



INTERNATIONAL JOURNAL OF
MODERN EDUCATION
(IJMOE)

www.gaexcellence.com/ijmoe



ADVANCING EFL SPEAKING ASSESSMENT: A SYSTEMATIC REVIEW OF METHODS, CHALLENGES, AND INNOVATIONS

Muhammad Ikhmal Mohd Jamil¹, Norlizawati Md Tahir^{2*}, Nurul Amira Azmi³

¹Academy of Language Studies, Universiti Teknologi MARA (UiTM), 08400 Merbok, Kedah, Malaysia

 ikhmaljamil@uitm.edu.my

 <https://orcid.org/0009-0006-3903-5093>

^{2*}Academy of Language Studies, Universiti Teknologi MARA (UiTM), 08400 Merbok, Kedah, Malaysia

 lizawati@uitm.edu.my

 <https://orcid.org/0000-0003-3124-4713>

³Faculty of Business and Management, Universiti Teknologi MARA (UiTM), 08400 Merbok, Kedah, Malaysia

 amiraazmi@uitm.edu.my

 <https://orcid.org/0000-0001-5779-0427>

*Corresponding Author

Article Info:

Article history:

Received date: 03.08.2026

Revised date: 30.08.2026

Accepted date: 17.09.2026

Published date: 24.09.2026

To cite this document:

Jamil, M. I. M., Tahir, N. M., & Azmi, N. A. (2026). Advancing EFL Speaking Assessment: A Systematic Review of Methods, Challenges, And Innovations. *International Journal of Modern Education*, 8(31), 1426-1440.

Abstract:

Despite the vast use of artificial intelligence (AI) in the teaching and learning of English as a Foreign Language (EFL), specific focus on the use, outcomes, and challenges of EFL speaking assessment and practices remain limited. This study addresses the gap by synthesising systematically empirical evidence on the AI-supported EFL speaking assessment and practice. Following the PRISMA 2020 statement, 13 peer-reviewed articles were identified and analysed thematically. The analysis revealed five main themes explaining the the AI-supported EFL speaking assessment and practice, namely AI chatbot-mediated speaking practice, automated speech scoring and analytics, affective factors in AI-mediated speaking, hybrid and collaborative approaches to AI-supported speaking, and feedback and metacognitive development. Based on the analysis, AI-supported speaking environments were found to provide opportunities for repeated speaking practice and may alleviate speaking anxiety for some learners. AI-supported assessment and feedback also provided information about learners' speaking performance and supported monitoring of their progress, particularly through automated speech scoring and AI-generated feedback. The findings further revealed that digital literacy, task design, teacher involvement, and learner engagement are important considerations in AI-supported speaking activities. The study concludes that AI-mediated approaches to EFL speaking assessment and practice should be guided by appropriate pedagogical design and teacher involvement rather than used as a replacement for teachers or human interaction. Future research should explore how AI-mediated speaking

assessment and practice work across different EFL contexts and learner groups, particularly whether gains from AI-supported activities extend to communication with human interlocutors.

DOI: 10.35631/IJMOE.831082

Keyword:

Artificial Intelligence, English as A Foreign Language (EFL),
Speaking Assessment, Speaking Practice



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact ijmoe@gaexcellence.com.

Introduction

Speaking is an important component of English learning, as it enables learners to develop oral proficiency for academic, professional, and social communication. However, performing speaking practice and assessment, specifically in English as a Foreign Language (EFL) context can be challenging because oral performance involves multiple dimensions, including pronunciation, fluency, vocabulary, grammar, and interaction. Conventional English as a Foreign Language (EFL) speaking assessment methods, such as oral interviews and presentations, often require teacher or human-rater involvement, while common EFL speaking practice activities, including role-plays and group discussions, typically require teacher involvement. These requirements pose challenges to both teacher and learner as the speaking practices can be time-consuming to be conducted, and speaking assessment may be affected by differences in rater judgement and scoring consistency (Luoma, 2004; Fulcher, 2015).

