Applied Research on English Language

V. 15 N. 4 2026
pp: 201-224
http://are.ui.ac.ir

DOI: 10.22108/are.2026.149375.2794
Document Type: Research Article

Mapping the Research Landscape of Artificial Intelligence Applications in English Speaking Assessment: A Bibliometric Analysis from 2015 to 2025

Weng Weifeng

Universiti Utara Malaysia, Malaysia

Received: 2026/07/17 **Accepted: 2026/08/24**

Abstract: This study carries out a bibliometric analysis of research on artificial intelligence (AI) applications in English speaking assessments published between 2015 and 2025. Drawing on 1,448 Scopus-indexed publications and using VOSviewer for analysis, the study examines publication trends, country/author contributions, keyword co-occurrence, citation networks, and thematic developments. Research output increased markedly after 2023, driven by rapid advances in deep learning-based speech processing, the wider adoption of automated speech evaluation technologies, and the growing demand for scalable digital assessment systems capable of supporting learning in increasingly technology-mediated educational environments. The US, China, the UK, and Japan are the most productive countries. They also exhibit the highest betweenness centrality in the international collaboration network. Thematic analysis shows continued fragmentation in this field, with unsolved challenges including construct validity, ethical issues, and learner-centered factors. The analysis reveals a rapid expansion of research after 2023, identifies the most productive countries, institutions, authors, and journals, and maps the evolution of research themes from automated speech recognition and pronunciation assessment to generative AI-supported speaking assessment. These findings establish the current knowledge structure of the field and identify emerging research directions that can inform the development, implementation, and evaluation of AI-supported English speaking assessment. This study contributes by delivering a systematic overview of the field and identifying key directions for future research on valid, transparent, and pedagogically meaningful AI-assisted speaking assessment systems.

Keywords: Centralization, English Language Teacher Education (ELTE), Historical Comparative Education, Artificial Intelligence, Spoken English, Automated Assessment, Autonomous Learning, Bibliometric Analysis.

* **Corresponding Author.**

Authors' Email Address:
[1] **Weng Weifeng** (568390715@qq.com)

2322-5343 © University of Isfahan
This is an open access article under the CC BY-NC-ND/4.0/ License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Speaking proficiency is widely recognized as a core component of English language competence, especially in academic, professional, and international settings (Ghafar & Raheem, 2023). Unlike receptive skills, speaking demands the real-time integration of linguistic knowledge, cognitive processing, and sociopragmatic awareness, making it one of the most complex language skills to acquire and assess (El Majidi et al., 2024; Leonard et al., 2017; Derakhshan et al., 2016). Speaking ability in an English as a Foreign Language (EFL) setting is usually considered one of the most important measures of communicative competence, as it affects learners' academic success, employability, and social participation (Nghia et al., 2024). Consequently, the accurate and fair assessment of speaking performance has long been a central concern in language testing research and practice (Fulcher, 2014; Kunnan, 2013; Xi, 2010).

Despite sustained methodological refinement, traditional speaking assessment continues to encounter methodological and practical challenges. Human-rated measures are resource-intensive, time-consuming, and cannot be scaled, especially in large-scale or high-stakes testing situations (Hommel, 2023; Chou, 2018). Furthermore, empirical studies over the past decade have consistently reported issues related to rater subjectivity, score variability, and rater bias, even when analytic rubrics and standardized rating procedures are employed (Li et al., 2025; Liu et al., 2025). Concerns regarding inter-rater reliability and score comparability continue to challenge the fairness and validity in speaking assessment practices (Karatay & Xu, 2025; Rajan et al., 2022). These challenges have prompted continued exploration of assessment approaches that can enhance consistency and efficiency while maintaining construct integrity.

In this context, the rapid development of artificial intelligence (AI) technologies has reshaped the landscape of language learning and assessment. Developments in automatic speech recognition (ASR), natural language processing (NLP), and machine learning (ML) have enabled the automated analysis of spoken language features, including fluency, pronunciation, lexical complexity, and discourse coherence (Zechner & Evanini, 2020). Since the mid-2010s, AI-driven speaking assessment systems have been adopted in both formative and summative contexts. These systems provide more immediate revised comments and more scalable scoring procedures than are typically feasible in human-only assessment models, contributing to increased scholarly interest across the fields of applied linguistics and educational technology (Vetrivel et al., 2025).

However, the burgeoning expansion of AI applications in English speaking assessment

has also generated new theoretical, methodological, and ethical questions. Scholars have raised concerns regarding construct underrepresentation, algorithmic transparency, and the alignment between automated scores and theoretically grounded models of speaking ability (Andersen et al., 2025; Chen, 2025). Research in this field has proliferated across disciplines, including applied linguistics, educational technology, computer science, and speech processing, resulting in increasingly fragmented literature. Despite increasing interest, individual empirical studies and narrative reviews offer valuable insights, but there remains a lack of systematic, large-scale mapping of research trends, intellectual structures, and thematic evolution of AI-assisted English speaking assessment using large-scale bibliometric methods (Xing & Saeed, 2026; Richardson & Clesham, 2021).

The purpose of this study is to conduct a bibliometric analysis of scholarly publications on the application of AI technologies in English speaking assessment from 2015 to 2025. Specifically, the study aims to map the growth trajectory of this research area, identify key contributors and outlets, and uncover thematic and methodological trends that have shaped the field over the past decade. By synthesizing a large body of literature, this study seeks to provide both researchers and practitioners with a clearer understanding of the knowledge structure and research directions in AI-assisted speaking assessment.

Guided by this purpose, the study considers the following four research questions:

RQ1: How has research on AI-assisted English speaking assessment evolved between 2015 and 2025?

RQ2: Which countries, authors, and journals have shaped the intellectual structure of AI-assisted English speaking assessment research?

RQ3: What are the dominant research themes and technological approaches identified through keyword and co-citation analyses?

RQ4: What emerging trends and potential research gaps can be identified to inform future studies on AI-based English speaking assessment?

Literature Review

Speaking Assessment in Language Education

Speaking proficiency occupies a central position within models of communicative language ability, requiring the simultaneous deployment of phonological, lexical, grammatical, and pragmatic resources under real-time processing constraints (Fulcher, 2014; Bachman & Palmer, 2010). Assessment of this multifaceted construct has conventionally relied on the interaction-based procedures involving oral proficiency interviews and discussion tasks

scored by human raters using either analytic or holistic rubrics (Luoma, 2004). While these approaches show a good deal of face validity, their use is constrained by various limitations well-established in the literature. Chief among these are the resource-intensive nature of test administration, inconsistencies arising from rater subjectivity, and score variability attributable to differences in rater severity, task difficulty, and interlocutor effects (Kunnan, 2013; Xi, 2010).

