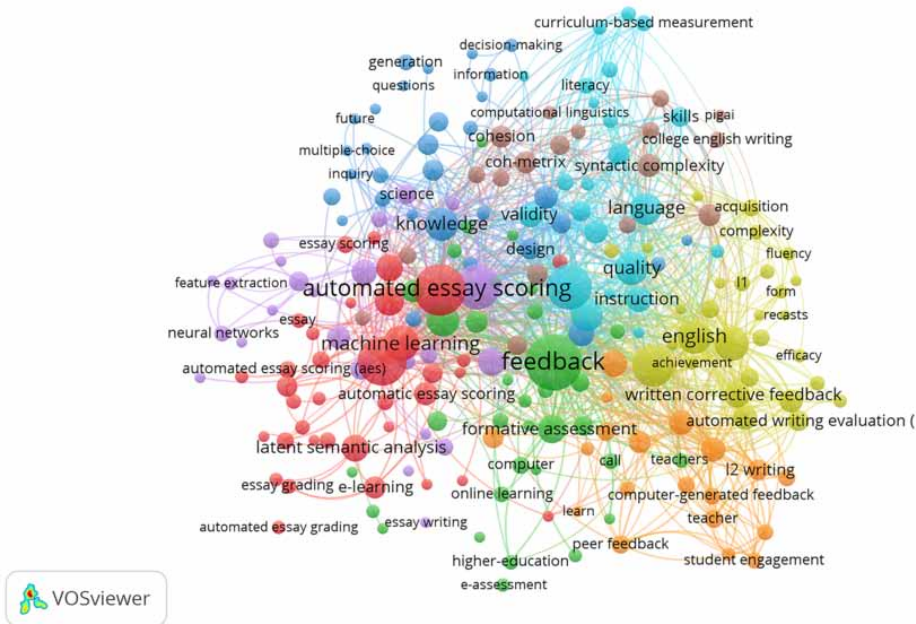**Figure 6. Cluster mapping based on keywords**

Table 5 explains the detailed information about these clusters. The clusters were ranked according to size, descending from the biggest one to the smallest one. The frequently used themes include feedback (Cluster 2), automated essay scoring (Cluster 1), natural language processing (Cluster 1),

students (Cluster 6), writing (Cluster 5), automated writing evaluation (Cluster 4), machine learning (Cluster 1), English (Cluster 4), accuracy (Cluster 4), and quality (Cluster 6), as well as other relevant terms.

Table 5. Information about Keywords of Six Clusters

Cluster	Color	Number	Main items	Percentage
1	Red	36	Automated essay scoring, natural language processing, machine learning, latent semantic analysis, text	16.8%
2	Green	34	Feedback, assessment, evaluation, higher education, model	15.9%
3	Blue	31	Agreement, academic writing, argumentation	14.5%
4	Yellow	29	Automated writing evaluation, accuracy, English, written corrective feedback	13.6%
5	Violet	24	Science, automated feedback	11.2%
6	Cyan	23	Performance, validity	10.7%
7	Orange	19	Teacher feedback, peer feedback, perceptions	8.9%
8	Brown	18	Linguistic features, syntactic complexity, cohesion	8.4%

Cluster 1 includes 36 items, e.g., "automated essay scoring" (N = 75, link strength = 187) "natural language processing" (N = 68, link strength = 195). These keywords indicated that there were a number of studies covering AWE technologies. Moreover, the term "machine learning" (N = 36, link strength = 119) in this cluster is also a most-used keyword. As one of the approaches to writing assessment, the machine-learning approach could not only automatically recognize and classify the written texts (Liu et al., 2017; Wulff et al., 2021), but also evaluate students' essays (Beggrow et al., 2014).

Featured by "feedback" (N = 87, link strength = 361), cluster 2 consists of 34 items. Other keywords included "assessment", "evaluation", "higher education", and "model". It was found that researchers preferred to use the word "feedback" rather than "assessment" (N = 30, link strength = 107) and "evaluation" (N = 11, link strength = 18). The term "feedback" had a dense link with "higher education", suggesting that research on AWE used in education mostly focused on higher education. A recent example of such a link could be found in the study by Zhang and Hyland (2022), examining the effect of AWE integration on student engagement in Chinese tertiary contexts. Furthermore, many studies investigated the potential and validity of deep learning in automated essay scoring, thus proposing various deep learning models to improve the accuracy of scoring (e.g., Yuan et al., 2020; Kumar & Boulanger, 2021).