These challenges have encouraged the use of technology, particularly artificial intelligence (AI) to better mediate EFL speaking practice and assessment. In terms of speaking practices, AI chatbots and other speech-enabled applications can facilitate repeated spoken interaction, while AI-based systems can support automated speech scoring and feedback during speaking assessment (Chapelle & Voss, 2016). However, existing studies on the teaching and learning of EFL mostly examine learner experiences and outcomes, including speaking anxiety, engagement, and self-regulation, across different technologies and learning contexts, rather than addressing the specific focus on AI-supported EFL speaking practice and assessment, particularly on the outcomes, challenges, and emerging approaches.

Therefore, this systematic review synthesises empirical research and explains AI-supported EFL speaking assessment and practice based on five identified themes, namely AI chatbot-mediated speaking practice, automated speech scoring and analytics, affective factors in AI-

mediated speaking, hybrid and collaborative approaches to AI-mediated speaking, and feedback and metacognitive development. As a result, this review provides a structured account of the emerging evidence and identifies directions for further research on the integration of AI into EFL speaking assessment and practice.

Methodology

Review Protocol

A systematic review was deemed appropriate for this study because the identified studies varied in their technologies, speaking activities, assessment approaches, learner contexts, and research designs, as asserted by Xiao & Watson (2019) and Snyder (2019). To ensure transparent reporting of the review process, this systematic literature review (SLR) followed the recommendation by Page et al. (2021) by adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, which guided the search and selection of studies, study quality appraisal, and data extraction and analysis. For the review, the search focused on studies published between 2021 and 2026 from Scopus database addressing methods and approaches, challenges, and emerging innovations in relation to the use of AI in supporting EFL speaking assessment and practice. All the identified studies were then screened against predefined eligibility criteria and assessed in full text. The selected studies quality was later appraised using a framework adapted from Abouzahra et al. (2020), and relevant data were extracted and thematically analysed following Braun and Clarke (2006) to identify patterns across the included studies.

Research Question

According to Xiao and Watson (2019), clearly formulated research questions in systematic literature reviews is essentially significant, as they determine the scope of the review and guide the review process, including the search strategy, screening and eligibility assessment, data extraction, and thematic analysis of relevant evidence. For this study, the following research questions were formulated:

- **RQ1:** What methods and approaches are commonly used in AI-supported EFL speaking assessment and practice across existing studies?
- **RQ2:** What are the major challenges and emerging innovations associated with AI-supported EFL speaking assessment and practice?

Systematic Search and Study Selection

The systematic search and study selection process was conducted to identify empirical studies relevant to AI-supported EFL speaking assessment and practice. The process comprised identification, screening, and eligibility assessment, followed by the inclusion of studies that met the predefined criteria. Figure 1 summarises the overall search and study selection process.