Several empirical studies over the past decade have shown that reliability across multiple raters continues to be problematic even with the application of a structured rubric and rater training procedures (Li et al., 2025; Liu et al., 2025). Karatay and Xu (2025) highlight that rater cognitive load and rater fatigue negatively influence scoring reliability in oral language proficiency assessment tests. These continuing issues have led to attempts to find more efficient assessment procedures that would ensure standardization, lower costs, and scalability without loss of construct validity (Hommel, 2023; Chou, 2018). These methodological limitations have increased interest in AI-assisted assessment.

AI Technologies in Language Assessment through a Pedagogical Lens

AI-assisted spoken language assessment has evolved through several developmental stages. Earlier automated speech score estimation tools, exemplified by Phone Pass (later Versant) and Speech Rater by ETS, utilized primarily acoustic-phonetic cues and relatively basic NLP to generate proficiency scores (Bernstein et al., 2023; Zechner & Evanini, 2020). While demonstrating reasonable correlations with human scores on limited response tasks, these technologies were somewhat restricted in their ability to assess more complex constructs, such as extended discourse, pragmatic appropriateness, or interactivity.

Due to the breakthroughs in deep learning techniques in the mid-2010s, especially the emergence of recurrent neural networks and transformer models, the scope of spoken language characteristics that can be automatically assessed has been greatly expanded (Chen, 2025). Modern AI-powered assessment systems can analyze multiple features of speech at once, such as fluency markers like speech rate and pauses, segmental and suprasegmental pronunciation accuracy, vocabulary range, syntax complexity, and discourse cohesion (Kang & Johnson, 2018; Vetrivel et al., 2025). Such technologies enable the use of AI for formative purposes, providing timely feedback to learners, and for summative assessment.

The emergence of large language models and generative AI chatbots has also contributed to a change in the field. Fathi et al. (2024) state that the use of AI for conversation significantly improves EFL learners' performance in terms of speaking skills

and willingness to speak. Likewise, Tai and Chen (2024) stress that generative AI chatbots positively contribute to speaking at the elementary level of proficiency when used in individual or pair interactions. Overall, the use of AI in the field is shifting from mere scoring instruments to interactive agents that can elicit and facilitate spoken output.

However, the inclusion of AI in the process of speaking assessment has also caused many theoretical and ethical problems. For example, Bulut et al. (2024) point out the problem of conflict between algorithmic efficiency and construct validity because AI scoring models generally focus on quantitative characteristics of speech while neglecting its interactive, sociopragmatic, and strategic aspects. Andersen et al. (2025) questioned algorithmic fairness by demonstrating that there are certain biases encoded in automatic scoring models connected with accent, dialect, or linguistic background. At the same time, issues related to data privacy, informed consent, and scoring algorithms remain unexplored in current literature despite their significance for high-stakes contexts (Richardson & Clesham, 2021).

Hybrid Assessment Models and Learner-Centered Perspectives

Considering the limitations inherent in automated approaches, modern literature has developed mixed evaluation frameworks in which the advantages of machine learning in terms of scalability and consistency coexist with the expertise of human raters, who employ their judgment (Järvelä et al., 2025; Xi, 2023). In such settings, AI works as a device but not as a replacement for humans, which can perform routine tasks, detect irregularities in students' performance, and provide initial feedback on the assessments. Human-AI co-evaluation is based on the recognition of the fact that the construct of speaking competence includes aspects that still lie out of reach for automated methods of analysis, particularly those related to negotiation of meaning and sociolinguistic and cultural appropriateness.

Parallel to technological developments, a growing body of research has examined learner-centered factors that mediate the effectiveness of AI-assisted assessment. El Shazly (2021) revealed that AI-mediated speaking practice significantly reduced foreign language anxiety while enhancing self-perceived speaking performance. These affective dimensions are consequential because anxiety and willingness to communicate directly influence both test performance and the validity of score interpretations. Despite their importance, learner variables have been neglected in the research on AI-assisted assessments, as reflected in their relatively low representation in keyword co-occurrence networks (Xing & Saeed, 2026).

Research Gap and Rationale

Despite the proliferating body of research on AI-assisted English speaking assessment, no comprehensive bibliometric study has been undertaken to map the field's intellectual structure and developmental trajectory systematically. Previous studies conducted a narrative review based on particular subtopics (Richardson & Clesham, 2021) and used systematic review techniques focused more on pedagogy than on the structure of knowledge (Xing & Saeed, 2026). An absence of a thorough and extensive bibliometric analysis prevents answering many key questions regarding collaboration trends, development of themes, fragmentation of knowledge, and emerging research frontiers. This becomes particularly important given the recent rapid development of the field after 2023 and its increasing interdisciplinary character, including applied linguistics, educational technology, computer science, and speech engineering. This study addresses this gap by conducting a bibliometric analysis of Scopus articles published from 2015 to 2025.

Materials and Methods

Research Design

This study adopts a bibliometric research design to systematically map the intellectual structure, developmental trajectory, and thematic evolution of scholarship on artificial intelligence (AI) applications in English speaking assessment between 2015 and 2025. By quantitatively analyzing publication patterns, citation networks, and keyword co-occurrences, bibliometric methods enable researchers to identify influential works, dominant research themes, and emerging trends in English speaking assessment (Donthu et al., 2021; Aria & Cuccurullo, 2017). Applying this approach to the study of AI-based English speaking assessment provides a comprehensive overview of how the field has evolved.

Data Source

The Scopus database was selected as the primary data source in this study. Scopus indexes a larger number of journals than other alternative databases and provides more extensive coverage of recent and interdisciplinary publications. To date, the Scopus database has been more widely adopted in social science research (Shaheen, 2025; Singh et al., 2021). It also offers one of the most comprehensive and authoritative collections of searchable materials, covering a broad range of publication titles, authors, abstracts, document types, affiliations, keywords, and citation counts (Pranckutė, 2021; Baas et al., 2020). This database ensures a

comprehensive capture of relevant literature while minimizing the risk of omitting key studies published in diverse outlets.

Inclusion Criteria

This study examined scholarly journal publications on AI applications in English speaking assessment published between 2015 and 2025, with the aim of providing a comprehensive overview of recent research trends in this field. The inclusion criteria were as follows:

a) Articles published in English,

b) Articles published between 2015 and 2025,

c) Articles focusing on AI applications in English speaking assessment,

d) Articles published in Arts and Humanities and Computer Science.