Cluster 3 is composed of 31 terms, located at the top left of the map. The term "agreement" (N = 13, link strength = 32) in many AWE studies referred to the agreement between either two human raters or human and machine. "Academic writing" (N = 11, link strength = 28) emerging as a keyword revealed that previous studies mainly explored AWE implementation in academic writing (e.g., Zhang & Hyland, 2022). There was a connection between "academic writing" and "argumentation" (N = 7, link strength = 42), showing that argumentative writing received the most attention among other genres. For example, Zhu et al. (2017) confirmed the positive influence of AWE on students' scores in scientific argumentation.

Cluster 4 comprises 29 items represented by "automated writing evaluation" (N = 48, link strength = 201). This term had a dense link with "accuracy" (N = 32, link strength = 166), which suggested

research examining either the accuracy of AWE or the impact of AWE on writing accuracy (Barrot, 2021). "English" (N = 33, link strength = 186) and "English writing" (N = 9, link strength = 15) were another two keywords closely associated with "automated writing evaluation", indicating most studies concentrated on English writing (e.g., Woodworth & Barkaoui, 2020). In addition, the term "written corrective feedback" (N = 19, link strength = 76) had connections with all of the aforementioned terms in cluster 4 since it was one of the most important parts provided by AWE systems.

Co-occurring keywords in the remaining four clusters are intertwined with each other. The keyword "science" (N = 11, link strength = 58) in cluster 5 showed the application of AWE in the science curriculum, for example, teaching climate change (Zhu et al., 2020). The frequent occurrence of "automated feedback" (N = 22, link strength = 116) and its connection to terms in cluster 7, i.e., "teacher feedback" (N = 17, link strength = 111) and "peer feedback" (N = 10, link strength = 51), revealed that researchers have integrated these three types of feedback into writing (Tian & Zhou, 2020; Zhang & Hyland, 2022). However, research on the contrasts between automated feedback, teacher feedback, and peer feedback was still relatively limited.

"Performance" (N = 28, link strength = 146) in cluster 6 had wide connections with terms in other clusters. To explore the "validity" (N = 17, link strength = 66) of AWE, some researchers assessed students' writing performance in terms of fluency and accuracy (Shang, 2022). Additionally, the links between "performance", "motivation" (N = 10, link strength = 44), and "perceptions" (N = 17, link strength = 102) were worth noting. Some studies examined the impact of AWE on students' psychological states, i.e., motivation (Sherafati et al., 2020) and engagement (Zhang & Hyland, 2022), while other studies focused on participants' perceptions of AWE tools (e.g., Miranty & Widiati, 2021). Cluster 8 illustrated researchers frequently used "linguistic features" (N = 11, link strength = 50), including lexical sophistication and diversity (Goh et al., 2020), syntactic complexity (Lee et al., 2021), and cohesion (Goh et al., 2020) to measure writing quality.

The results indicated that many studies discussed AWE techniques and methods such as machine learning and natural language processing to improve the validity of AWE systems mainly because of their educational value. The study also suggested that AWE could be used in science education, which was not identified by previous reviews as they did not develop the coding scheme for disciplines. In addition, the keywords revealed previous contrasts among different types of feedback, which supported the previous review by Fu et al. (2022).

The frequently occurring keywords revealed the focus of research literature. It was found that AWE research primarily targeted higher education and English writing. This was consistent with the results done by Stevenson and Phakiti (2014) but different from the findings of Nunes et al. (2022). It might be due to the reason that Nunes et al. excluded the studies if they did not focus on children. The co-occurrence analysis also pointed out academic writing, particularly argumentative writing, was the main target task. It was in line with the research of Fu et al. (2022). They concluded that researchers most frequently investigated persuasive essays among six genres of writing texts. The possible explanation was that argumentative features correlated with analytic rubrics on which AWE technology in most cases was based (Davies et al., 2021).