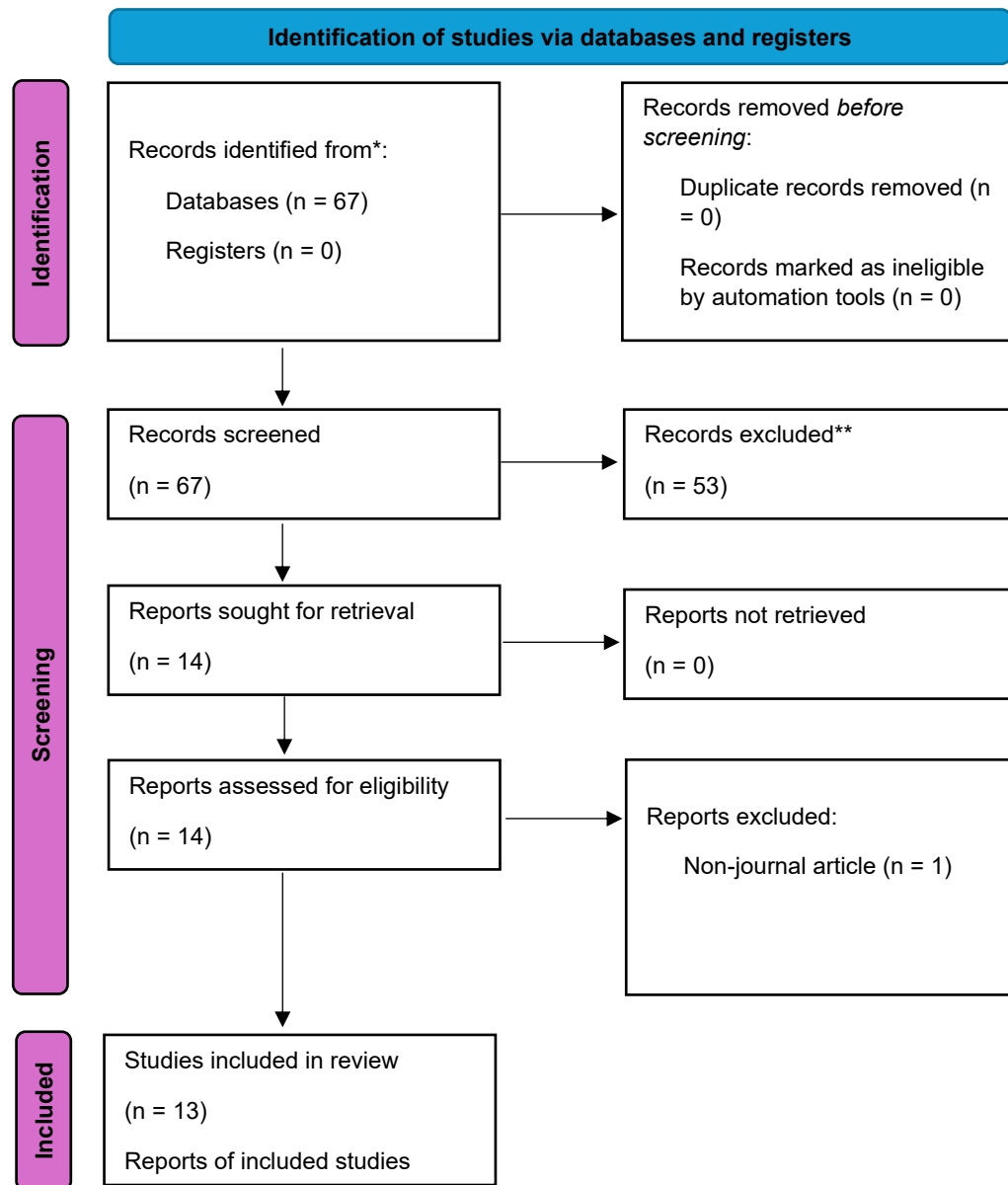


Figure 1: Systematic Study and Selection Process

Identification

The search was conducted exclusively in the Scopus database using predefined search terms. Boolean operators (AND/OR), phrase searching, truncation, and field-specific searching were used to identify relevant records. The search was limited to English-language publications published between 2021 and 2026 to capture recent developments in technology-mediated language learning and speaking assessment (García-Peñalvo, 2022). The search strategy was also informed by previous research on AI and technology-mediated language learning (Yan et al., 2023).

The Scopus search string was:

TITLE-ABS-KEY (("ESL" OR "English as a Second Language" OR "EFL" OR "English language learner") AND ("speaking assessment*" OR "oral assessment*" OR "speaking evaluation*" OR "oral proficiency assessment*")) AND (LIMIT-TO (LANGUAGE , "English")) AND PUBYEAR > 2020 AND PUBYEAR < 2027.*

The search identified 67 records, which were exported for screening and duplicate checking.

Screening

The identified records were screened by examining their titles and abstracts against the predefined inclusion and exclusion criteria. The criteria focused on studies addressing EFL speaking assessment or practice through technology-mediated approaches, including AI-supported tools, automated assessment, digital feedback, or related technologies. Studies focusing exclusively on written language skills or lacking a clear connection to speaking assessment or practice were excluded. Non-peer-reviewed publications, such as white papers and editorials, were also excluded.

No duplicate records were identified. Of the 67 records screened, 53 were excluded because they did not meet the selection criteria. The remaining 14 reports were retrieved for full-text assessment.

Eligibility

The 14 retrieved reports were assessed against the inclusion and exclusion criteria through full-text examination. Following this assessment, 13 studies met the eligibility criteria and were included in the review. One report was excluded because it was a non-journal article.