Data Collection

A comprehensive search strategy was developed to explore research on English speaking assessment. The search focused on two primary keywords, “English speaking assessment” and “English speaking evaluation,” along with related terms. This step was necessary because researchers do not always use the same terms, such as “spoken English assessment” and “English speaking test,” which are often used to describe the same concept.

A systematic screening was conducted to identify the most appropriate keywords for searching for articles. The search query was confined as follows:

TITLE-ABS-KEY (English speaking assessment) OR TITLE-ABS-KEY (English speaking evaluation) OR TITLE-ABS-KEY (spoken English assessment) OR TITLE-ABS-KEY (English speaking test)) AND PUBYEAR > 2014 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "ARTS") OR LIMIT-TO (SUBJAREA, "COMP") OR EXCLUDE (SUBJAREA, "NURS") OR EXCLUDE (SUBJAREA, "MATH") OR EXCLUDE (SUBJAREA, "MEDI") OR EXCLUDE (SUBJAREA, "NEUR") OR EXCLUDE (SUBJAREA, "IMMU") OR EXCLUDE (SUBJAREA, "PHAR") OR EXCLUDE (SUBJAREA, "CHEM") OR EXCLUDE (SUBJAREA, "PHYS") OR EXCLUDE (SUBJAREA, "ENVI") OR EXCLUDE (SUBJAREA, "ECON") OR EXCLUDE (SUBJAREA, "BUSI") OR EXCLUDE (SUBJAREA, "ENGI") OR EXCLUDE (SUBJAREA, "HEAL") OR EXCLUDE (SUBJAREA, "ENER") OR EXCLUDE (SUBJAREA, "BIOC") OR EXCLUDE (SUBJAREA, "CENG") OR EXCLUDE (SUBJAREA, "DECI") OR EXCLUDE (SUBJAREA,

"PSYC") OR EXCLUDE (SUBJAREA, "DENT") OR EXCLUDE (SUBJAREA, "VETE") OR EXCLUDE (SUBJAREA, "AGRI") OR EXCLUDE (SUBJAREA, "EART")) AND (LIMIT-TO (LANGUAGE, "English")) AND (EXCLUDE (EXACTKEYWORD, "Healthcare Communication"))

To identify as many relevant studies as possible and to conduct a comprehensive yet targeted search for English speaking assessments, the keywords were combined using the Boolean operators “OR” and “AND”.

Figure 1 visually illustrates the entire process, from identifying the study’s main focus to creating the final visual map via bibliometric analysis. The first search process, the data retrieval, was conducted on 04 January 2026 and uncovered 1,528 documents from the Scopus database. During the screening process, the results were filtered to include only English-language publications. To validate the reliability and academic value of the results obtained in this research, only peer-reviewed sources were used. As a result, only articles are included in this study. Conference papers, reviews, books, and book chapters were excluded. These subject areas include but are not limited to Computer Science, and Arts and Humanities. Ultimately, a total of 1,448 documents were selected for the bibliometric analysis.

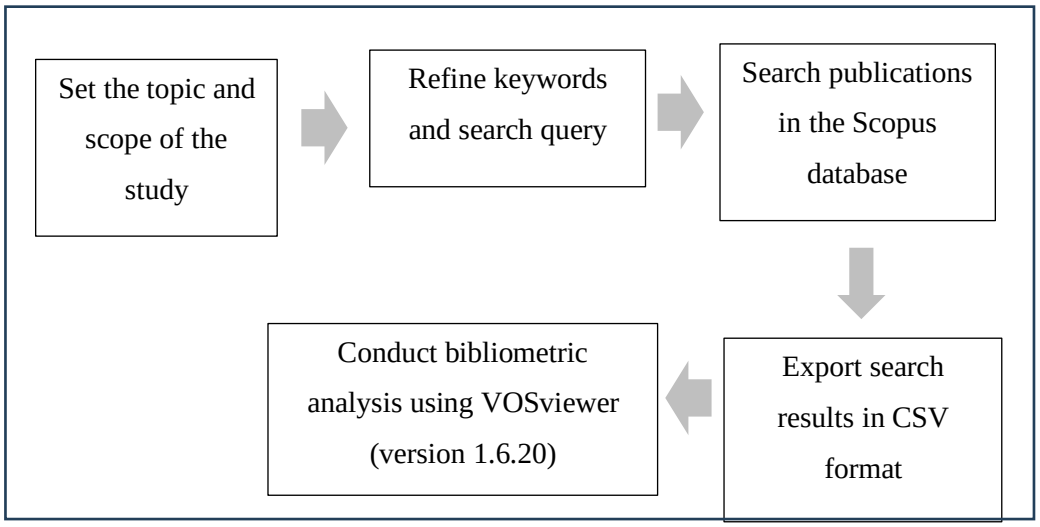


Figure 1. The Brief Process of Conducting a Bibliometric Analysis

Data Analysis

The study adopts bibliometric analysis as its main method to study research publications of English speaking assessments. The aim is to gain a broad overview of the development in this field during the period from 2015 to 2025, including the trajectory of publication output, the

most-cited countries, international collaboration, leading journals and notable articles, the co-occurrence network of keywords, etc. The bibliographic information of 1,448 documents was extracted from the Scopus database and saved in an Excel file for further analysis. It includes document title, publication year, author name, author affiliation, keywords, abstract, source title, publisher, and various citation information. The Scopus database allows for the analysis and visualization of descriptive statistics.

VOSviewer is a robust software used for constructing and visualizing bibliometric networks. In particular, it handles large data sets efficiently and offers different types of visualizations, providing extensive analyses (Van Eck & Waltman, 2014). The study adopted three major bibliometric techniques: co-authorship analysis, keyword co-occurrence analysis, and citation analysis. Co-authorship analysis was used to understand patterns of contribution and international collaboration. Keyword co-occurrence analysis, also known as co-word analysis, was used to identify the most frequently used words in article titles, author keywords, and abstracts (Zupic & Cater, 2015). Citation analysis was used to determine the number of times the journals and documents in the dataset were cited in the Scopus database.

Results

To respond to the research questions presented in the preceding section, this study conducts a comprehensive analysis of a corpus of 1,448 documents obtained from the Scopus database. In total, 1,448 documents from 160 sources were included between 2015 and 2025.