Using the clustering function in CitNetExplorer software, the author obtained the citation network. Non-matching cited references were included, and the minimum number of citations was set at 10. Thus, the program identified altogether 972 publications with 4133 citation links during the period from 1960 to 2022. Clustering analysis resulted in six main groups, visualizing the 100 most cited publications on automated writing evaluation. Due to the minimum size requirement, 199 publications did not belong to a group. Table 6 shows citation network information for the six main groups, ranked by group size from the biggest to the smallest. The first group is the largest one

with 349 publications, 1561 citation links, and 55 studies in 100 most cited publications. The second and the third group consist of 234 publications and 1263 citation links, and 91 publications and 256 citation links respectively. Each of the other remaining three groups accounted for less than 8% of total number of publications and less than 3% of all citation links.

Table 6. Information about citation networks of six groups

Group	Color	Number of publications	Number of citation links	Citation score Median (Range)	Number of publications with citation score ³ 10	Number of publications in 100 most cited publications
1	Blue	349	1561	1 (0-73)	85	55
2	Green	234	1263	1 (0-53)	68	32
3	Purple	91	256	1 (0-32)	24	9
4	Orange	70	118	0 (0-44)	10	4
5	Yellow	16	17	0 (0-14)	1	0
6	Brown	13	14	0 (0-11)	1	0

Figure 7 presents a timeline-based citation network including the top 100 cited references. The longitudinal axis shows the year when the document was published, while the distance between two articles on the horizontal axis represents the extent of their relatedness (van Eck & Waltman, 2014). It was found that the first three groups (blue, green, and purple) included most of the publications and citation links. Considering these findings and the small number of publications in Groups 4, 5, and 6, the last three groups would not be further analyzed. The first three groups, rooted in three classical publications, were hereafter referred to as the Cohen-group, the Vygotsky-group, and the Halliday-group.

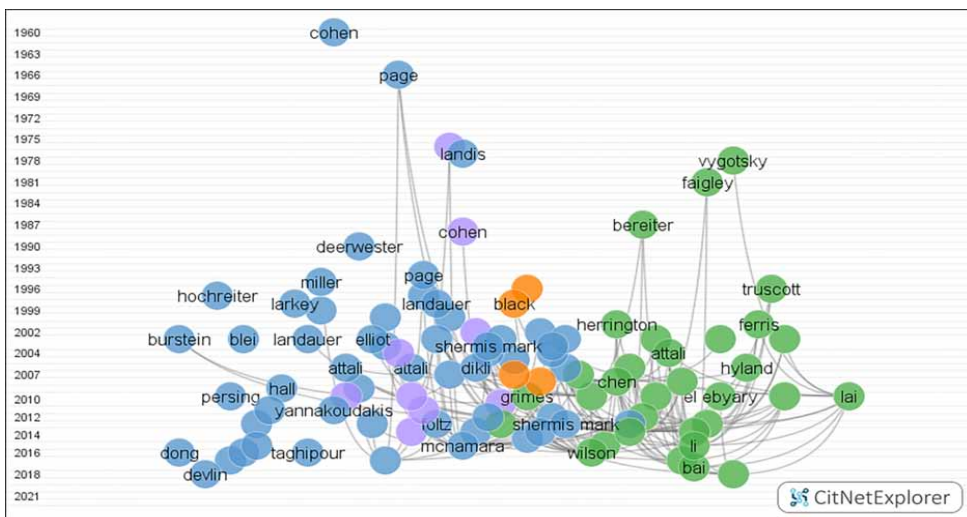**Figure 7. Citation networks about AWE studies**

Table 7 summarizes the representative works of three biggest groups. The Cohen-group started to publish articles in 1960, whereas the Vygotsky-group and the Halliday-group pioneered their research in 1978 and 1976 respectively. Thus, these three groups did not interact with each other in the first 40 years. The article written by Chen (2008) marked the beginning of the interaction between the Cohen-group and the Vygotsky-group. Since then, both two groups moved closer to each other, indicating more communication with each other. Furthermore, the Halliday-group started to cite publications in the Cohen-group from 2010 onwards, leading to no clear boundaries among these three groups.