Quality Appraisal

The 13 included studies underwent a structured quality appraisal to assess their methodological and reporting quality. Quality appraisal is an important component of systematic reviews because it allows the methodological strengths and limitations of included studies to be considered when interpreting the evidence (Kitchenham & Charters, 2007; Grant & Booth, 2009). The appraisal framework was adapted from Abouzahra et al. (2020) and comprised six criteria:

- **QA1:** Is the purpose of the study clearly stated?
- **QA2:** Is the interest and usefulness of the work clearly presented?
- **QA3:** Is the study methodology clearly stated?
- **QA4:** Are the concepts of the approach clearly defined?
- **QA5:** Is the work compared or measured against other similar studies?
- **QA6:** Are the limitations of the work mentioned?

Each criterion was scored as Yes (1.0) when fully met, Partly (0.5) when partially addressed, and No (0) when the criterion was not addressed. The maximum possible score was 6.0. Studies scoring 3.0 or above were considered to have met the minimum quality threshold for inclusion in the final synthesis. The appraisal process supported a consistent assessment of the included

studies and reduced the influence of individual judgement during appraisal (Moher et al., 2009).

Data Extraction

Following the selection and quality appraisal stages, data were systematically extracted from the 13 included studies. The extraction process was guided by the research questions and focused on information relevant to EFL speaking assessment and practice, technology-mediated approaches, reported outcomes, challenges, and emerging innovations (Boland et al., 2017). The principal author extracted information primarily from the findings and discussion sections of the included studies. The methodology and other sections were consulted when additional information about the study context, participants, technologies, or research design was required. A co-author subsequently checked the extracted information against the original studies to support the accuracy and consistency of the extraction process (Gough et al., 2017).

The extracted findings were synthesised using thematic analysis, which was considered appropriate for identifying recurring patterns across studies employing different research designs (Snyder, 2019). The analysis followed the six phases proposed by Braun and Clarke (2006):

1. Familiarisation: Repeated reading of the selected studies to become familiar with the reported findings.
2. Initial coding: Identifying relevant concepts and findings, including automated scoring, speaking anxiety, chatbot-mediated practice, and AI-generated feedback.
3. Theme development: Grouping related codes into potential themes.
4. Theme review: Reviewing the potential themes against the coded data and the complete dataset.
5. Theme definition: Refining and defining the content and focus of each theme.
6. Reporting: Organising the themes into a coherent synthesis and relating the findings to the research questions.

The analysis resulted in five themes: (1) AI Chatbot-Mediated Speaking Practice; (2) Automated Speech Scoring and Analytics; (3) Affective Factors in Digital Assessment; (4) Hybrid and Collaborative Assessment Approaches; and (5) Feedback and Metacognitive Development. These themes formed the basis for the synthesis of findings addressing the two research questions concerning the methods and approaches used in AI-mediated EFL speaking assessment and practice and the challenges and emerging innovations associated with these practices.

Results and Discussion

Study Background

This section presents the bibliographic and contextual characteristics of the 13 studies included in the final synthesis. The descriptive findings provide an overview of publication trends, geographical distribution, research contexts, and methodological characteristics of studies examining technology-mediated ESL speaking assessment.

Distribution by Publication Year

The 13 included studies were published between 2021 and 2026, with a greater concentration of publications appearing from 2024 onwards. This recent increase indicates growing research interest in the application of digital technologies and artificial intelligence (AI) to speaking assessment. The inclusion of studies published in 2025 and 2026, including early-access or in-press publications where applicable, further indicates that technology-mediated speaking assessment remains an active and developing area of research.

The concentration of recent studies is also consistent with the increasing availability of generative AI and AI-supported language learning tools. However, given the relatively small number of studies included in this review, the observed increase should be interpreted as an indication of recent research activity rather than evidence of a definitive publication trend.

Geographical Location and Study Context

The included studies were predominantly conducted in Asian and Middle Eastern contexts, including China, Indonesia, Taiwan, Thailand, Turkey, Saudi Arabia, and Egypt. This geographical distribution indicates considerable research attention to AI-supported speaking assessment and practice within EFL contexts in these regions.