Annual Production

The development of English speaking assessment research, as represented in Figure 2, is discussed in terms of the yearly distribution of published studies. The number of publications increases slightly between 2015 and 2023. Publication volume remained modest during the early period. Nevertheless, the number of scholarly publications surged dramatically between 2023 and 2025, coinciding with a broader range of technological advancements in deep learning and speech recognition. This period witnessed a shift toward data-driven modeling approaches, particularly AI capable of processing large-scale speech corpora.

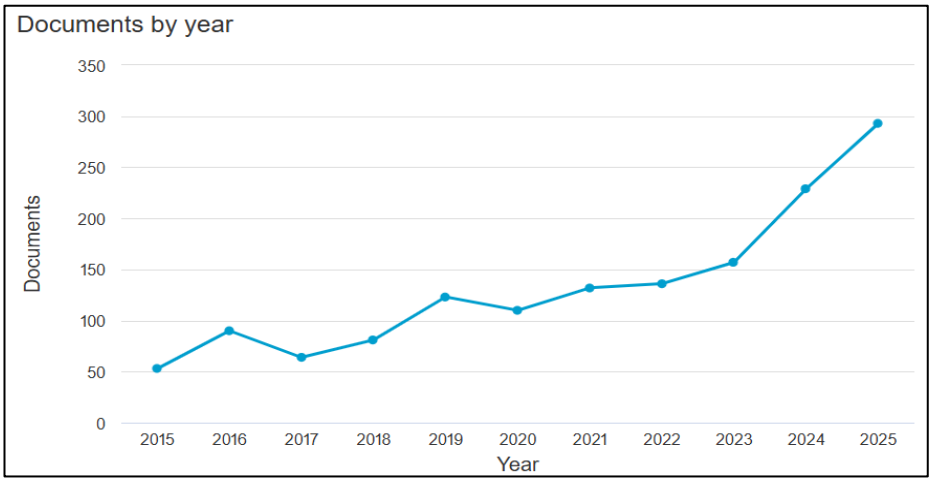


Figure 2. Annual Production of Articles (2015–2025)

Contributions and Collaborations by Country/Region

An overview of the global research contributions on AI applications in English speaking assessment practices reveals a variety of contributing nations, though it is evident that there is a difference in the number of publications as well as their citation influence. To date, 94 countries and regions have contributed to this research field. The top 20 countries and regions publishing articles in this field are shown in Figure 3. Countries and regions with big and bright circles were those with large numbers of publications. The United States has emerged as the most influential contributor, leading in both the number of publications and citations. China ranks second; the United Kingdom, Japan, Saudi Arabia, and Indonesia also demonstrate notable contributions to the publication output in this field.

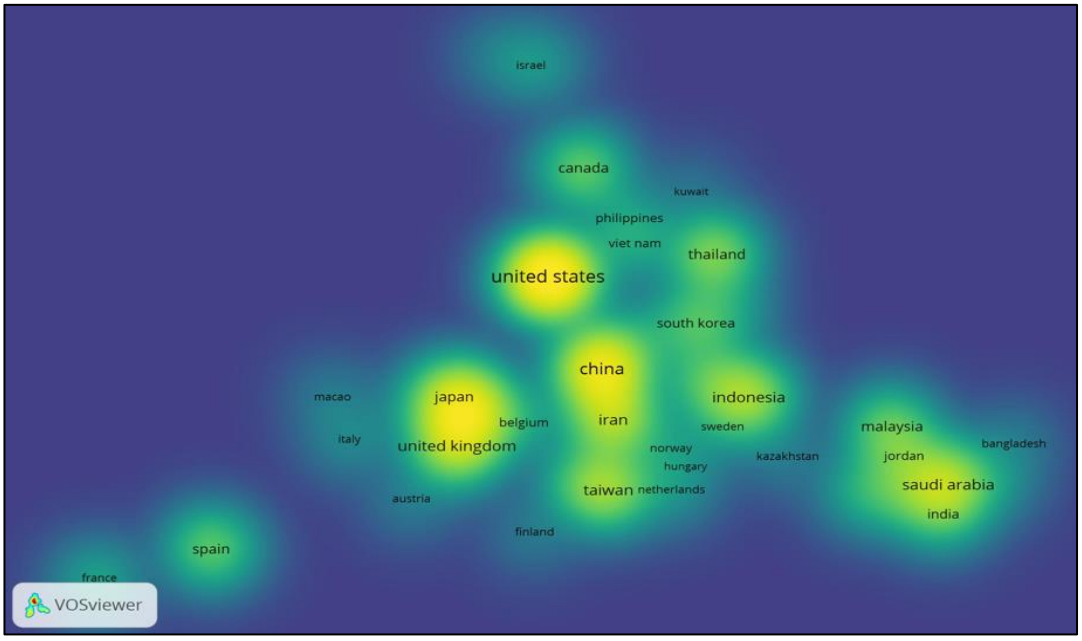


Figure 3. Density Map of the Top 20 Countries/Regions

Regarding country collaborations, the network diagram created by the VOSviewer software (version 1.6.20) reveals the international collaborations in the field of AI application in English speaking assessment, presenting the interconnectedness of 94 countries in cross-border research, indicating a high level of global academic interactions in this field. The visualization in Figure 4 depicts seven different clusters, distinguished by color, implying high international collaborations and information flows in different geographic areas.