Table 7. Representative publications of three biggest groups

Group		Publications		
		Pioneers	Most cited	Most recent
	Citation score	19	73	0
	Year	1960	2003	2022
	First author	Cohen	Shermis	Saha
Cohen-group	Title	A Coefficient of Agreement for Nominal Scales	Automated Essay Scoring: A Cross-Disciplinary Perspective	Development of a Practical System for Computerized Evaluation of Descriptive Answers of Middle School Level Students
	Topic	The design and development of AWE technology		
	Citation score	18	53	2
	Year	1978	2008	2021
	First author	Vygotsky	Chen	Wilson
Vygotsky-group	Title	Mind in Society: The Development of Higher Psychological Processes	Beyond the Design of Automated Writing Evaluation: Pedagogical Practices and Perceived Learning Effectiveness in EFL Writing Classes	Automated Feedback and Automated Scoring in the Elementary Grades: Usage, Attitudes, and Associations with Writing Outcomes in a Districtwide Implementation of MI Write
	Topic	The use of AWE in education		
	Citation score	17	32	0
	Year	1976	2004	2021
	First author	Halliday	Graesser	Mahadini
Halliday-group	Title	Cohesion in English	Coh-Metrix: Analysis of Text on Cohesion and Language	Using Conventional Rubric and Coh-Metrix to Assess EFL Students' Essays
	Topic	The effect of AWE on linguistic features		

The publications in the Cohen-group mainly focused on the design and development of AWE technology such as automated essay scoring systems. The pioneering publication was an article titled "A Coefficient of Agreement for Nominal Scales", published in *Educational and Psychological Measurement* in 1960 (Cohen, 1960). This article presented the concept of Cohen's kappa to measure the degree of agreement between two judges or scorers. Since then, this metric was often used to evaluate human-machine agreement in the field of AWE studies. The most highly cited publication in Group 1 was a book published in 2003 (Shermis & Burstein, 2003). The authors contributed to providing an overview of the development of AWE technology in many disciplines, including cognitive science, education, testing and measurement, language, and computer science. They also analyzed specific AWE systems and psychometric issues regarding reliability, validity, norming and scaling, and Bayesian analysis. The most recent publication that has been cited by other studies came from Saha and Rao (2022). This article proposed a grading system to evaluate long or descriptive answers based on reference answers written by experts. The experimental results confirmed the high accuracy of the system in the field of social science. Thus, the authors called on more experimental evidence to verify the applicability in other disciplines.

Studies in the Vygotsky-group centered on the use of AWE in education. The pioneering publication in this group was a book written by Vygotsky (1978). This book was essentially a collection of Vygotsky's essays, emphasizing the basic theories and data in children's development. It also applied theories and main methodological points to issues in cognitive psychology, trying to provide educational implications for researchers and practitioners. The most frequently cited publication in Group 2 was an article published in 2008 (Chen & Cheng, 2008). At that time, the majority of studies evaluated the validity of AWE software. However, the authors decided to go beyond AWE technology itself. They discussed different ways of AWE implementation in three English writing classes and students' perceived effectiveness of it. The findings indicated that students held mixed feelings about the AWE tool, i.e., MY Access!, favouring a combination of computer-generated feedback and human feedback. The most recent cited article, published in 2021, also focused on AWE implementation (Wilson, Huang, et al., 2021). This study examined the effect of MI Write usage on elementary students' writing quality and state test writing performance. The results revealed that the AWE program exerted limited and mixed predictive influence on writing outcomes when controlling other factors, including writing self-efficacy and prior writing performance. Although AWE tools were not fully utilized, teachers and students were positive towards it.

The Halliday-group focused on the effect of AWE on linguistic features. The publication that appeared first in this group was from Halliday (1976). This is a book titled "Cohesion in English" highlighting the importance of cohesion. It mainly discussed textual cohesion and common cohesive devices, i.e., reference, substitution, ellipsis, and conjunction. Moreover, the author also introduced a method to analyze and code sentences. The most highly cited article with a citation score of 32 in Group 3 was published in 2004 (Graesser et al., 2004). The authors developed a computer tool, Coh-Metrix, which was designed to analyze texts in terms of cohesion, language, and readability. They introduced the way to use Coh-Metrix, identifier information, and measures provided by the computer tool. Nevertheless, they did not evaluate its validity in this study. A recent publication in this group written by Mahadini et al. (2021) examined the validity of Coh-Metrix. The traditional scoring rubric was overshadowed by this automated tool in time consumption and information quantification. However, the combination of conventional rubric and Coh-Metrix provided a comprehensive understanding of essay evaluation.

