Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society*, *26*(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009

Trends, Research Issues and Applications of Artificial Intelligence in Language Education

Xinyi Huang[1], Di Zou[2*], Gary Cheng[1], Xieling Chen[1] and Haoran Xie[3]

[1]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong // [2]Department of English Language Education, The Education University of Hong Kong, Hong Kong // [3]Department of Computing and Decision Sciences, Lingnan University, Hong Kong SAR // hxinyi@eduhk.hk // dizoudaisy@gmail.com // chengks@eduhk.hk // xielingchen0708@gmail.com // hrxie2@gmail.com

[*]Corresponding author

ABSTRACT: Artificial Intelligence (AI) plays an increasingly important role in language education; however, the trends, research issues, and applications of AI in language learning remain largely under-investigated. Accordingly, the present paper, using bibliometric analysis, investigates these issues via a review of 516 papers published between 2000 and 2019, focusing on how AI was integrated into language education. Findings revealed that the frequency of studies on AI-enhanced language education increased over the period. The USA and Arizona State University were the most active country and institution, respectively. The 10 most popular topics were: (1) automated writing evaluation; (2) intelligent tutoring systems (ITS) for reading and writing; (3) automated error detection; (4) computer-mediated communication; (5) personalized systems for language learning; (6) natural language and vocabulary learning; (7) web resources and web-based systems for language learning; (8) ITS for writing in English for specific purposes; (9) intelligent tutoring and assessment systems for pronunciation and speech training; and (10) affective states and emotions. The results also indicated that AI was frequently used to assist students in learning writing, reading, vocabulary, grammar, speaking, and listening. Natural language processing, automated speech recognition, and learner profiling were commonly applied to develop automated writing evaluation, personalized learning, and intelligent tutoring systems.

Keywords: Artificial Intelligence, Language Education, Bibliometric Analysis, Automated Writing Evaluation, Intelligent Tutoring System

1. Introduction

Humans have been steadily improving their learning by employing new technologies. One of the newest technologies in the modern era is Artificial Intelligence (AI), which is defined as “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments” (Organisation for Economic Co-operation and Development, 2019, p. 7). AI has great potential for education as it can generate predictive and diagnostic models for precision education, help visualize students at risk, provide timely intervention, and reduce dropout rates (Lu et al., 2018). Personalized learning systems, software agents, ontologies, and the semantic web are the major AI techniques for education (Hinojo-Lucena et al., 2019). Hwang et al. (2020) categorized AI-powered Education (AIEd) applications into four types. The first type is the intelligent tutor, which can satisfy students’ needs and promote positive learning outcomes. The intelligent tutee is another AIEd application that encourages learners to be tutors and participate in active learning. Intelligent learning tools or partners, the third type, collect and analyze students’ data to enhance learning. The fourth type, policy-making advisor applications, assist administrators in understanding educational trends and problems and help them make effective decisions (Hwang et al., 2020).

Researchers and practitioners of technology-enhanced language learning (TELL) have been applying a wide range of educational technology in language education for three decades (Zou et al., 2018). One of the challenges of using technologies for language learning is that students with different proficiency levels might not achieve the same learning outcomes (Shadiev & Yang, 2020). To solve this problem, machine learning algorithms and data analysis techniques can be used to develop personalized learning systems (Cui et al., 2018). Personalized learning systems allow learners with low language proficiency to learn at their own pace to maximize their progress (Chen et al., 2021a). Heil (2016) observed that many current applications for language learning are decontextualized, lacking authentic speech production. However, AI-enhanced approaches can address this limitation as well. For example, Chen et al. (2019) developed a context-aware ubiquitous language learning system. With a GPS function, this system can support location-based contextualized English learning. The results indicated that students showed high motivation while learning with this AI-enhanced contextualized system and achieved a satisfactory performance. Thus, it appears that AI has great potential for language education and can solve some existing problems and issues in TELL.

ISSN 1436-4522 (online) and 1176-3647 (print). DOI 10.30191/ETS. This article of *Educational Technology & Society* is available under Creative Commons CC-BY-NC-ND 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Editors at ets.editors@gmail.com.

The trend of integrating AI into education has hastened the need to analyze AI research. Previous reviews have mainly focused on AI in education in general (e.g., Chen et al., 2020a; Song & Wang, 2020; Zawacki-Richter et al., 2019), while few studies have been conducted examining AI in specific domains, such as language education. To fill this gap, the present study aims to provide a comprehensive review of research on AI-enhanced language learning by noting publication trends, the main research issues, and the most frequently used AI applications in language education during the period 2000-2019. The following research questions guided our study:

- What were the publication trends regarding AI in language education in terms of years, journals, countries, and institutions?
- What were the main research issues in AI-enhanced language education?
- What were the common applications of AI in language education?

2. Literature review

2.1. Review on AI in education

Among all the review studies we found on AI in education, five articles appear to be both the most representative and recent. Chen et al. (2020a), who analyzed the AIEd literature from 1999 to 2019, identified a rising frequency of articles in the area, with slow growth between 1999 and 2002, steady growth between 2003 and 2011, and rapid growth between 2012 and 2019. Concerning the key terms used in AIEd, “education,” “machine learning,” “robotics,” “artificial intelligence,” and “deep learning” were most frequently used. As for the terms that received growing attention from researchers in recent years, the top ones were “classification,” “STEM” (i.e., science, technology, engineering, and mathematics), “computational thinking,” “educational data mining,” and “neural networks.” Similar results were reported in another study (Chen et al., 2021b) on the past, present, and future of smart learning, an important sub-field of AIEd. This review conducted a topic-modeling analysis of 555 relevant articles from 1989 to 2019, identifying several important research issues, including interactive and multimedia learning, STEM education, smart learning analytics, software engineering for e-learning systems, the Internet of Things, and cloud computing.

Zawacki-Richter et al. (2019), who analyzed AI applications in higher education from 2007 to 2019 globally, found that profiling and prediction, assessment and evaluation, adaptive systems and personalization, and ITS were the major AI-enhanced education areas. Regarding profiling and prediction, most studies adopted machine learning methods to model students’ profiles and make predictions. For assessment and evaluation, automated grading systems were frequently used to grade assignments and provide feedback. Adaptive and personalized systems provided academic advice and personalized learning content. ITSs were mainly used to deliver course content and provide learning materials.

Song and Wang (2020) further broadened the scale of analysis by reviewing the development of Educational Artificial Intelligence (EAI) from 2000 to 2019. They proposed that EAI research could be conceptualized as having four stages. The first stage (2000-2004) concentrated on developing intelligent robots, computer programming, and Virtual Reality (VR). A breakthrough in AI occurred during the second stage (2005-2009), with foci on intelligent tutors and educational computing. During the third stage (2010-2014), deep neural networks led to the development of automatic pattern recognition, speech recognition, and image classification. AI infiltrated education at the final stage (2014-2018) when distance education, adaptive learning, e-learning, and data mining became popular.

2.2. Review on AI in language education

Several researchers have conducted reviews on AI in language education. Gamper and Knapp (2002) investigated 40 Intelligent Computer-Assisted Language Learning (ICALL) systems finding that AI techniques such as User Modelling, Natural Language Processing (NLP), Natural Language Generation, Automated Speech Recognition (ASR), and Machine Translation were the most frequently utilized in language learning systems. Ali (2020) reviewed the approaches to integrating AI in language education through content analysis. Ali’s review specifically focused on ASR, which recognizes human speech, identifies linguistic features, and assists in human-machine communication. Related to ASR, Chatbots can conduct intelligent conversations through a keyword matching technique that assesses students’ speaking abilities. AI-amalgamated flipped classrooms can also effectively enhance students’ learning performance and motivation. Therefore, researchers have generally displayed positive attitudes towards AI-enhanced language learning.

Pokrivcakova (2019) analyzed AI technologies from the language teachers' perspective. In the study, different forms of AI were employed in language education for diverse purposes, including: (1) providing personalized learning content; (2) translating a written/spoken text from one language to another; (3) correcting grammar errors by means of writing assistants; (4) conducting conversations using chatbots; (5) creating smart language learning platforms and apps; (6) enabling personalized language tutoring; and (7) developing intelligent VR for learners to practice speaking. Considering the increasing trend of using AI in education, Pokrivcakova (2019) noted the importance of teacher training in the AI age.

Chen et al. (2021a) focused on the sub-field of precision language education and identified research trends and issues in the domain of personalized language learning after reviewing 108 articles between 2000 and 2019. They found that personalized recommendations, feedback, and assessment were the most frequently investigated topics. Findings revealed that personalized language education was effective as it met different learners' needs and provided them with personalized diagnoses and adaptation.

In sum, although many of the existing reviews on AI in education have focused on general education (e.g., Chen et al., 2020a; Song & Wang, 2020; Zawacki-Richter et al., 2019), a few studies have investigated AI in language education; however, most of these have focused on AI tools and applications in language classrooms, with limited research on AI research trends in language education. Moreover, the sample sizes of most previous review studies have been relatively small (e.g., Ali, 2020; Gamper & Knapp, 2002). The wide application of AI in language classrooms suggests AI is playing a significant role in language education, which indicates there is a need to analyze the current research status of AI-enhanced language learning using a computational method that can provide a more comprehensive analysis of the literature. Accordingly, the present study provides an overview of the status of AI in language education by analyzing research trends and the most-discussed topics using bibliometric analysis.