The geographical concentration also limits the extent to which the findings can be generalised to other educational settings. In particular, the predominance of Asian and Middle Eastern contexts means that the findings may reflect the educational practices, technological infrastructure, and language-learning environments of these settings. In addition, all 13 included studies were situated in EFL contexts, with most involving learners in higher education.

Methodological and Academic Characteristics

The included studies used a range of research designs, with mixed-methods and quasi-experimental approaches appearing frequently. These designs allowed researchers to examine changes in speaking-related outcomes and, in some cases, learners' experiences of AI-supported activities. The studies therefore provide evidence on both the use of technology and learners' responses to it.

Most of the included studies were published as journal articles and involved learners in higher education. The emphasis on university students indicates that current research on AI-supported EFL speaking assessment and practice has largely been conducted in tertiary settings. Comparatively less evidence is available for primary and secondary learners.

The included studies also differed in their technological focus. Some examined AI chatbots and conversational tools, whereas others focused on automated speech scoring, AI-generated feedback, or technology-supported collaborative activities. This variation provides a broad view of how AI is currently being incorporated into EFL speaking-related activities.

Interpretation of Patterns

Overall, the descriptive findings show that research on AI-supported EFL speaking assessment and practice is relatively recent. The included studies were concentrated in Asian and Middle Eastern contexts and mainly involved university learners. The increasing number of recent publications also corresponds with the growing use of AI-based technologies in English language education.

However, the evidence base remains limited in size and scope. The concentration of studies in particular regions and educational levels mean that the findings cannot be assumed to apply equally to other learner populations or educational settings. Further research involving different age groups, learning environments, and assessment contexts is therefore needed.

Quality Appraisal Results

The methodological quality of the 13 selected studies was evaluated using six criteria adapted from Abouzahra et al. (2020). The criteria assessed the clarity of the study purpose (QA1), usefulness of the work (QA2), clarity of the methodology (QA3), definition of the core concepts (QA4), comparison with similar studies (QA5), and acknowledgement of limitations (QA6).

Each criterion was scored as Yes (1.0), Partly (0.5), or No (0). As specified in the review protocol, studies achieving a minimum score of 3.0 out of 6.0 were considered eligible for the final synthesis.

Table 1: Quality Appraisal Results for Selected Articles

Study	QA1	QA2	QA3	QA4	QA5	QA6	Total Marks	Decision
Susoy (2026)	Yes	Yes	Yes	Yes	Yes	Yes	6.0	Included
Yu & Qiu (2026)	Yes	Yes	Yes	Yes	Partly	Partly	5.0	Included
El Shazly (2021)	Yes	Yes	Yes	Yes	Yes	Partly	5.5	Included
Li (2025)	Yes	Yes	Yes	Yes	Partly	Yes	5.5	Included
Wang et al. (2025)	Yes	Yes	Yes	Yes	Yes	Partly	5.5	Included
Wu et al. (2025)	Yes	Yes	Yes	Yes	Partly	Partly	5.0	Included
Wu et al. (2026)	Yes	Yes	Yes	Yes	Yes	Yes	6.0	Included
Waluyo et al. (2025)	Yes	Yes	Yes	Yes	Partly	Partly	5.0	Included
Gong (2023)	Yes	Yes	Yes	Yes	Yes	Partly	5.5	Included
Limgomvilas et al. (2025)	Yes	Yes	Yes	Yes	Partly	Yes	5.5	Included
Dakhil et al. (2025)	Yes	Yes	Yes	Yes	Yes	Partly	5.5	Included
Zou et al. (2023)	Yes	Yes	Yes	Yes	Partly	Partly	5.0	Included
Wang & Zou (2025)	Yes	Yes	Yes	Yes	Yes	Yes	6.0	Included

Interpretation of Quality Appraisal Results

All 13 studies met the minimum quality threshold, with scores ranging from 5.0 to 6.0. All studies received a Yes rating for QA1, QA2, QA3, and QA4, indicating that the purpose,

usefulness, methodology, and core concepts were clearly reported. The lower ratings occurred mainly for QA5 and QA6. Several studies were rated Partly for comparison with similar studies and/or reporting of limitations. Thus, although the included studies generally provided clear information about their research purposes and methods, some provided less detailed discussion of their relationship to previous research or their study limitations.