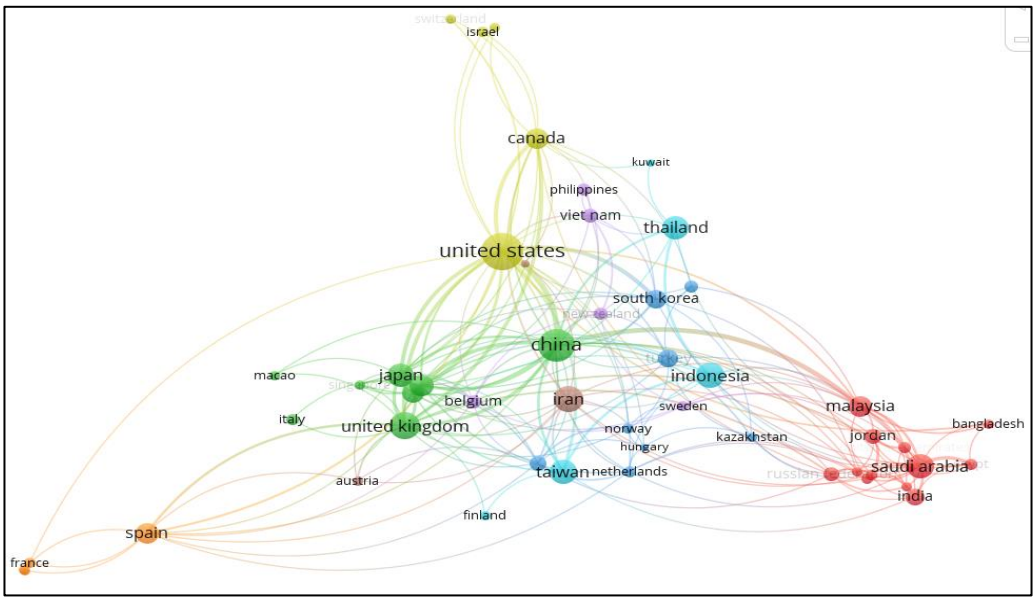


Figure 4. The Network Visualization Map of Country Collaboration

The United States plays a key role in the collaborative network of AI applications in English language assessment, having extensive connections with a total of 101 link strengths to various countries. The United Kingdom, China, Japan, and Saudi Arabia are also highly involved in this collaborative network.

Influential Authors and Documents

From the 1,448 documents analyzed, Figure 5 shows the top 10 most relevant authors having a major impact on AI application in English speaking assessment. The top 5 authors are Altakhaineh, A.R.M. and Trofimovich, P. with 10 articles, Shadiev, R. with 8 articles, Fathi, J., Waluyo, B. and Zou, B. with 7 articles.

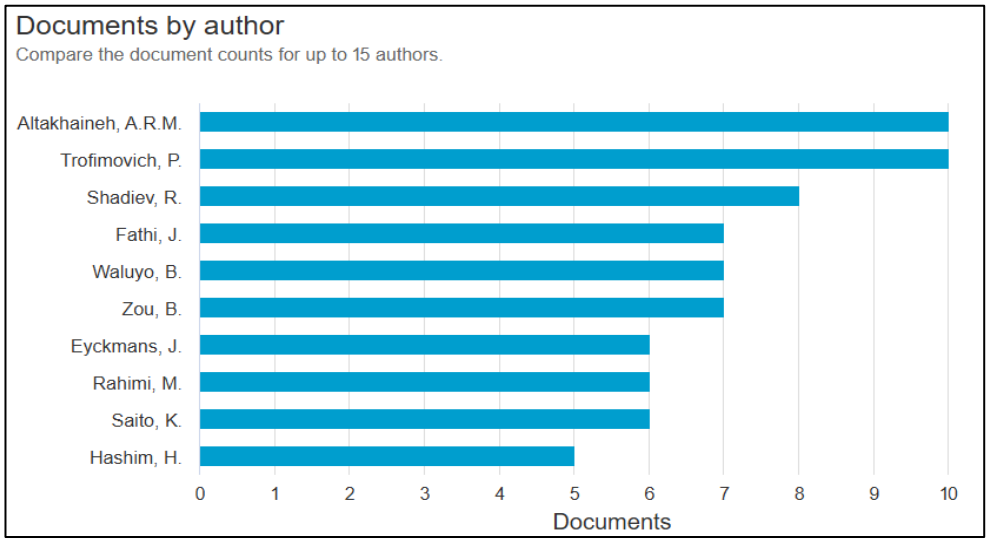


Figure 5. Most Productive Authors

Table 1 lists the top 10 most cited documents on AI application in English speaking assessment with information on author, title, year, citations, and source title. According to Table 1, *Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English speaking performance and learning perceptions* is the most widely cited document written by Chien S.-Y. as the first author, with 354 total citations. The second comes from Fathi J.'s article *Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions*, with 262 citations. The third one is *Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts,* by Hwang W.-Y as the first author, with 159 total citations.

Table 1. Most Highly Cited Publications on AI Applications in English Speaking Assessment (Scopus Database)

No	Author(s)	Title	Year	Citation	Source	Country
1	Chien S.-Y., et al.	Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English speaking performance and learning perceptions	2020	354	Computers and Education	Taiwan, China
2	Fathi J., et al.	Improving EFL learners' speaking skills and	2024	262	System	Iran

No	Author(s)	Title	Year	Citation	Source	Country
		willingness to communicate via artificial intelligence-mediated interactions				
3	Hwang W.-Y., et al.	Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts	2016	159	Computer Assisted Language Learning	Taiwan, China
4	Dizon G.	Evaluating intelligent personal assistants for L2 listening and speaking development	2020	149	Language Learning and Technology	Japan
5	Saito K.; Trofimovich P.; Isaacs T.	Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness: A Validation and Generalization Study	2017	145	Applied Linguistics	United Kingdom
6	Amiryousefi M.	The incorporation of flipped learning into conventional classes to enhance EFL learners' L2 speaking, L2 listening, and engagement	2019	116	Innovation in Language Learning and Teaching	Iran
7	Tai T.-Y., Chen H.H.-J.	Navigating elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions	2024	114	Computers and Education	Taiwan, China
8	Hwang W.-Y., et al.	Effects of storytelling to facilitate EFL speaking using a Web-based multimedia system	2016	113	Computer Assisted Language Learning	Taiwan, China

No	Author(s)	Title	Year	Citation	Source	Country
9	Chen M.-R.A., Hwang G.-J.	Effects of experiencing authentic contexts on English speaking performances, anxiety and motivation of EFL students with different cognitive styles	2022	96	Interactive Learning Environments	Japan
10	Crowther D., et al.	Does a speaking task affect second language comprehensibility?	2015	96	Modern Language Journal	United States

Most Relevant Sources and Publishers

Table 2 comprehensively presents the top 10 most relevant sources and publishers that have contributed to the development and growth of AI applications in English speaking assessments. The 1,448 articles have been published in 160 sources, and they mention the name of the source, the number of articles published, and the publishing company.

Table 2. Top 10 Publication Sources for Research on AI Applications in English Speaking Assessment

Sources	Articles	Publisher
Language Testing in Asia	45	Springer
Asian EFL Journal	42	Asia TEFL
System	42	Elsevier
Language Testing	39	SAGE Publications
Journal of Language Teaching and Research	37	Academy Publication
World Journal of English Language	37	Sciedu Press Quarterly
Theory and Practice in Language Studies	34	Academy Publication
Computer Assisted Language Learning	31	Taylor & Francis
Language Teaching Research	30	SAGE Publications
Sage Open	29	SAGE Publications

Most Cited Affiliations

This section is valuable for understanding the research output of various universities. It provides researchers with insights into the evolution of AI used in English speaking assessments. It provides opportunities for collaboration that will advance the development of this research area. Figure 6 represents the top 10 most cited institutions contributing to this field. They are the Chinese University of Hong Kong (20), Prince Sattam Bin Abdulaziz University (19), Concordia University (15), the University of Hong Kong, the University of Melbourne, and Universiti Kebangsaan Malaysia (14). It is worth noting that, after Europe and America, Asian countries/regions have the highest number of most-cited articles.