Overall, the analysis of citation patterns in AWE studies resulted in three major groups. Dated back to three seminal pioneering publications, the three groups published the majority of heavily cited documents. There were distinct characteristics of these three groups. The Cohen-group focused on the design and development of technologies associated with AWE, while both the Vygotsky group and the Halliday group discussed the application of AWE. However, compared with the Vygotsky group, the Halliday group focused on a more specific topic, i.e., linguistic features. Thus, the principal

difference was the discrepancy between the design and the application of AWE technology. AWE programs could not be used effectively until the technology has developed and been mature. That may be the reason why the Cohen-group appeared 15 years earlier than the Vygotsky group and the Halliday group. It also revealed that there should be a time difference between development and validation of AWE technology.

CONCLUSION

Major Contributions

This study adopted the bibliometric analysis to provide an overview of AWE research literature by using VOSviewer and CitNetExplorer software. First, it was found that the prolific authors, institutions, and countries mainly came from North America, Asia, and Europe. The keywords in AWE studies ranged from technological terms to educational terms, with eight clusters. Three major groups in citation network analysis showed that different themes of AWE research were complementary to each other. Second, this study linked the missing literature on the bibliometric analysis of previous AWE studies. It also followed the suggestion of van Eck and Waltman (2017) to combine VOSviewer and CitNetExplorer, providing a solid foundation for researchers and practitioners.

LIMITATIONS

As with other research articles, this study still has limitations. First, the CitNetExplorer software could merely import Web of Science (WoS) files, meaning that research literature from other databases may not be analyzed. Thus, the researchers were likely to miss related publications that were not included in WoS database. Second, this study analyzed only representative publications in AWE research. It was impossible to comment on all the publications in the citation network. In addition, the naming of each group theme depends on qualitative interpretations.

IMPLICATIONS

Cluster mapping and citation networks enable researchers to identify topics for future research. Based on keyword co-occurrence analysis, it is suggested that future research could pay more attention to K12 education than higher education. Since most AWE studies focused on language teaching and learning, and some centered on science education, other disciplines are expected to be investigated as well in the future. In terms of genres, apart from argumentative writing, other genres should also be given enough attention. Based on citation network analysis, the development of new models or programs is encouraged to be validated in pedagogical use. It is also recommended that future studies could examine the influence of AWE technology on students' daily life, their learning practices, goals, cultures, and social communities (Roll & Wylie, 2016).

The visualizations also provide some practical implications for researchers. They could refer to the most cited countries and organizations, the most influential researchers, and representative works. This, for researchers, is instrumental in standing on the shoulders of giants to do research that is conducive to scientific development. Otherwise, researchers may well repeat previous works, leading to the waste of time and energy. Considering past successful experiences in other countries, researchers could propose a new framework or theory for their countries (Brika et al., 2021). Moreover, a bibliometric analysis enables readers to get a holistic view of AWE research even if they are new to this field.

There are some practical implications for practitioners as well. First, educational policymakers may realize the status of AWE in education and are expected to take measures supporting future trends in AWE implementation. Second, teachers and instructors are encouraged to integrate AWE technology

into their classes. Third, program developers could deepen their understanding of the history of AWE technology. The pioneering publications could help developers understand the foundation of AWE technology, while the most recent publications may update them on newly developed models and tools, which contributed to improving their own programs.

CONFLICT OF INTEREST

The author has no conflicts of interest to declare that are relevant to the content of this article.

FUNDING AGENCY

REFERENCES

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (2nd ed., pp. 316–329). Academic Press.

Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 1–24. Advance online publication. doi:10.1080/09588221.2021.1936071

Beggrow, E. P., Ha, M. S., Nehm, R. H., Pearl, D., & Boone, W. J. (2014). Assessing scientific practices using machine-learning methods: How closely do they match clinical interview performance? *Journal of Science Education and Technology*, *23*(1), 160–182. doi:10.1007/s10956-013-9461-9

Brika, S. K. M., Algamdi, A., Chergui, K., Musa, A. A., & Zouaghi, R. (2021). Quality of higher education: A bibliometric review study. *Frontiers in Education*, *6*, 666087. Advance online publication. doi:10.3389/feduc.2021.666087

Chen, C.-F. E., & Cheng, W.-Y. E. (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. http://llt.msu.edu/vol12num2/chencheng/

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104

Davies, P., Passonneau, R., Muresan, S., & Gao, Y. (2021). Analytical techniques for developing argumentative writing in STEM: A pilot study. *IEEE Transactions on Education*, 1–11. Advance online publication. doi:10.1109/TE.2021.3116202

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. doi:10.1016/j.jbusres.2021.04.070

Fellnhofer, K. (2019). Toward a taxonomy of entrepreneurship education research literature: A bibliometric mapping and visualization. *Educational Research Review*, *27*, 28–55. doi:10.1016/j.edurev.2018.10.002

Fu, Q., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 1–43. Advance online publication. doi:10.1080/09588221.2022.2033787

Gaviria-Marin, M., Merigo, J. M., & Popa, S. (2018). Twenty years of the Journal of Knowledge Management: A bibliometric analysis. *Journal of Knowledge Management*, *22*(8), 1655–1687. doi:10.1108/JKM-10-2017-0497

Goh, T., Sun, H., & Yang, B. (2020). Microfeatures influencing writing quality: The case of Chinese students' SAT essays. *Computer Assisted Language Learning*, *33*(4), 455–481. doi:10.1080/09588221.2019.1572017

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. doi:10.3758/BF03195564 PMID:15354684

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge., doi:10.4324/9781315836010

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Heersmink, R., van den Hoven, J., van Eck, N. J., & van den Berg, J. (2011). Bibliometric mapping of computer and information ethics. *Ethics and Information Technology*, *13*(3), 241–249. doi:10.1007/s10676-011-9273-7

Hibert, A. I. (2019). Systematic literature review of automated writing evaluation as a formative learning tool. In *Proceedings of 14th European Conference on Technology Enhanced Learning* (pp. 199-212). Cham: Springer. doi:10.1007/978-3-030-29736-7_15

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

Huang, X. Y., Zou, D., Cheng, G. R., Chen, X. L., & Xie, H. R. (2021). Trends, research issues and applications of artificial intelligence in language education. *Journal of Educational Technology & Society*, *24*(3), 238–255.

Kumar, V. S., & Boulanger, D. (2021). Automated essay scoring and the deep learning black box: How are rubric scores determined? *International Journal of Artificial Intelligence in Education*, *31*(3), 538–584. doi:10.1007/s40593-020-00211-5

Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, *97*, 102461. Advance online publication. doi:10.1016/j.system.2021.102461

Liu, M., Liu, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. doi:10.1109/TLT.2016.2612659

Mahadini, M. K., Setyaningsih, E., & Sarosa, T. (2021). Using conventional rubric and Coh-Metrix to assess EFL students' essays. *International Journal of Language Education*, *5*(4), 260–270. doi:10.26858/ijole.v5i4.19105

Miranty, D., & Widiati, U. (2021). Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts. *Pegem Egitim Ve Orgetim Dergisi*, *11*(4), 126–137. doi:10.47750/pegegog.11.04.12

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 Statement. *Systematic Reviews*, *4*(1), 1. Advance online publication. doi:10.1186/2046-4053-4-1 PMID:25554246

Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. (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. doi:10.1111/jcal.12635

O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*(1), 42–56. doi:10.14742/ajet.3795

Parra, L., & Calero, X. (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

Perelman, L. (2014). When "the state of the art" is counting words. *Assessing Writing*, *21*, 104–111. doi:10.1016/j.asw.2014.05.001

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. doi:10.1007/s40593-016-0110-3

Saha, S. K., & Rao, C. H. D. (2022). Development of a practical system for computerized evaluation of descriptive answers of middle school level students. *Interactive Learning Environments*, *30*(2), 215–228. doi:10.1080/10494820.2019.1651743

Shang, H. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, *30*(1), 4–16. doi:10.1080/10494820.2019.1629601

Sherafati, N., Largani, F. M., & Amini, S. (2020). Exploring the effect of computer-mediated teacher feedback on the writing achievement of Iranian EFL learners: Does motivation count? *Education and Information Technologies*, *25*(5), 4591–4613. doi:10.1007/s10639-020-10177-5

Shermis, M. D., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Routledge., doi:10.4324/9781410606860

Shermis, M. D., Burstein, J., & Bursky, S. A. (2013). Introduction to automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 1–15). Routledge.