Overall, the appraisal results support the inclusion of all 13 studies in the synthesis. The appraisal also indicates that the findings should be interpreted with attention to the differences in reporting quality across studies. This is consistent with the importance of considering methodological quality and study limitations when synthesising evidence from different sources (Higgins et al., 2019).

The Themes

The thematic analysis identified five themes: (1) AI Chatbot-Mediated Speaking Practice, (2) Automated Speech Scoring and Analytics, (3) Affective Factors in Digital Assessment, (4) Hybrid and Collaborative Approaches to AI-Mediated Speaking, and (5) Feedback Loops and Metacognitive Development.

The first, second, and fourth themes primarily address how AI is used in EFL speaking activities, including chatbot-mediated practice, automated speech scoring, and hybrid and collaborative approaches. These themes therefore contribute mainly to RQ1 by identifying the approaches used in AI-mediated EFL speaking activities. The third and fifth themes focus on learners' affective experiences, feedback, and metacognitive development, while also highlighting factors that may influence learners' experiences with AI-supported speaking activities. Together, the five themes contribute to RQ2 by identifying the outcomes, experiences, and factors associated with AI-mediated EFL speaking activities.

Theme 1: AI Chatbot-Mediated Speaking Practice

Wang et al. (2025) and Wu et al. (2025) found that AI chatbots, featured in GPT-based conversational activities and Think-Pair-Share, contributed to improved EFL learners' fluency and lexical diversity during speaking practice. These findings suggest that AI chatbots provide learners with opportunities for repeated speaking practice and exposure to spoken English without requiring a human interlocutor for every interaction. These findings answer RQ1 by identifying AI chatbot-mediated practice as an approach used in AI-supported EFL speaking activities.

Despite the benefits of AI chatbots, several studies reported varied learner experiences when using them for speaking practice. Susoy (2026) and El Shazly (2021) reported anxiety among EFL learners when interacting with chatbots, possibly due to the absence of human facilitation and the unpredictable responses generated by AI chatbots. Another concern expressed by EFL learners when using AI chatbots for speaking practice is reliance on AI-supported translation and comprehension functions (Limgomolvilas and Sukserm, 2025). These findings indicate that the benefits of AI chatbot-mediated speaking practice may depend on learners' level of anxiety, technology-related difficulties, and reliance on AI represent, which address RQ2.

These findings can be interpreted through the Optimal Input Hypothesis and Translanguaging Theory, particularly in relation to learners' use of AI-supported interaction and language

resources to support comprehension and speaking practice. However, the findings do not establish that AI chatbots are inherently more effective than human interaction. Rather, they suggest that AI chatbots can provide an additional space for speaking practice when their use is appropriately structured.

Theme 2: Automated Speech Scoring (ASS) & Analytics

The findings revealed that speaking can assessment can be performed using Automated speech scoring and analytics, particularly in terms of EFL learners' pronunciation and fluency. In the study by Li (2025), the AI-supported oral assessment system used to assess speaking reported accuracy rates of 94.7% for native speakers and 86.6% for non-native speakers. Yu and Qiu (2026) also demonstrated the use of AI-mediated EFL speaking assessment when examining AI-based speech recognition and corrective functions to identify EFL learners' needs and their speaking performance. These findings address RQ1 by identifying automated speech scoring and analytics as an approach used in AI-mediated EFL speaking assessment.

However, Gong (2023) identified concerns regarding automated scoring when examining the washback of SpeechRater in the TOEFL iBT context and. The study revealed that automated assessment may influence not only how speaking performance is evaluated but also what learners choose to practise, as learners may focus on improving their scores rather than developing broader communicative ability. The findings also indicate that automated scores may not represent all dimensions of speaking ability, particularly aspects involving interaction and spontaneous communication. These findings answer RQ2 by highlighting the washback effects and limitations associated with automated speaking assessment.

The findings can be related to Communicative Language Ability (CLA), which views speaking ability as involving multiple dimensions of communication. Although automated systems can provide information about features such as pronunciation and fluency, the reviewed evidence suggests that these measures should not necessarily be treated as a complete representation of speaking ability. Automated scoring may therefore be useful as part of speaking assessment, but its use needs to be considered alongside the communicative purpose of the assessment. Future research should examine how automated systems can better address dimensions of speaking that are more difficult to assess automatically, including interactional and pragmatic aspects of communication.