3. Research method

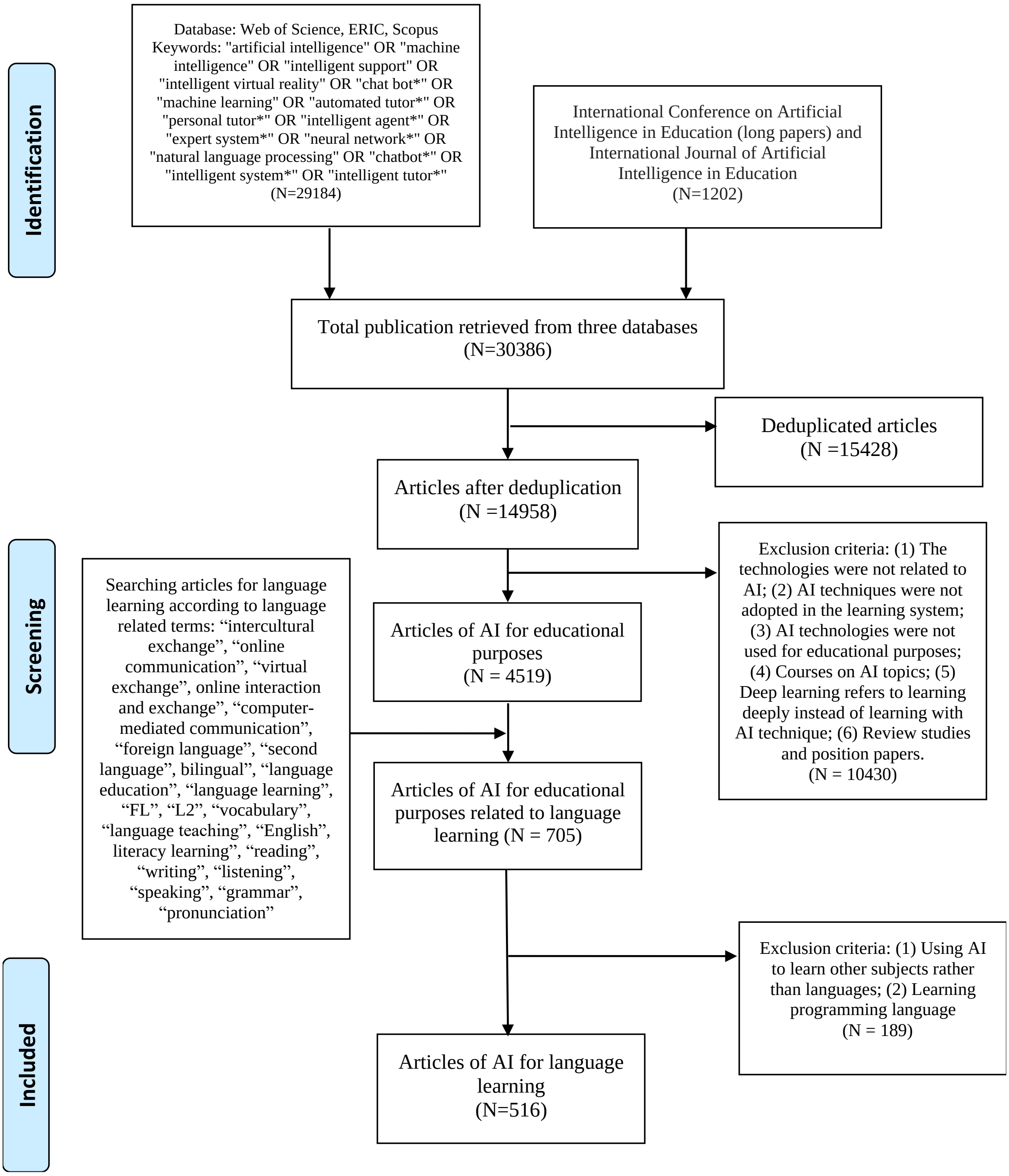
Bibliometric analysis was adopted in this research because it can effectively evaluate the academic status of a particular research area (Chen et al., 2020b; Chen et al., 2020c). Many researchers have employed this method to investigate research trends in different areas. Studies have also applied it to analyze AI in education (e.g., Hinojo-Lucena et al., 2019; Song & Wang, 2020) and language learning (e.g., Chen et al., 2018; Gong et al., 2018). Hence, this method was considered applicable for analyzing AI research trends in language education.

3.1. Data retrieval

Web of Science (WoS), Education Resource Information Center (ERIC), and Scopus were chosen for the databases. Many previous review studies have also included data from these services (e.g., Fu et al., 2022; Wang et al., 2019; Zou et al., 2019; Zou et al., 2020). Following Tran et al. (2019), we selected AI-related keywords in education to search for target papers (Figure 1). A total of 29,184 papers related to AI in education (original research articles) and 1,202 publications from the International Conference on Artificial Intelligence in Education (ICAIE) and the International Journal of Artificial Intelligence in Education (IJAIED) were retrieved. Thus, in total, we identified 30,386 publications. We included conference papers in this review as they are the main source of research on AI in education (Hinojo-Lucena et al., 2019). ICAIE is an important conference in the field, so we also included its proceedings papers to present a comprehensive overview of the whole research area.

Figure 1 illustrates the detailed data retrieval process. After deduplication (N = 15,428), two domain experts screened the remaining papers (N = 14,958) based on the following criteria: (1) The papers had to focus on AI technologies; (2) AI technologies had to be used to support learning and teaching; and (3) the studies had to be empirical. Upon completion of this initial screening, the inter-coder agreement was 91%, with differences being decided via discussion, resulting in 4,519 remaining papers. We then consulted previous review studies on technology-enhanced language learning and identified 22 language-related keywords (Chen et al., 2021c; Fu et al., 2022; Su & Zou, 2020; Wang et al., 2019; van den Berghe et al., 2019; Zhang & Zou, 2020). Using these keywords (see Figure 1), we searched the titles, abstracts, and keywords of the 4,519 papers and selected those that applied AI for language learning purposes. In total, we found 705 papers. At the final stage, two domain experts examined these papers and excluded those that used AI to learn other subjects or programming languages. The inter-coder agreement was 95%, with differences being resolved via discussion. A total of 516 papers were finalized for review.

Figure 1. Process of data retrieval



3.2. Structural topic modeling

Structural topic modeling (STM) (Roberts et al., 2014) was adopted to identify the latent topics from the 516 papers. STM can identify the principal features of a corpus using machine learning algorithms (Grajzl & Murrell, 2019). We applied this method to extract terms from the titles, abstracts, and keywords. As suggested by Chen et al. (2022), 0.4, 0.4, and 0.2 were respectively assigned as the weights to the terms from keywords, titles, and abstracts. We also employed Term Frequency-Inverse Document Frequencies (TF-IDF) to filter terms according to their importance. Originally, there were 5,582 terms. We set the threshold of TF-IDF as 0.03, 0.04 and 0.05 and found 5,463, 4,935 and 3,807 terms, respectively. We selected terms with 0.04 TF-IDF because 0.03 TF-IDF included terms that were not very relevant (i.e., admissible, diagramming), while 0.05 TF-IDF did not include some important terms (e.g., learn, read). Thus, 0.04 TF-IDF appeared most appropriate. Following previous research (Chen et al., 2020c; Chen et al., 2020d), we ran a set of 16 models by setting the number of topics ranging from 5 to 20. We compared each model by examining the representative terms and articles according to the following criteria. First, a meaningful topic had to be formed based on the representative terms; second, all articles had to be highly related to the identified topic; third, all topics within a topic model had to be different; and fourth, all crucial dimensions of AI in language education had to be included.

After comparing the 16 models with different numbers of topics, we chose the 10-topic model. We then generated the statistical results based on the level of importance of the topics and obtained the key terms from the topics following the distribution matrix to label the topics. Thereafter, two domain experts interpreted the semantic meanings of each key term and analyzed the representative articles for each topic. Finally, two researchers summarized each topic’s labels independently and compared the labeling results to ensure consistency.

3.3. Performance analysis

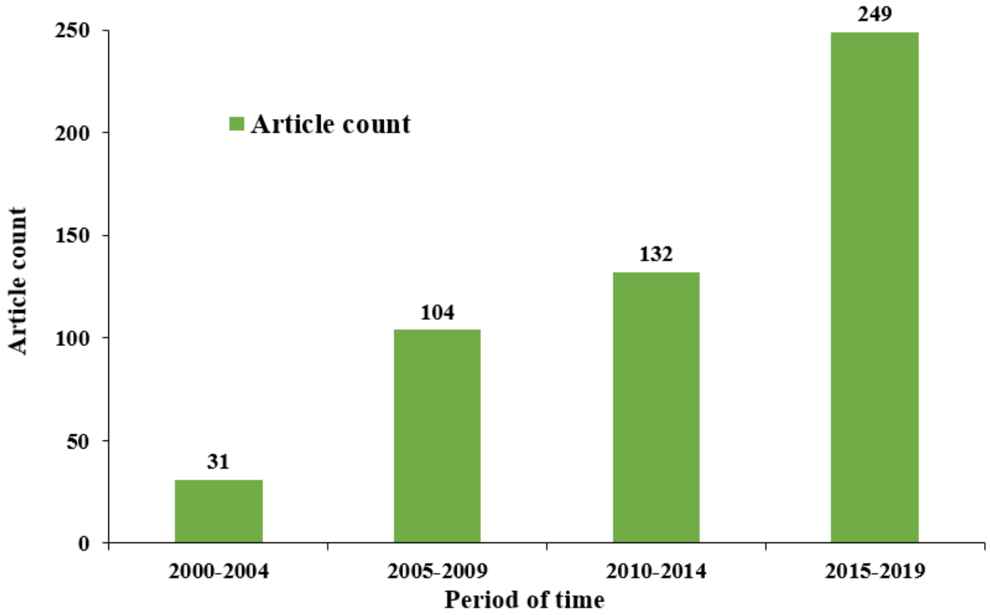
A performance analysis was conducted to investigate the academic outputs of the journals, institutions, and countries/regions. Several indicators were used, including the Hirsch index (H-index), Article count (A), Citation count (C), and Average Citation per Article (ACP). H-index considers both the number of works that have been published as well as the citations of those published papers, which indicates the relevance of the research (Hirsch & Buela-Casal, 2014) and is one of the most recognized indicators of academic impact (Svensson, 2010). We calculated the citation counts using Google Scholar on 30th May 2020. Google Scholar identifies the most relevant academic information from a given query and offers the citation data, and it is widely considered reliable (Martín-Martín et al., 2018). Many previous review studies have also used the citation counts of Google Scholar for bibliometric analysis (e.g., Chen et al., 2020a; Dey et al., 2018; Wang & Preminger, 2019).

4. Research results

4.1. Publication trends

Figure 2 shows the frequency of articles published on AI-enhanced language learning from 2000 to 2019. A rising trend can be observed, indicating that researchers have paid increasing attention to the field. Researchers paid comparatively little attention to AI-enhanced language learning between 2000 and 2004, while there was a sharp increase during 2005 and 2009. The number of publications kept increasing in the third period (2010-2014) and reached the highest number in the last period (2015-2019).

Figure 2. AI-related publication frequency



4.2. Influential publication sources

Figure 3 presents the top 15 sources that contributed to the research field. The three most influential sources based on the H-index were the *IJAIED, Computers & Education*, and *ICAIE*, with H-indexes of 25, 24, and 23 respectively.