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

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, *91*, 102247. Advance online publication. doi:10.1016/j.system.2020.102247

van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, *8*(4), 802–823. doi:10.1016/j.joi.2014.07.006

van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053–1070. doi:10.1007/s11192-017-2300-7 PMID:28490825

van Eck, N. J., & Waltman, L. (2021, July 22). *VOSviewer manual*. Universiteit Leiden. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.17.pdf

Vogel, R., & Masal, D. (2015). Public Leadership: A review of the literature and framework for future research. *Public Management Review*, *17*(8), 1165–1189. doi:10.1080/14719037.2014.895031

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (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. Advance online publication. doi:10.1016/j.compedu.2021.104208

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. doi:10.1007/s40593-020-00236-w

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87–125. doi:10.1177/0735633119830764

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

Wulff, P., Buschhuter, D., Westphal, A., Nowak, A., Becker, L., Robalino, H., Stede, M., & Borowski, A. (2020). Computer-based classification of preservice physics teachers' written reflections. *Journal of Science Education and Technology*, *30*(1), 1–15. doi:10.1007/s10956-020-09865-1

Yilmaz, R. M., Topu, F. B., & Tulgar, A. T. (2019). An examination of the studies on foreign language teaching in pre-school education: A bibliometric mapping analysis. *Computer Assisted Language Learning*. Advance online publication. doi:10.1080/09588221.2019.1681465

Yu, Z. (2020a). Visualizing artificial intelligence used in education over two decades. *Journal of Information Technology Research*, *13*(4), 32–46. doi:10.4018/JITR.2020100103

Yu, Z. (2020b). Visualizing co-citations of technology acceptance models in education. *Journal of Information Technology Research*, *13*(1), 77–95. doi:10.4018/JITR.2020010106

Yu, Z. (2021). A literature review on MOOCs integrated with learning analytics. *Journal of Information Technology Research*, *14*(2), 67–84. doi:10.4018/JITR.2021040104

Yu, Z., Gao, M., & Wang, L. (2021). The effect of educational games on learning outcomes, student motivation, engagement and satisfaction. *Journal of Educational Computing Research*, *59*(3), 522–546. doi:10.1177/0735633120969214

Yu, Z., Xu, W., & Yu, L. (2022). Constructing an online sustainable educational model in COVID-19 pandemic environments. *Sustainability*, *14*(6), 3598. Advance online publication. doi:10.3390/su14063598

Yuan, S., He, T., Huang, H., Hou, R., & Wang, M. (2020). Automated Chinese essay scoring based on deep learning. *Computers Materials & Continua*, *65*(1), 817–833. doi:10.32604/cmc.2020.010471

Zaini, A. (2018). Word processors as monarchs: Computer-generated feedback can exercise power over and influence EAL learners' identity representations. *Computers & Education*, *120*, 112–126. doi:10.1016/j.compedu.2018.01.014

Zhang, L., Carter, R. A. Jr, Qian, X., Yang, S., Rujimora, J., & Wen, S. (2022). Academia's responses to crisis: A bibliometric analysis of literature on online learning in higher education during COVID-19. *British Journal of Educational Technology*, *53*(3), 620–646. Advance online publication. doi:10.1111/bjet.13191 PMID:35600420

Zhang, L., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. Advance online publication. doi:10.1016/j.asw.2021.100586

Zhu, M., Lee, H., Wang, T., Liu, O. L., Belur, V., & Pallant, A. (2017). Investigating the impact of automated feedback on students' scientific argumentation. *International Journal of Science Education*, *39*(12), 1648–1668. doi:10.1080/09500693.2017.1347303

Zhu, M., Liu, O. L., & Lee, H. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, *143*, 103668. Advance online publication. doi:10.1016/j.compedu.2019.103668