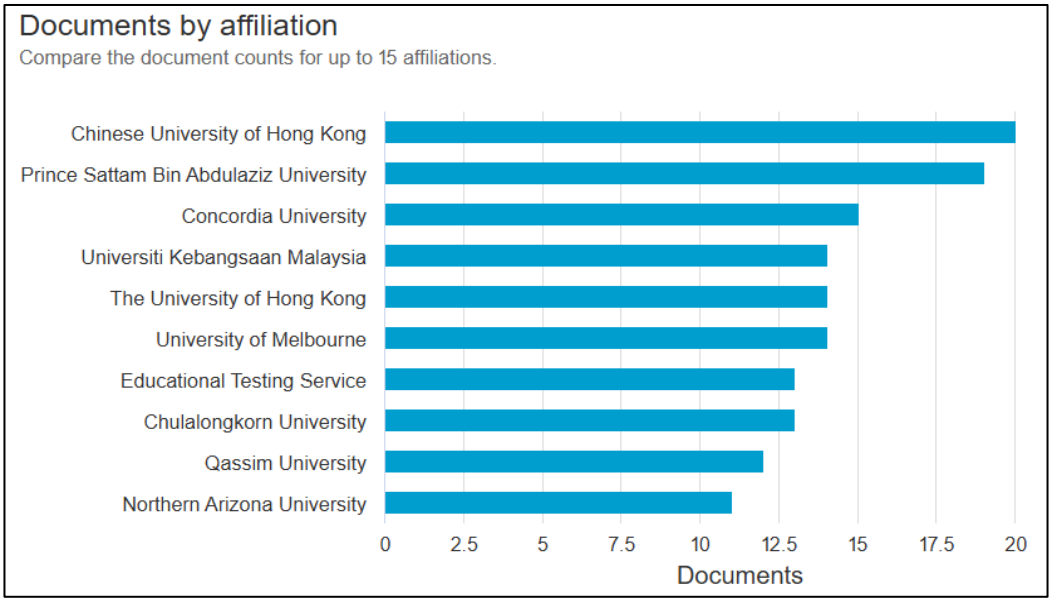


Figure 6. Most Cited Affiliations

Keyword Co-occurrence Network

Figure 7 represents these 11 clusters, comprising 163 items, 1,082 links, and a total link strength of 1,396. A minimum keyword occurrence threshold of 5 was applied. It depicts the clustering of the 163 co-occurring network keywords extracted from our dataset. The clusters are indicated by different colors and densities, offering an extensive overview of the connections and correlations between the keywords. The figure illustrates the main nodes revealing the topic of this research: speaking, speaking skills, assessment, EFL, fluency, etc.

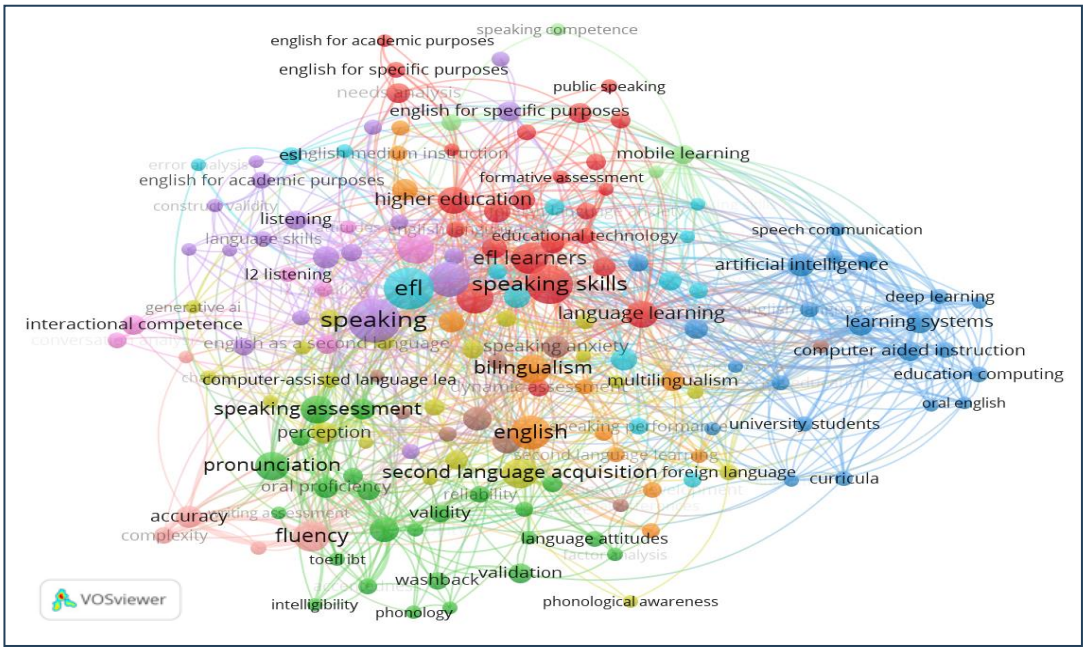


Figure 7. Keyword Co-occurrence Network on English Speaking Assessment

The overall keyword co-occurrence provides a broad perspective on this field. Table 3 presents the 10 most frequent keywords among the co-occurrence networks. These keywords, such as 'speaking (frequency = 61)', 'speaking skills (frequency = 46)', 'assessment (frequency = 38)', and 'English as a foreign language (frequency = 31)', highlight the relevance of language factors in the academic literature.

Table 3. The Top Ten Frequent Keywords on English Speaking Assessment in Scopus (2015–2025)

Ranking	Keywords	Frequency	Link Strengths
1	speaking	61	94
2	speaking skills	46	61
3	assessment	38	59
4	English as a foreign language	31	45
5	speaking skill	30	39
6	fluency	29	65
7	second language acquisition	28	31
8	pronunciation	26	46
9	bilingualism	25	22
10	speaking assessment	24	20

Simultaneously, the network also emphasizes several emerging and underexplored areas that require closer attention. In particular, AI and deep learning seem to receive comparatively limited focus in the context of English speaking assessment. This suggests that their potential has not been sufficiently explored in language assessment research.