ICAIE and *IJAIED* respectively published 264 and 104 articles on AI in language education and accounted for 71% of the total number (516) (see Figure 4).

As for citation counts (Figure 5), the most influential sources were the *ICAIE* (3,294), *IJAIED* (2,540), and *CALICO Journal* (2,234).

As shown in Figure 6, *Bilingualism: Language and Cognition* had the highest ACP (109.92), followed by *Educational Technology & Society* (53.62) and *Language Learning* (52).

Figure 3. Top 15 sources: H-index

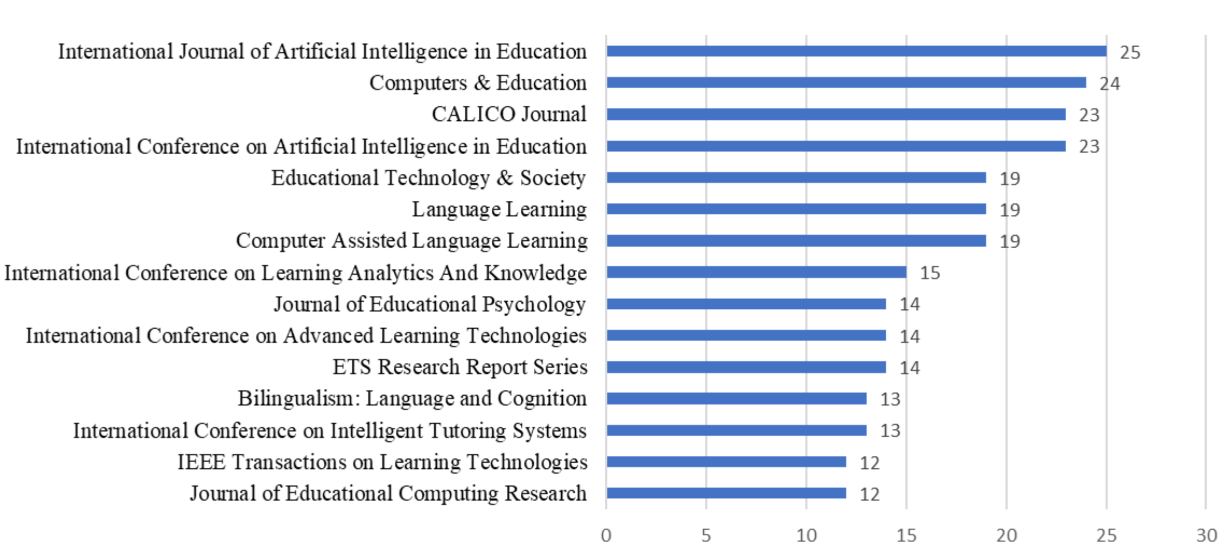


Figure 4. Top 15 publications: Article counts

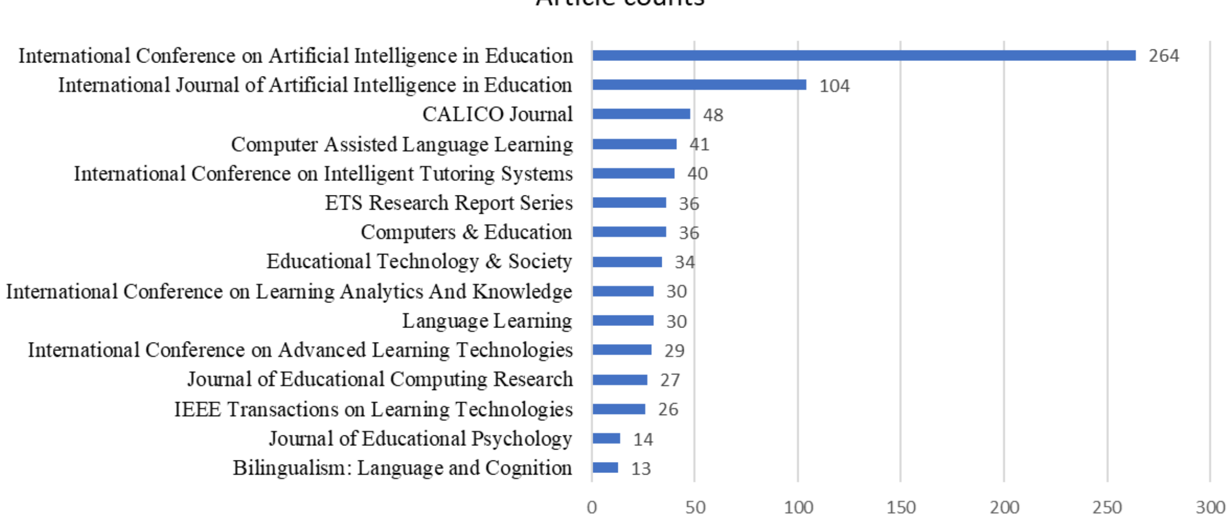


Figure 5. Top 15 publications: Citation counts

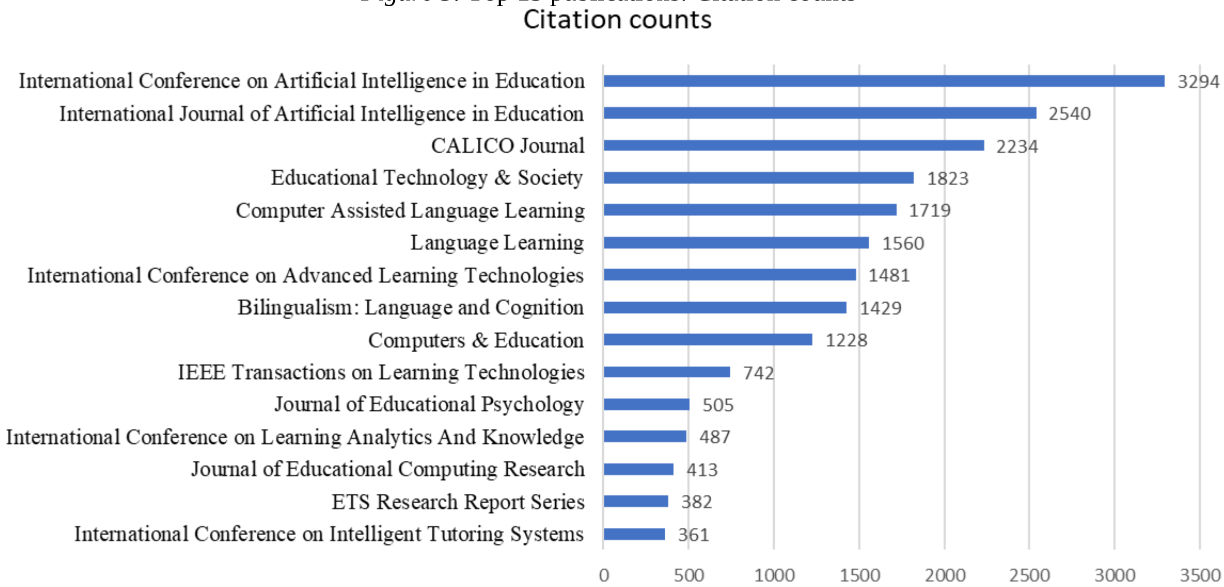
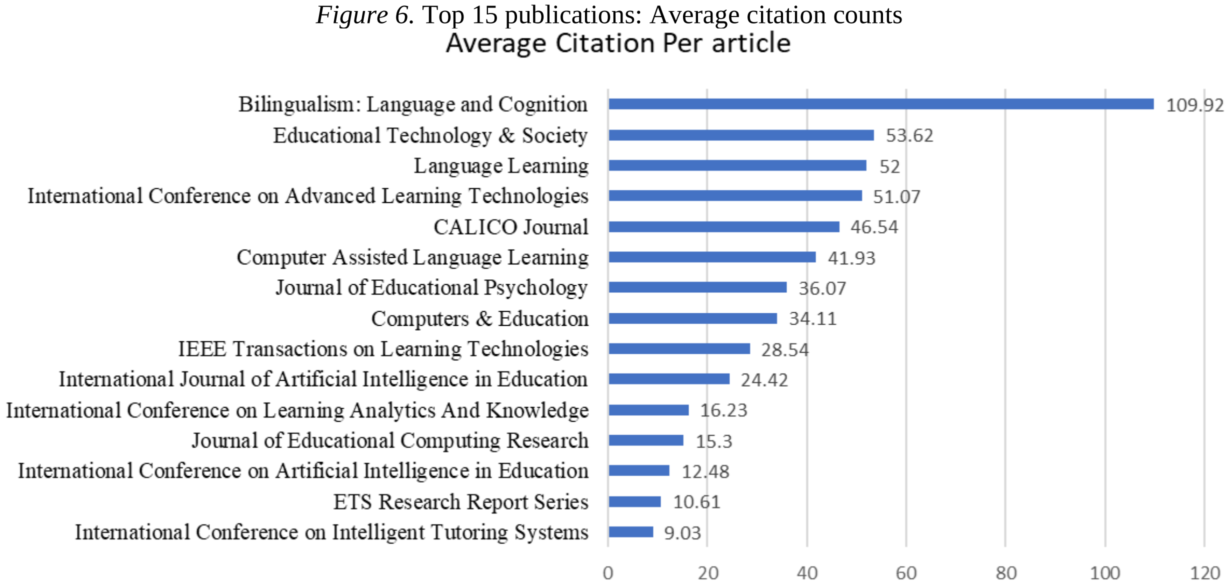


Figure 6. Top 15 publications: Average citation counts



4.3. Representative articles

We also ranked the articles according to their total and annual citations until September 13rd 2020 to identify the most representative studies and analyze their main findings.

- The 10 studies with the highest total citations were Stockwell (2007), Chen et al. (2006), Chen and Li (2010), Johnson et al. (2005), Grimes and–Warschauer (2010), Johnson (2007), Calvo et al. (2010), McNamara et al. (2013), Roscoe and McNamara (2013), and McNamara et al. (2015).
- The 10 studies with the highest annual citations were Stockwell (2007), Chen et al. (2006), Chen and Li (2010), McNamara et al. (2015), Alexopoulou et al. (2017), McNamara et al. (2013), Kyle and Crossley (2018), Grimes and-Warschauer (2010), Roscoe and McNamara (2013), and Vajjala (2018).