Discussion

RQ1: Developmental Trajectory of Research (2015–2025)

The upward trajectory of publications indicates that AI-assisted speaking assessment has transitioned from a peripheral research interest to an increasingly established domain within educational technology and language assessment. The greatest growth in output occurs after 2023, when digital learning increased internationally. Publication output is expected to continue expanding given sustained technological development. This shift from initial exploration to accelerated development appears to be fueled by the convergence of AI technology and contextual factors. In particular, advances in deep learning-based automatic speech recognition (ASR) have enhanced the capacity for large-scale analysis of spoken language data (Chen, 2025). Simultaneously, the rapid transition to online learning during the COVID-19 pandemic intensified the demand for scalable and automated assessment solutions (Xi, 2023; Forrester, 2020).

The expansion of research in this field reflects not only technological innovation but also shifts in educational practices that emphasize flexibility, scalability, and remote accessibility. Notably, these developments do not indicate the replacement of human evaluators; rather, they point to a shift of language assessment practices toward hybrid models in which AI systems supplement human evaluation (Järvelä et al., 2025). This trend aligns with broader developments in language assessment, which aim to balance efficiency with considerations of validity and fairness.

RQ2: Most-Cited Countries and Authors

The results reveal a geographically uneven yet highly networked research landscape. The United States, China, the United Kingdom, and Australia serve as indicators of established trends of global knowledge production in this field. These countries are central nodes of collaboration networks and influence the priorities of research as well as the methodological approaches. Although this concentration can lead to innovation, the issue of representativeness of the findings of the research can also be brought up, especially

concerning different linguistic and educational backgrounds. At the individual scholar level, the appearance of a relatively steady group of highly cited authors implies the formation of an intellectual core. Notably, significant research in this field focuses more on learner variables and pedagogical effectiveness than on algorithm performance. It reveals that the field is becoming more influenced by applied linguistic perspectives rather than being merely technology-driven (Fathi et al., 2024; El Shazly, 2021).

RQ3: Dominant Research Themes and Technological Approaches

The keyword co-occurrence analysis reveals that the field is still largely grounded in foundational research like language, speech, and language development. This suggests that studies on the application of AI-assisted speaking assessment are grounded in well-established linguistic and pedagogical theories and frameworks (Zhou & Chiu, 2026). Technological terms appear less frequently, which may indicate a potential disconnection between methodological innovation and linguistic theories. While AI techniques, particularly ASR and data-driven modeling, are central to current research practices, their integration into the conceptual discourse of language assessment appears partial. This leads to significant concerns about the construct representation whereby automated systems may privilege measurable linguistic features, e.g., fluency and pronunciation, over interactional and pragmatic aspects of communicative competence (Bulut et al., 2024).

RQ4: Emerging Trends and Research Gaps

The findings indicate new directions in addition to the gaps. Evidence is mounting that more advanced AI systems can integrate multimodal data, potentially enabling a more comprehensive assessment of speaking ability (Gaballo, 2024). At the same time, critical challenges remain. Foremost among these is the issue of construct validity, particularly the extent to which automated scoring aligns with theoretically grounded models of communicative competence (Zechner & Evanini, 2020; Kang & Johnson, 2018). In addition, ethical considerations, including algorithmic bias, transparency, and data privacy, are not yet systematically embedded within the research agenda, despite their significance in high-stakes assessment contexts (Chen, 2025). The relative marginalization of affective variables within the keyword network further suggests that learner-centered dimensions, such as anxiety and willingness to communicate, remain under-integrated at a structural level (El Shazly, 2021). Moreover, the limited development of explainable AI frameworks presents a practical barrier

to adoption, as opaque scoring processes may undermine both interpretability and stakeholder trust.

In short, the findings reveal that although there is significant methodological progress with AI-assisted English speaking assessment, further theoretical combination and reflection are required. Further studies are needed to advance the field toward culturally and ethically sound, pedagogically significant progress to tackle problems of construct alignment, transparency, and the applicability of the context. Such efforts are crucial if AI technologies prove helpful in improving the effectiveness of assessment practices, validity, and educational value.

Conclusion, Implication, Limitation, and Suggestion for Future Study

This research contributes to the literature by providing a bibliometric analysis of AI applications in English-language assessment and by stressing the need for greater alignment between technological advances and pedagogical practices. The results showed that this field has been rapidly growing since 2020, driven by advances in deep learning-based speech technologies and the need for scalable assessment in digitally mediated learning environments. However, the global research landscape remains uneven; only a few countries and regions influence knowledge production and collaboration networks.

Despite significant methodological innovations, the field remains conceptually grounded in traditional linguistics and assessment models, indicating that, on the theoretical level, the integration of AI into English speaking assessment seems not to keep pace with technological innovation. Indeed, the imbalance raises concerns about construct validity, transparency, algorithmic fairness, etc. As a result, current AI-driven approaches are more inclined to prioritize measurable linguistic features over communicative competence.

This study has some limitations that should be recognized. First, the search was limited to the Scopus database; relevant studies indexed only in Web of Science or ERIC may have been omitted. Second, bibliometric analysis emphasizes citation impact over qualitative research quality; highly cited works are not necessarily methodologically sound. Third, our subject-area restriction (arts, humanities, and computer science) may have led to the underrepresentation of pedagogical studies published in other social science journals. Future research should include a broader range of databases and subject categories.

Overall, despite the above limitations, the findings indicate that alongside technological developments, theoretical, pedagogical, and ethical concerns are significant to the application of AI-assisted speaking assessment. Therefore, further research can contribute to prioritizing

construct alignment, explainability, and inclusivity of AI-assisted assessment systems that are not only efficient but also valid, fair, transparent, and relevant to the application of AI technologies in English speaking assessment.

Acknowledgement

This research was conducted without financial support from any public, commercial, or non-profit funding agencies.

Declaration of Conflicting Interests

The authors affirm that there is no conflict of interest related to the publication of this article.

References

Andersen, N., Mang, J., Goldhammer, F., & Zehner, F. (2025). Algorithmic fairness in automatic short answer scoring. *International Journal of Artificial Intelligence in Education*, *35*(5), 3128–3165. https://doi.org/10.1007/s40593-025-00495-5

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386. https://doi.org/10.1162/qss_a_00019

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.