Many studies with the highest total citations also had the highest annual citations, four of which were on AI-enhanced writing. NLP techniques were applied for essay quality evaluation and immediate feedback (Fu et al., 2022). Grimes and Warschauer (2010) investigated teachers' and students' attitudes towards an Automated Writing Evaluation (AWE) tool called MY Access!. This system grades students' essays and provides automated feedback. Results showed that teachers regarded the automatic scoring function as useful because it saved them time, and students considered it helpful for revising and enhancing their writing skills. iWrite was used in Calvo's et al. (2010) study to support collaborative writing activities by helping students revise their group work.

It was found that students spent more time on collaborative writing because the system enabled all group members to view their work, which promoted individual participation. Similarly, McNamara et al. (2013) developed an ITS (Writing Pal) to teach students writing strategies such as generating ideas, organizing essays, and revising essays. This ITS also evaluated essay quality and generated automatic feedback for students. Results indicated that the ratings of this system were similar to that of human graders. Roscoe and McNamara (2013) further examined the feasibility of using this system in writing classrooms. Results from their surveys indicated that students perceived the lessons given by the system as beneficial and informative.

NLP technologies were also used for language feature analysis in AWE and Automated Essay Scoring (AES) systems in the reviewed studies. McNamara et al. (2015) applied a hierarchy classification approach to the AES system that could evaluate essays according to their length and quality and predict scores. Results showed that this approach had a higher accuracy than other AWE systems since it used a set of thresholds to predict essay scores. Alexopoulou et al. (2017) investigated the effects of tasks on learners' written language by analyzing their work using NLP techniques. The results revealed that learners' writing of professional tasks, i.e., writing a job advertisement, had lower error rates than narrative tasks, i.e., storytelling. This is perhaps because professional tasks are normally in bullet-point form. In Kyle and Crossley's (2018) study, NLP was employed to extract language features from the essays of the Test of English as a Foreign Language (TOEFL) and analyze the syntactic complexity of learners' writing. They found that the fine-grained indices of phrasal complexity were the best predictors of learners' writing quality scores because they provided complimentary explanatory power. Similarly, Vajjala (2018) identified the most predictive features in different AES and AWE systems adopting NLP techniques to build predictive models. The researchers concluded that document length played an important role in predicting TOEFL writing scores, and discourse features were an important predictor in Cambridge First Certificate in the English dataset.

In the representative studies, both NLP and ASR were used to enhance communication in game settings. Johnson et al. (2005) integrated AI and serious games into the Tactical Language Training System (TLTS) for language and cultural learning. Learners interacted with the Non-Player Characters (NPC) to complete missions in a simulated world. ASR techniques were used to identify the intended meanings of players' utterances, and NLP was adopted to generate dialogues between the players and the NPCs in the game. However, Johnson et al. (2005) did not evaluate the effectiveness of this game, so it is uncertain whether and to what extent students benefited from learning to use this approach. In a follow-up study, Johnson (2007) evaluated the usefulness of the software by inviting users to rate the system with scores from 0 to 5. Findings revealed that 78% of the participants perceived the training positively and they also felt they had acquired some functional ability of the target language.

Learner profiling, fuzzy item theory, and context-aware techniques have also been integrated into ITS to promote vocabulary learning and reading ability. Stockwell (2007) developed a mobile ITS to enhance students' vocabulary learning. This system keeps logs of students' access to the system, creates learner profiles to record the vocabulary with which students were unfamiliar, and presents these words more frequently. Chen et al. (2016) developed a Personalized Mobile Learning System (PLMS) to recommend English articles to students based on their reading ability. The students' reading ability was evaluated by fuzzy item response theory, and articles were retrieved from websites via a crawler agent. The proposed system was beneficial for students as it provided personalized learning. Chen and Li (2010) designed a personalized context-aware ubiquitous system to provide students with relevant vocabulary learning materials according to their locations, ability, learning time, and leisure time. Results showed that students who applied the learning systems with context awareness outperformed those who did not, as the content was appropriate.

4.4. Productive regions and institutions

Figure 7 lists the top 15 countries/regions ranked by the H-index. The most influential country was the USA (H-index = 38), followed by Taiwan (H-index = 15) and Canada, UK, and Japan (H-index = 11).

Figure 8 shows the USA, Taiwan, and Japan also had the highest citation counts, which were 5,808, 1,333, and 678, respectively.

Figure 9 shows the USA (n = 228), Japan (n = 44), and Taiwan (n = 39) produced the greatest number of studies with the USA contributing 44% of the total publications.

New Zealand had the highest ACP (37.14), followed by Taiwan (34.18) and Hong Kong (32.88) (Figure 10).

Figure 7. Top 15 countries/regions: H-index

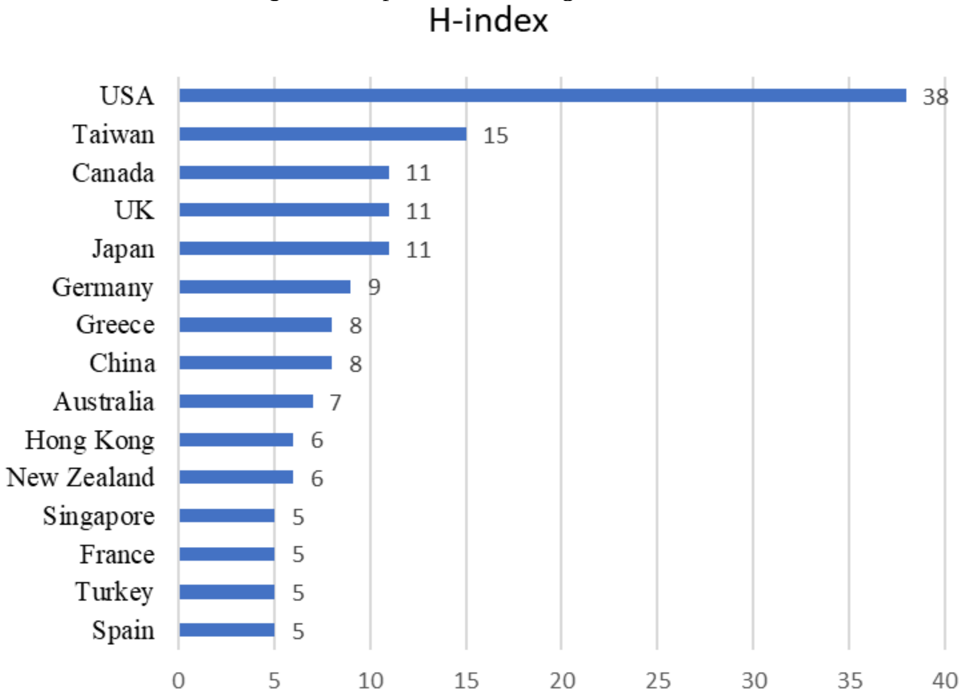


Figure 8. Top 15 countries/regions: Citation counts

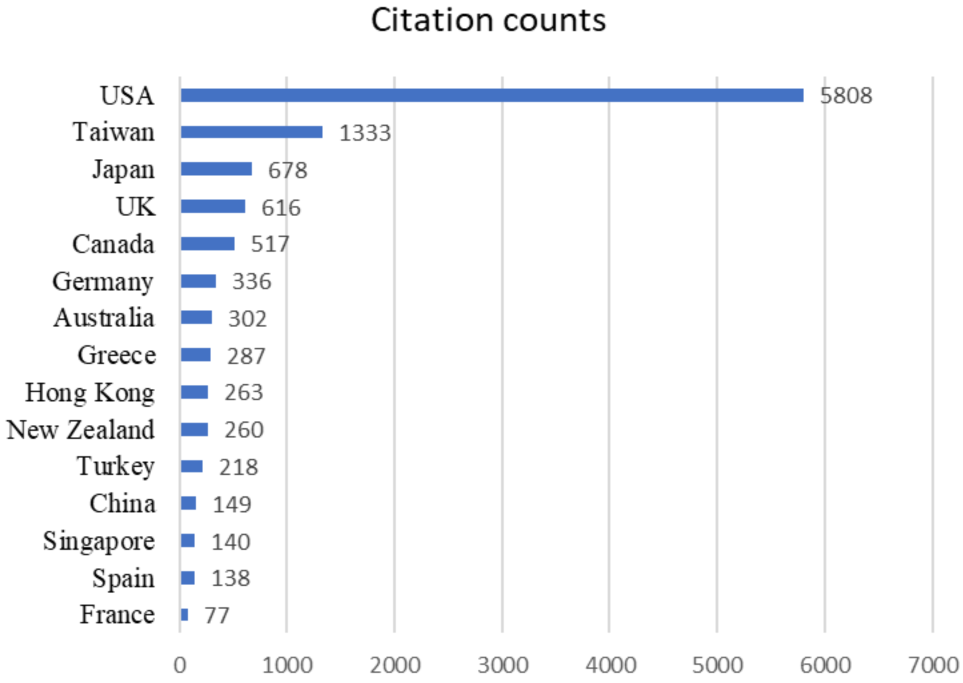


Figure 9. Top 15 countries/regions: Article counts

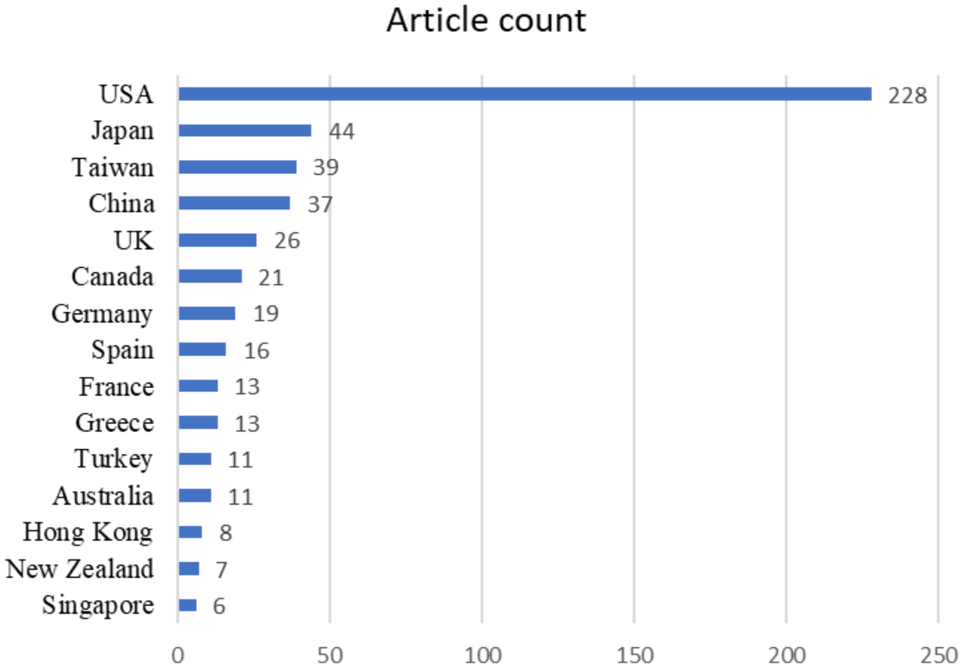
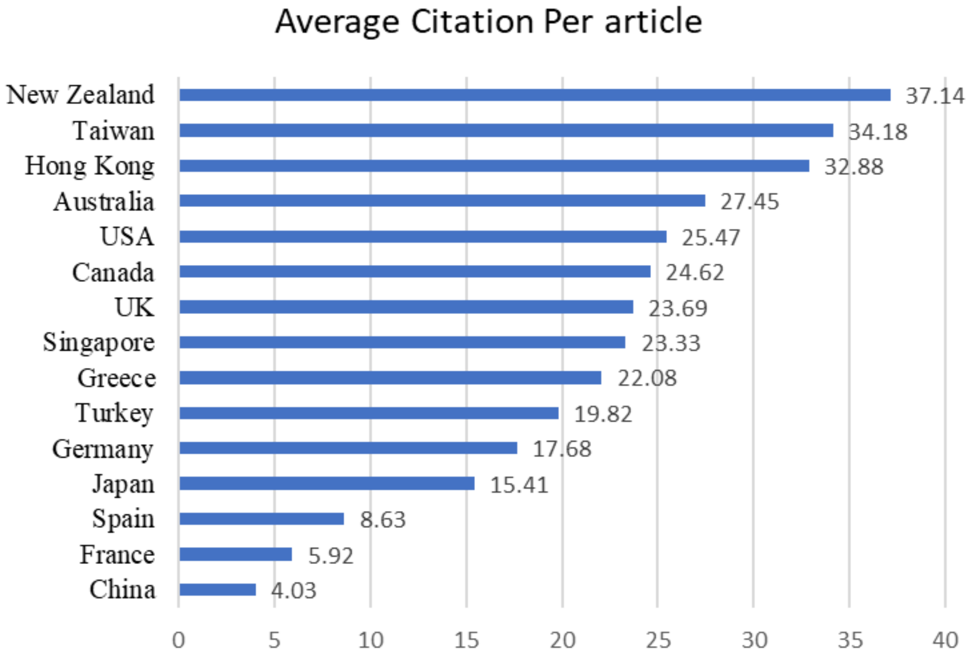


Figure 10. Top 15 countries/regions: ACP



The USA had the highest H-index and the greatest article and citation counts, and 10 out of the 13 top institutions ranked by H-index were from the USA (Figure 11). The other institutions were in Greece (*University of Piraeus*) which ranked seventh, Germany (*University of Tubingen*) which ranked ninth, and Australia (*University of Technology Sydney*) which ranked 13th.

The top three institutions were *Arizona State University* (H-index = 17), *Georgia State University* (H-index = 13), and *Carnegie Mellon University* (H-index = 13). These three also had the greatest article counts, which were 56, 29, and 27, respectively (Figure 12).

Arizona State University was the university that had their papers most frequently cited by researchers (n = 1,093), followed by *Pennsylvania State University* (n = 910) and *Georgia State University* (n = 800) (Figure 13).

The universities with the highest ACP were Pennsylvania State University (n = 82.73), University of Southern California (n = 63.44), and University of Pittsburgh (n = 41.54) (Figure 14).

Figure 11. Top 13 institutions: H-index

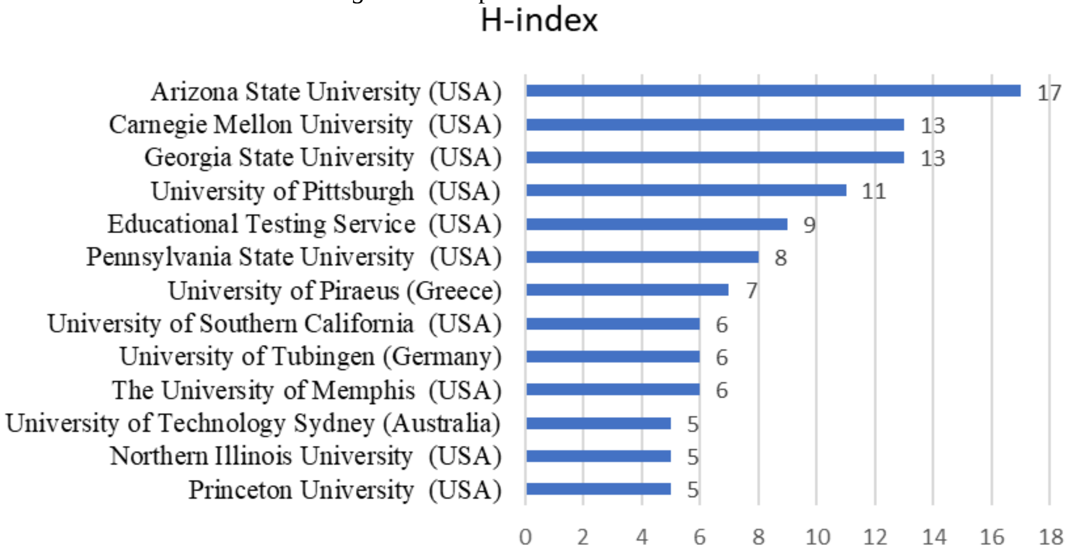


Figure 12. Top 13 institutions: Article count

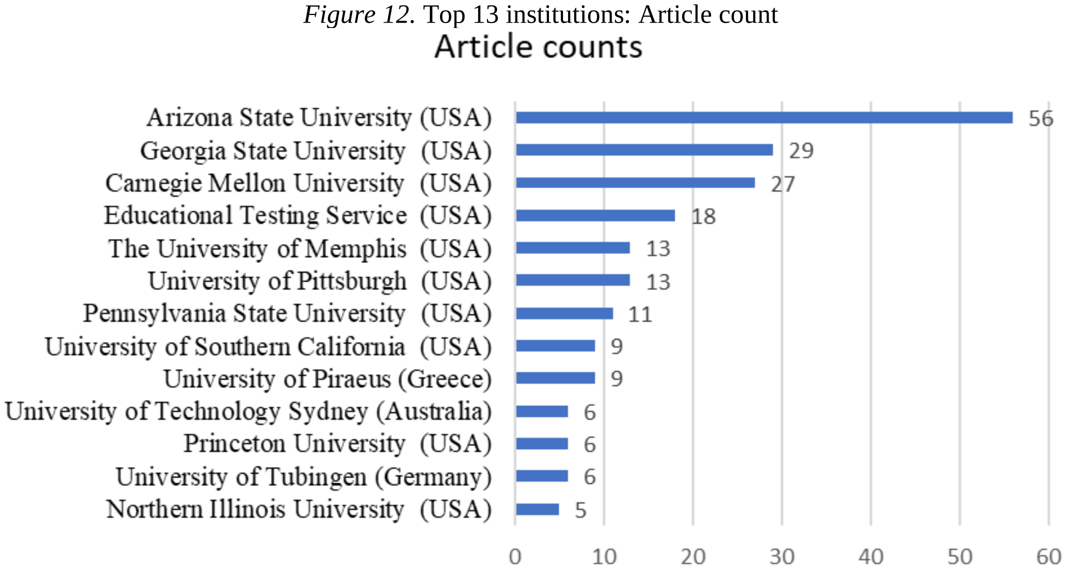


Figure 13. Top 13 institutions: Citation count

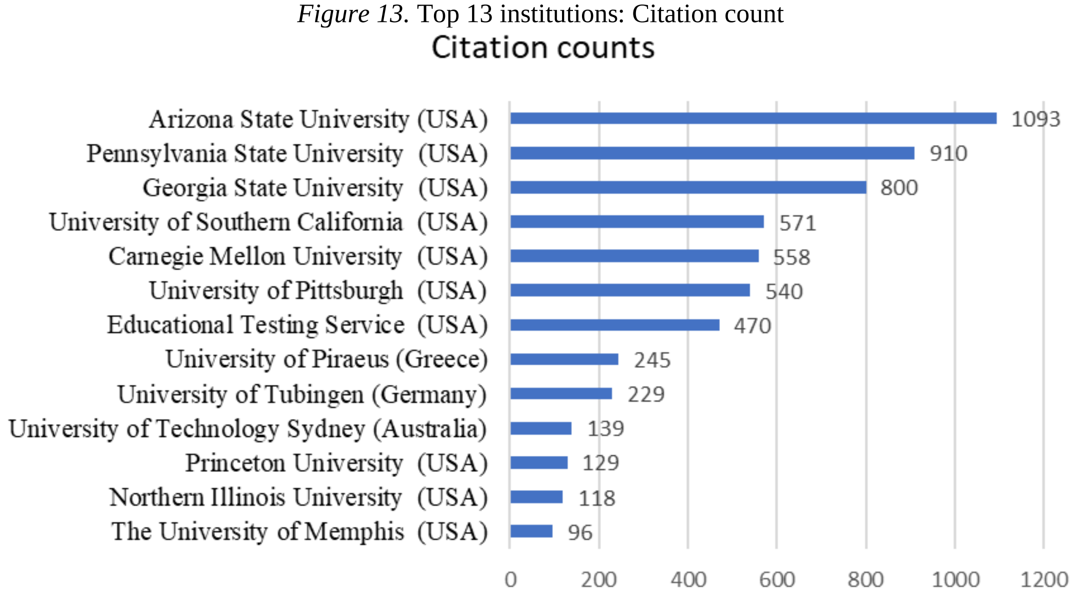
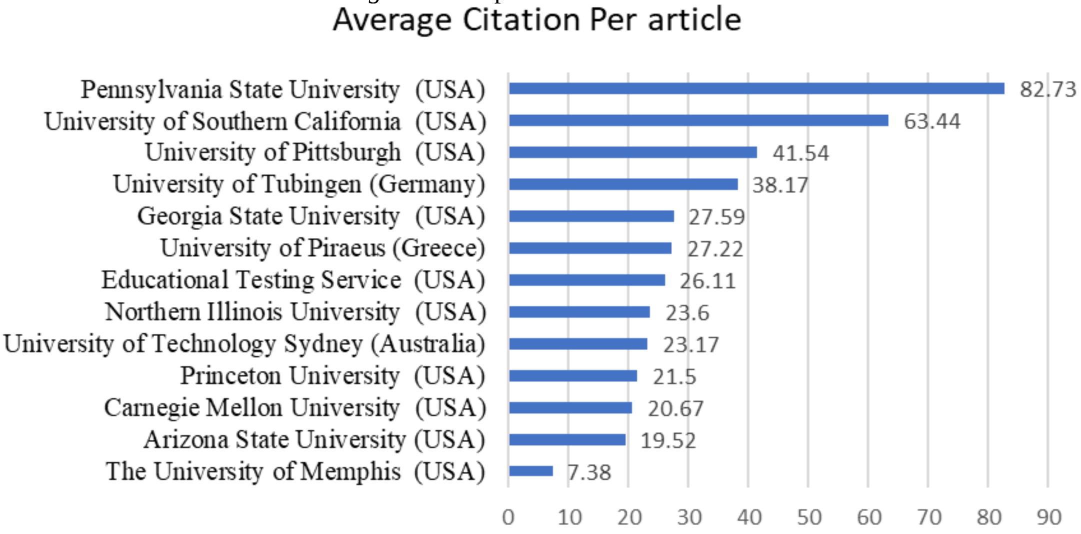