Bernstein, J. C., & Cheng, J. (2023). Speech analysis in assessment. In *Advancing Natural Language Processing in Educational Assessment* (pp. 31–57). Routledge. https://doi.org/10.4324/9781003278658-4

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C. M., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilova, P. (2024). *The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges*. arXiv. https://doi.org/10.48550/arXiv.2406.18900Chen, Y. (2025). Transforming English language education with AI-driven deep learning models for scalable adaptive and inclusive assessment. *International Journal of Information and Communication*

Technology, *26*(12), 15–31. https://doi.org/10.1504/IJICT.2025.146163

Chou, M.-H. (2018). Speaking anxiety and strategy use for learning English as a foreign language in full and partial English-medium instruction contexts. *TESOL Quarterly*, *52*(3), 611–633. https://doi.org/10.1002/tesq.455

Derakhshan, A., Khalili, A. N., & Beheshti, F. (2016). Developing EFL learner's speaking ability, accuracy and fluency. *English Language and Literature Studies*, *6*(2), 177. https://doi.org/10.5539/ells.v6n2p177

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. https://doi.org/10.1016/j.jbusres.2021.04.070

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3), e12667. https://doi.org/10.1111/exsy.12667

El Majidi, A., de Graaff, R., & Janssen, D. (2024). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*, *28*(6), 1717–1735. https://doi.org/10.1177/13621688211050619

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*(103254), 103254. https://doi.org/10.1016/j.system.2024.103254

Forrester, A. (2020). Addressing the challenges of group speaking assessments in the time of the Coronavirus. *International Journal of TESOL Studies*, *2*(2), 74–88. https://doi.org/10.46451/ijts.2020.09.07

Fulcher, G. (2014). *Testing Second Language Speaking*. Routledge. https://doi.org/10.4324/9781315837376

Gaballo, V. (2024). Innovation in multimodal language teaching and learning: AI-powered oral performance assessment. *Rassegna Italiana di Linguistica Applicata*, *(2/3)*, 247–267. https://doi.org/10.1400/301426

Ghafar, Z. N., & Raheem, B. R. (2023). Factors Affecting Speaking Proficiency in English Language Learning: A general overview of the speaking skill. *Journal of Social Science (JoSS)*, *2*(6), 507–518. https://doi.org/10.57185/joss.v2i6.107

Hommel, B. E. (2023). Expanding the methodological toolbox: Machine-based item desirability ratings as an alternative to human-based ratings. *Personality and Individual Differences*, *213*(112307), 112307. https://doi.org/10.1016/j.paid.2023.112307

Järvelä, S., Zhao, G., Nguyen, A., & Chen, H. (2025). Hybrid intelligence: Human– AI

coevolution and learning. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, *56*(2), 455–468. https://doi.org/10.1111/bjet.13560

Kang, O., & Johnson, D. (2018). The roles of suprasegmental features in predicting English oral proficiency with an automated system. *Language Assessment Quarterly*, *15*(2), 150–168. https://doi.org/10.1080/15434303.2018.1451531

Karatay, Y., & Xu, J. (2025). Exploring the potential of conversational AI for assessing second language oral proficiency. *TESOL Quarterly*, *59*(S1), S220–S243. https://doi.org/10.1002/tesq.70003

Kunnan, A. J. (2013). Fairness and justice in language assessment. In *The Companion to Language Assessment* (pp. 1098–1114). Wiley. https://doi.org/10.1002/9781118411360.wbcla144

Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *Modern Language Journal*, *101*(1), 179–193. https://doi.org/10.1111/modl.12382

Li, J., Huang, J., & Sheeran, T. (2025). ChatGPT4o as an AI peer assessor in EFL speaking classrooms: Examining scoring reliability and feedback effectiveness. *SAGE Open, 15*(3), 1–29. https://doi.org/10.1177/21582440251369938

Liu, X., Huang, J., Deng, Y., & Spiridakis, J. (2025). AI versus human assessment in EFL speaking classrooms: a comparative study in China. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2025.2530555

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511733017

Nghia, T. L. H., Anh, N. P., & Kien, L. T. (2024). English language skills and employability: a theoretical framework. In T. L. H. Nghia, L. T. Tran, & M. T. Ngo (Eds.), *English language education for graduate employability in Vietnam* (pp. 71–93). Springer. https://doi.org/10.1007/978-981-99-4338-8_4

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1), 12. https://doi.org/10.3390/publications9010012

Rajan, S. S., Udeshi, S., & Chattopadhyay, S. (2022). AequeVox: Automated fairness testing of speech recognition systems. In E. B. Johnsen & M. Wimmer (Eds.), *Fundamental approaches to software engineering* (pp. 245–267). Springer. https://doi.org/10.1007/978-3-030-99429-7_14

Richardson, M., & Clesham, R. (2021). Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *London Review of Education*, *19*(1), 1-13. https://doi.org/10.14324/lre.19.1.09

Shaheen, H. (2025). Social media marketing research: a bibliometric analysis from Scopus. *Future Business Journal*, *11*(1), 41. https://doi.org/10.1186/s43093-025-00465-2

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, *220*, 105112. https://doi.org/10.1016/j.compedu.2024.105112

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer. https://doi.org/10.1007/978-3-319-10377-8_13

Vetrivel, S. C., Arun, V. P., Ambikapathi, R., & Saravanan, T. P. (2025). Automated grading systems. In T. Murugan, K. Periasamy, & A. M. Abirami (Eds.), *Adopting artificial intelligence tools in higher education: Student assessment* (pp. 41–61). CRC Press. https://doi.org/10.1201/9781003470304-3

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, *27*(2), 147–170. https://doi.org/10.1177/0265532209349465

Xi, X. (2023). Advancing language assessment with AI and ML–leaning into AI is inevitable, but can theory keep up? *Language Assessment Quarterly*, *20*(4–5), 357–376. https://doi.org/10.1080/15434303.2023.2291488

Xing, C., & Saeed, M. A. (2026). A Systematic Review on Artificial Intelligence (AI) Technologies in ESL/EFL Speaking Skills. *International Journal of TESOL Studies, 8*(2), 240–270. https://doi.org/10.58304/ijts.250908

Zechner, K., & Evanini, K. (Eds.). (2020). *Automated speaking assessment: Using language technologies to score spontaneous speech*. Routledge.

Zhou, X., & Chiu, K. F. T. (2026). The role of epistemic beliefs in predicting deep learning strategies in an AI-assisted English approach. *Language Testing in Asia*, *16*(1), 2–22. https://doi.org/10.1186/s40468-025-00416-2

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629