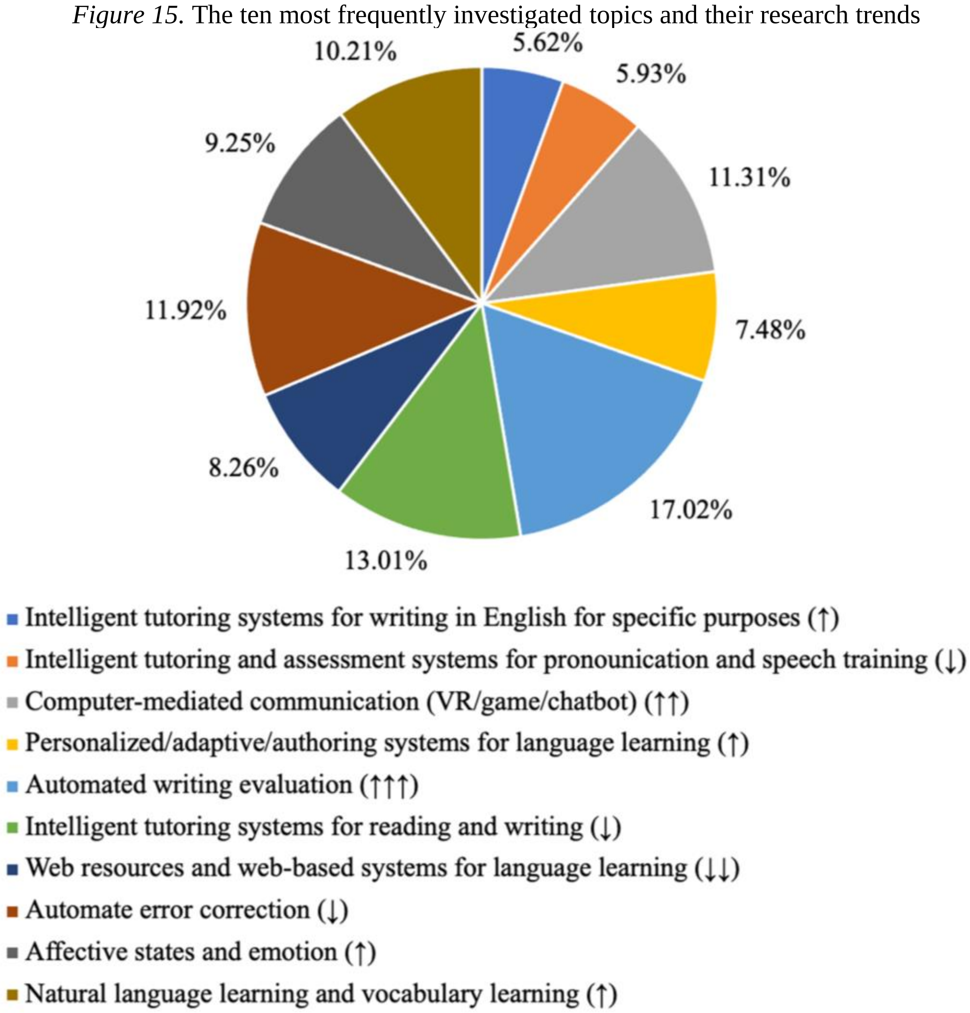
Figure 14. Top 13 institutions: ACP



4.5. Research foci and trends

Figure 15 presents the 10 most frequently investigated topics in AI-assisted language learning and their research trends (indicated by arrows).

Figure 15. The ten most frequently investigated topics and their research trends

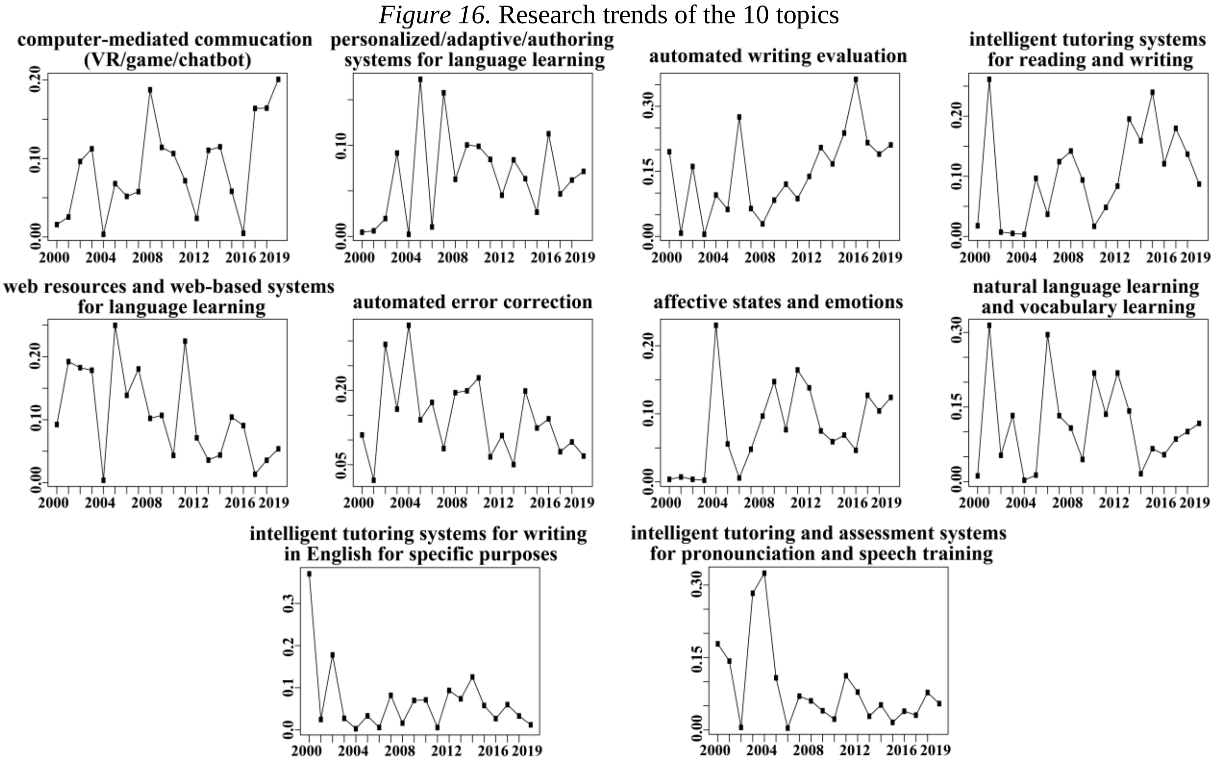


Note. Arrows represent increasing (up) and decreasing (down) interest over time with intensity indicated by the number of arrows.

AWE was the most popular topic accounting for 17.02% of the articles. ITS for reading and writing was comparatively less explored at 13.01%. Automated Error Detection and CMC shared a similar proportion at 11.92% and 11.31%, respectively. The fifth major issue was Natural Language Learning and Vocabulary Learning, which accounted for 10.21% of the reviewed articles. Most of these topics continued to attract research interest towards the end of the review period apart from four topics, i.e., Web resources and Web-based systems for language learning, Automated Error Detection, ITS for writing in English for Specific Purposes, and ITS and Assessment System for pronunciation and speech training. The topics, Web Resources and Web-based systems for language learning, exhibited a significant decreasing trend.

Figure 16 presents the annual article counts of each topic. Both CMC and AWE significantly increased in the later years. Studies on CMC generally increased with fluctuations throughout the period having the smallest number of articles between 2004 and 2016 but having the most in 2019. From 2008, the number of articles on AWE increased and achieved its highest point in 2016 but dropped in the final three years. Web resources and Web-based systems for language learning gradually decreased with a sharp decrease in 2004; in later years, it fluctuated.

Figure 16. Research trends of the 10 topics



5. Discussion

5.1. Research trends of AI in language learning

The results reveal that the number of articles on AI in language learning increased greatly in 2016. Zawacki-Richter et al. (2019), Chen et al. (2020a), and Chen et al. (2022) also showed that research on AI in education increased greatly from 2016, based on which, AI applications developed rapidly. Concerning the number of articles published in each journal, *ICAIE* had the greatest number from 2000 to 2019, followed by *IJAIED*, which is generally consistent with Zawacki-Richter et al. (2019); however, they found that *IJAIED* contributed the most articles from 2007 to 2019. This is likely because we also included conference papers in our review.

Similar to the review findings of Song and Wang (2020) and Zawacki-Richter et al. (2019), we found that the USA had the most publications from 2000 to 2019. However, while Song and Wang (2020) found that the UK and China ranked second and third, respectively, our review ranked Japan and Taiwan in these positions. This is likely because our research focused mainly on AI in language education, whereas they investigated AI in education in general.

Our review indicated that AWE systems were frequently used in language learning, given their potential to reduce teachers' workload and assist students in writing and revision. Additionally, AI may be integrated into VR technology to help students practice their target language in simulated environments (Mirzaei et al., 2018). This innovative approach to CMC drew increasing attention in the later years of our review. With the rising number of ITSs, conventional web-based learning systems drew less research interest in the later years. Similar findings were also reported in Johnson et al. (2017).

5.2. Common applications of AI in language learning

5.2.1. AI applications in learning writing

AI was used to assist students' writing via AWE systems and ITS. These systems evaluate students' work using NLP techniques to diagnose and comment on students' errors so that they have a comprehensive understanding of language use. In Lee et al. (2015), a correction system called Genie Tutor was designed to improve English writing by detecting grammar mistakes and suggesting appropriate expressions. This system guided learners to correct their mistakes in real-time, which is useful for language development. An ITS (i.e., EJP-Write) was also developed to facilitate academic journal writing in Lin et al. (2017). The system's functions, such as citing references and searching for templates, were helpful and effective in providing students with phrase and paragraph templates for better language use.

5.2.2. AI applications in learning reading

ITS was also used to enhance language learners' reading comprehension. For example, Johnson et al. (2017) developed an ITS, Interactive Strategy Training for Active Reading and Thinking (iSTART), for adult literacy learners. iSTART offered instructional videos and exercises for learning comprehension strategies. It also taught summarization strategies and provided learners with interactive narratives to read. The results indicated that learners had positive attitudes towards the narratives. Another example was Wijekumar et al. (2017), who developed an ITS to teach Structure Strategy (ITSS) for enhancing reading comprehension. The ITSS helped students identify text structures and provided hints and feedback in an assessment exercise. The results indicated that the students who used the ITS outperformed those who did not as the system helped organize textual information.

5.2.3. AI applications in learning vocabulary and grammar

One example of AI for vocabulary learning was a study by Chen and Li (2010), who developed a context-aware vocabulary learning system. The system could also suggest new words to be learned based on students' leisure time, i.e., new words would be suggested if the students had more time to learn. Results showed that the students who used the system with context awareness outperformed those who did not. In another study, Pandarova et al. (2019) developed an ITS for practicing English tenses. This system applied dynamic difficulty adaption to adjust the difficulty levels of grammar exercises. The results showed that the system could provide materials of appropriate difficulty levels and allow students to learn grammar at their own pace, making it conducive to effective learning.

5.2.4. AI applications in learning speaking and listening

AI was frequently used to facilitate speaking and listening. Ayedoun et al. (2019) developed a conversational agent to foster communication. The agent was designed based on communication strategies and affective backchannels. The learners could practice and improve their conversation skills by asking the AI agent questions which were then answered. In Johnson (2007), learners practiced speaking skills in games, i.e., the Mission and Arcade. While playing the Arcade Game, players are required to give spoken commands to move their avatars, and in the Mission game, the players speak on behalf of their avatars to complete their mission. ASR techniques were embedded in the games to enable learners to interact with the NPC to practice speaking and listening. The results showed most participants felt the game helped them acquire functional abilities in the target language.

5.3. Advantages of using AI in language learning

5.3.1. Providing personalized learning experiences

AI can suggest appropriate content for learners according to their level, needs, and preferences with advanced algorithms. Pandarova's et al. (2019) system could adjust the difficulty of grammar learning content according to students' language abilities, which allowed students to learn at their own pace, optimizing the learning outcomes. Similarly, Chen et al. (2006) designed a PIMS to enhance reading development. The system recommended English news articles based on a learner's language proficiency. Results showed that using this personalized system for facilitating students' reading was effective as it reduced cognitive overload by aligning articles with students' level of competence. In Chao et al. (2012), the Affective Tutoring System recommended lessons based on learners' emotional state. The system monitored students' moods and customized learning materials to help students avoid learning anxiety. If the system detected negative emotions, it provided relatively easier learning tasks. In this way, students' self-confidence was enhanced, thereby encouraging them to learn.

5.3.2. Enabling immediate adjustment

AI enabled language learners to adjust their learning after receiving automated feedback. As discussed, the NLP techniques used in the AWE systems can detect errors and provide learners with rich feedback, which allows them to take immediate action. For example, the Bengali Handwriting Education System used in Khatun and Miwa (2016) recognized learners' errors such as stroke production errors and stroke sequence errors. Students received timely feedback and made immediate adjustments using this system. In this way, students' language proficiency could be enhanced by repeatedly making modifications and improving their work. As for the quality of feedback, Gierl et al. (2014) showed that AWE systems can provide rich formative feedback, which can overcome teachers' preference for summative feedback due to time constraints with large-sized classes. Gierl et al. (2014) offered students AI-based rich and individualized feedback, enabling them to adjust their learning behavior during their learning process instead of at the final stage.

5.3.3. Rich opportunities using AI in language learning

Using AI techniques, the limited opportunities to practice the target language can be resolved. ITS allows students to learn anywhere and anytime. Stockwell (2007) developed a mobile-based ITS that could record difficult words by presenting them more frequently to increase learning opportunities. Learners could also practice the target language by interacting with a digital human. Mirzaei et al. (2018) introduced Virtual Reality Conversation Envisioning for learners to interact with an AI agent in an immersive context under which simulated scenarios, e.g., bargaining and interviewing, could be created. Students had more opportunities to practice their speaking skills by conducting conversations in different contexts and had more frequent use of the language without going abroad.

5.4. Challenges using AI in language learning

5.4.1. Reliability of AI technology

Although we have discussed the effectiveness of applying AI in language learning in previous sections, its reliability remains a concern. Many researchers have expressed uncertainty about whether this technology is ready for use in the classroom. Grimes and Warschauer (2010) doubted the accuracy of AWE as it could not evaluate subjective features of natural languages. The computational semantic analysis mainly focuses on the denotative meanings of words, while the connotative meanings may not be fully captured. In such cases, the author's intent is unlikely to be evaluated by the system resulting in improper grading of essays. Similarly, Johnson (2007) noted the challenges of evaluating ASR accuracy. Since ASR performance varies across contexts, students may not have smooth interactions with the NPC. As the quality of interactions directly impacts students' learning effectiveness, uncertainty regarding quality can pose challenges for using AI to learn languages. More advanced AI technology is needed to address this problem, and system developers should extensively test their designs before launching new systems.

5.4.2. Acceptance by teachers and students

The uncertain effectiveness of using AI in language education is sometimes caused by teachers' and students' reluctance to use the technology. For example, students in Roscoe and McNamara's study (2013) complained that some feedback given by the writing system was confusing. As the quality of AI cannot be guaranteed, students and teachers may have little motivation to use it as prior negative experiences in using technology can discourage them. Lin et al. (2017) found that users who had little experience using e-learning tools had lower satisfaction with ITS and had negative perceptions of the system due to its differences from traditional technology. Such challenges were also noticed by Pokrivcakova (2019), who showed that a lack of experience with Information Communication Technology (ICT) resulted in teachers' reluctance to use AI-related technologies. Thus, the acceptance of instructors and learners could be improved by developing better AI-enhanced learning systems that provide better teaching and learning experiences and help build positive attitudes. Teacher training programs should also be conducted to help teachers understand the potential benefits of AI in language education.

5.4.3. Social issues of AI in language education

Discourse analysis conducted by AI may be biased if the data and algorithms used for training contain societal biases (Yang et al., 2021). Algorithms may include unbalanced and disproportionate information (Luan et al., 2020) which could lead to social inequities or social cohesion. Further, as some developing countries cannot afford basic ICTs, they may be unlikely to adopt newly developed AI-based technologies possibly leading to a more expansive digital divide and contributing to educational inequality (Luan et al., 2020). Hwang et al. (2020) and Zhang and Aslan (2021) have also suggested that AIEd ethics be developed to address privacy issues from all stakeholders. For example, principles and ethical codes could be established before using AI to avoid leaking personal information. Researchers need to screen out biased information or set up keywords to filter sensitive information when selecting resources for natural language processing. International organizations could also support developing countries by providing essential communication technologies (Luan et al., 2020).

Putting humans at the center of AI applications is an important consideration. AI needs to be shifted from technology-oriented applications, which emphasize the development of production and performance, to human-oriented ones, which accentuate the integration of human and machine intelligence (Yang, 2021; Yang et al., 2021). Yang et al. (2021) called this new trend of AI, Human-centered AI (HAI), suggesting that some of the limitations of AIEd can be solved by HAI. HAI algorithms such as Bidirectional Encoder Representations from Transformers and Generative Pre-Training can be adopted for natural language processing to achieve performances close to those of humans (Yang et al., 2021). This can help increase the accuracy of grading on student writing. HAI also allows researchers to understand users' perceptions and requirements when using AI-enhanced language tools (e.g., translation applications and voice assistants). It can identify students' motivations and engagement and provide them with timely assistance and intervention during the learning process, which is essential for effective language learning (Huang et al., 2020).

6. Conclusion

The present review provides comprehensive coverage AI research trends in language education by analyzing publications from 2000 to 2019. According to our results, the number of articles related to AI in language education showed an increasing trend over the period reflecting researchers' growing interest in using AI tools to assist language learning. Notably, an increasing number of new journals on AI, such as *Computers & Education*, *Artificial Intelligence*, *International Journal of Learning Analytics* and *Artificial Intelligence for Education*, and *IJAIED* emerged during the period. *IJAIED* was the most influential journal, and the USA and the Arizona State University were the country and institution that contributed the most research. We also found that AWE is the most investigated AI application, and its interest grew over the years.

As for the limitations of our review, because it was limited to articles found in only three sources (i.e., WoS, ERIC, and Scopus), not all academic research related to AI in language education was included. Thus, future research may consider including more sources to provide a more comprehensive analysis. Regarding the citation count, the data retrieved from Google Scholar might have included citations from non-academic resources, which may have led to multiple counts when the publications were released on different platforms. Future research may consider using other approaches for citation counting. Additionally, the research methodology applied in the current research was bibliometric; future research may apply different methods to further investigate the literature on AI in language education from other perspectives. Other suggestions for future research on AI in language education include Yang et al. (2021), who recommended investigating AI's potential for improving teaching and learning outcomes, and Hwang et al. (2020) who also suggested that future research investigate the possibility of using AI for language courses.

Acknowledgement

An abstract entitled "Artificial Intelligence in Language Education" based on this paper was presented at the International Conference on Education and Artificial Intelligence 2020, The Education University of Hong Kong, 9-11 November 2020, Hong Kong. Gary Cheng's work in this research is supported by the Research Cluster Fund (RG 78/2019-2020R) of The Education University of Hong Kong and the Dean's Research Fund 2019/20 (IDS-2 2020) of The Education University of Hong Kong. Haoran Xie's work in this research is supported by the Faculty Research Fund (DB21A9) and the Lam Woo Research Fund (LWI20011) of Lingnan University, Hong Kong.

References

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A Large-scale learner corpus analysis employing natural language processing techniques. *Language Learning,* 67(S1), 180-208.

Ali, Z. (2020). Artificial Intelligence (AI): A Review of its uses in language teaching and learning. *IOP Conference Series: Materials Science and Engineering, 769*(1), 012043. https://doi.org/10.1088/1757-899x/769/1/012043

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, *29*(1), 29-57.

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2010). Collaborative writing support tools on the cloud. *IEEE Transactions on Learning Technologies*, *4*(1), 88-97.

Chao, C. J., Lin, H. K., Huang, T. C., Hsu, K. C. & Hsieh, C. Y. (2012). The Application of affective tutoring systems (ATS) in enhancing learners' motivation. In *Workshop Proceedings of the 20th International Conference on Computers in Education* (ICCE) (pp. 58-66). Asia-Pacific Society for Computers in Education.

Chen, C. M., Hsu, S. H., Li, Y. L., & Peng, C. J. (2006). Personalized intelligent m-learning system for supporting effective English learning. In *2006 IEEE International Conference on Systems, Man and Cybernetics* (Vol. 6, pp. 4898-4903). IEEE. https://doi.org/10.1109/ICSMC.2006.385081

Chen, C. M., & Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments, 18*(4), 341-364.

Chen, M. P., Wang, L. C., Zou, D., Lin, S. Y., & Xie, H. (2019). Effects of caption and gender on junior high students' EFL learning from iMap-enhanced contextualized learning. *Computers & Education, 140,* 103602. https://doi.org/10.1016/j.compedu.2019.103602

Chen, X., Hao, J., Chen, J., Hua, S., & Hao, T. (2018). A bibliometric analysis of the research status of the technology enhanced language learning. In *International Symposium on Emerging Technologies for Education* (pp. 169-179). Springer, Cham.

Chen, X., Xie, H., & Hwang, G. J. (2020a). A Multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence, 1*, 100005. https://doi.org/10.1016/j.caeai.2020.100005

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020b). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chen, X., Zou, D., & Xie, H. (2020c). Fifty years of British Journal of Educational Technology: A Topic modeling based bibliometric perspective. *British Journal of Educational Technology, 51*(3), 692-708.

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020d). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A Retrospective of all volumes of computer & education. *Computers & Education*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Chen, X., Zou, D., Xie, H., & Cheng, G. (2021a). Twenty years of personalized language learning. *Educational Technology & Society*, *24*(1), 205-222.

Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021b). Past, present, and future of smart learning: A Topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, *18*(1), 1-29.

Chen, X., Zou, D., Xie, H. R., & Su, F. (2021c). Twenty-five years of computer-assisted language learning: A Topic modeling analysis. *Language Learning & Technology, 25*(3), 151-185.

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of Artificial Intelligence in education. *Educational Technology & Society, 25*(1), 28-47.

Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. (2018). A Systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI, 5*, 37. https://doi.org/10.3389/frobt.2018.00037

Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2022). A Review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 1-43. https://doi.org/10.1080/09588221.2022.2033787

Gamper, J., & Knapp, J. (2002). A Review of intelligent CALL systems. *Computer Assisted Language Learning*, *15*(4), 329-342.

Gierl, M. J., Latifi, S., Lai, H., Boulais, A. P., & De Champlain, A. (2014). Automated essay scoring and the future of educational assessment in medical education. *Medical Education, 48*(10), 950-962.

Gong, Y., Lyu, B., & Gao, X. (2018). Research on teaching Chinese as a second or foreign language in and outside mainland China: A Bibliometric analysis. *The Asia-Pacific Education Researcher*, *27*(4), 277-289.

Grajzl, P., & Murrell, P. (2019). Toward understanding 17th century English culture: A Structural topic model of Francis Bacon's ideas. *Journal of Comparative Economics, 47*(1), 111-135.

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), 1-44. https://ejournals.bc.edu/index.php/jtla/article/view/1625

Heil, C. R., Wu, J. S., Lee, J. J., & Schmidt, T. (2016). A Review of mobile language learning applications: Trends, challenges, and opportunities. *EuroCALL Review, 24*(2), 32–50. https://doi.org/10.4995/eurocall.2016.6402

Hinojo-Lucena, F. J., Aznar-Díaz, I., Cáceres-Reche, M. P., & Romero-Rodríguez, J. M. (2019). Artificial intelligence in higher education: A Bibliometric study on its impact in the scientific literature. *Education Sciences*, *9*(1), 1-9. https://doi.org/10.3390/educsci9010051

Hirsch, J. E. & Buela-Casal, G. (2014). The Meaning of the H-index. *International Journal of Clinical and Health Psychology, 14*(2), 161-164.

Huang, A. Y., Lu, O. H., Huang, J. C., Yin, C. J., & Yang, S. J. (2020). Predicting students' academic performance by using educational big data and learning analytics: Evaluation of classification methods and learning logs. *Interactive Learning Environments*, *28*(2), 206-230.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1,* 100001. https://doi.org/10.1016/j.caeai.2020.100001

Johnson, A. M., Guerrero, T. A., Tighe, E. L., & McNamara, D. S. (2017). iSTART-ALL: Confronting adult low literacy with intelligent tutoring for reading comprehension. In *International Conference on Artificial Intelligence in Education* (pp. 125-136). Springer. https://doi.org/10.1007/978-3-319-61425-0_1

Johnson, W. L. (2007). Serious use of a serious game for language learning. *Frontiers in Artificial Intelligence and Applications*, *158*, 67-74.

Johnson, W. L., Vilhjálmsson, H. H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI? In *Artificial Intelligence in Education* (pp. 306-313). IOS Press.

Khatun, N., & Miwa, J. (2016). An Autonomous learning system of Bengali characters using web-based intelligent handwriting recognition. *Journal of Education and Learning*, *5*(3), 122-138.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333-349.

Lee, K., Kwon, O. W., Kim, Y. K., & Lee, Y. (2015). A Hybrid approach for correcting grammatical errors. In F. Helm, L. Bradley, M. Guarda, & S. Thouesny (Eds.), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 362-367). Research-publishing.net. https://doi.org/10.14705/rpnet.2015.000359

Lin, C. C., Liu, G. Z., & Wang, T. I. (2017). Development and usability test of an e-learning tool for engineering graduates to develop academic writing in English: A Case study. *Educational Technology & Society, 20*(4), 148-161.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C. C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.580820

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*(2), 499-515.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A Hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35-59.

Mirzaei, M. S., Zhang, Q., van der Struijk, S., & Nishida, T. (2018). Language learning through conversation envisioning in virtual reality: A Sociocultural approach. In P. Taalas, J. Jalkanen, & S. Thouësny (Eds.), *Future-Proof CALL: Language Learning as Exploration and Encounters-Short Papers from EUROCALL* (pp. 207-213). http://doi.org/10.14705/rpnet.2018.26.838

Organisation for Economic Co-operation and Developmen (OECD). (2019). *Artificial Intelligence in Society*. OECD Publishing. https://dx.doi.org/10.1787/eedfee77-en

Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R. D., & Brefeld, U. (2019). Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education*, *29*(3), 342-367.

Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, *7*(3), 135-153.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064-1082. https://doi.org/10.1111/ajps.12103

Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, *105*(4), 1010-1025.

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), 524. https://doi.org/10.3390/su12020524

Song, P., & Wang, X. (2020). A Bibliometric analysis of worldwide educational artificial intelligence research development in recent twenty years. *Asia Pacific Education Review*, *21*(3), 473-486.

Stockwell, G. (2007). Vocabulary on the move: Investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning*, *20*(4), 365-383.

Su, F., & Zou, D. (2020). Technology-enhanced collaborative language learning: theoretical foundations, technologies, and implications. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1831545

Svensson, G. (2010). SSCI and its impact factors: A "Prisoner's dilemma"? *European Journal of Marketing, 44*(1/2), 23–33.

Tran, B. X., Latkin, C. A., Vu, G. T., Nguyen, H. L. T., Nghiem, S., Tan, M. X., Lim, Z.-K., Ho, C. S. H., & Ho, R. (2019). The Current research landscape of the application of Artificial Intelligence in managing cerebrovascular and heart diseases: A Bibliometric and content analysis. *International Journal of Environmental Research and Public Health*, *16*(15), 2699. https://doi.org/10.3390/ijerph16152699

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., Van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A Review. *Review of Educational Research*, *89*(2), 259-295.

Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education*, *28*(1), 79-105.

Wang, C. P., Lan, Y. J., Tseng, W. T., Lin, Y. T. R., & Gupta, K. C. L. (2019). On the effects of 3D virtual worlds in language learning–A Meta-analysis. *Computer Assisted Language Learning*, 1-25.

Wang, F., & Preininger, A. (2019). AI in health: State of the art, challenges, and future directions. *Yearbook of medical informatics*, *28*(01), 16-26. https://doi.org/10.1055/s-0039-1677908

Wijekumar, K. K., Meyer, B. J., & Lei, P. (2017). Web-based text structure strategy instruction improves seventh graders' content area reading comprehension. *Journal of Educational Psychology*, *109*(6), 741-760.

Yang, S. J. (2021). Guest Editorial: Precision education-A New challenge for AI in education. *Educational Technology & Society*, *24*(1), 105-108.

Yang, S. J., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence, 2,* 100008. https://doi.org/10.1016/j.caeai.2021.100008

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhang, R., & Zou, D. (2020). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, 1-47. https://doi.org/10.1080/09588221.2020.1744666

Zou, D., Huang, Y., & Xie, H. (2019). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning*, *34*(5-6), 751-777. https://doi.org/10.1080/09588221.2019.1640745

Zou, D., Luo, S., Xie, H., & Hwang, G. J. (2020). A Systematic review of research on flipped language classrooms: theoretical foundations, learning activities, tools and research topics and findings. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1839502

Zou, D., Xie, H., & Wang, F. L. (2018). Future trends and research issues of technology-enhanced language learning: A Technological perspective. *Knowledge Management & E-Learning: An International Journal*, *10*(4), 426-440.