Theme 3: Affective Factors in Digital Assessment

Affective factors were identified across several studies examining AI-supported speaking assessment and practice. These included foreign language speaking anxiety, self-efficacy, willingness to communicate, enjoyment, and perceived communicative competence. Susoy (2026) found differences in the relationship between speaking anxiety and achievement across human-facilitated and AI-facilitated conditions. Wu et al. (2025, 2026) reported reductions in speaking anxiety and changes in foreign language enjoyment following AI-supported speaking activities, while Wang and Zou (2025) reported positive effects on learners' self-perceived communicative competence. These findings indicate that learners' affective experiences may vary according to how AI is incorporated into speaking activities.

However, the findings were not consistently positive. El Shazly (2021) reported that difficulties with chatbot interaction and technical problems could contribute to learner anxiety. Yu and Qiu

(2026) also indicated that learners' technological proficiency may influence their experiences with AI-supported learning. These findings suggest that AI does not automatically reduce affective difficulties. Although AI-mediated interaction may reduce some interpersonal pressure associated with speaking, technology-related difficulties and differences in digital literacy may create additional challenges. This theme therefore contributes mainly to RQ2 by identifying affective responses and digital literacy as factors that may influence learners' experiences with AI-mediated speaking assessment and practice.

The findings can be considered from a Positive Psychology perspective, which emphasises positive experiences that may support learners' engagement in language learning. The variation in learners' affective experiences across studies suggests that the value of AI-supported speaking activities may depend on how learners experience and engage with the technology. From a practical perspective, this highlights the need to provide appropriate support to help learners engage effectively with AI-supported speaking activities.

Theme 4: Hybrid and Collaborative Approaches to AI-Mediated Speaking

Hybrid and collaborative approaches were identified across several studies examining AI-mediated speaking activities. These approaches involved integrating AI chatbots with peer interaction, collaborative learning, and teacher support rather than using AI as a standalone tool. Wu et al. (2025) found that integrating AI chatbots into Think-Pair-Share (TPS) activities led to greater improvements in speaking performance and language enjoyment than conventional collaborative methods. Similarly, Wu et al. (2026) found that chatbot-supported collaborative learning positively influenced student engagement, which in turn predicted improvements in speaking skills. These findings indicate that AI can be incorporated into collaborative speaking activities to support learners' engagement and speaking development.

However, the way AI was integrated into learning activities varied across studies. Li (2025) reported that teacher-AI interaction was associated with learning and self-regulation, whereas Lingomolvilas and Sukserm (2025) found that students also used AI tools independently outside teacher-led activities. Susoy (2026) and Wu et al. (2025) examined AI-supported speaking activities within structured EFL learning contexts, while Li (2025) examined a more technologically integrated approach involving Generative AI and IoT. These findings suggest that hybrid and collaborative approaches can range from structured teacher-supported activities to more learner-initiated use of AI. This theme therefore contributes mainly to RQ1 by identifying hybrid and collaborative integration as an approach used in AI-supported EFL speaking activities.

The findings can be considered from a Social Constructivist perspective, which emphasises learning through interaction and collaboration with others. The integration of AI with peer and teacher interaction suggests that AI-mediated speaking activities may support learning through a combination of technological and social interaction. From a practical perspective, this highlights the importance of integrating AI into structured collaborative activities rather than relying solely on individual interaction with AI.

Theme 5: Feedback Loops & Metacognitive Development

The final theme concerns the use of AI-generated feedback to provide learners with information about their speaking performance and support their reflection on learning. Li (2025) reported

improvements in metacognitive knowledge and self-regulation following the use of an AI-supported oral assessment system. Wang and Zou (2025) also reported positive effects of AI-supported speaking instruction on learners' self-perceived communicative competence. These findings suggest that AI-generated feedback may provide learners with information that can support monitoring of their speaking performance and identification of areas for improvement. This theme contributes to RQ1 by identifying feedback as an approach used in AI-supported speaking assessment and practice and to RQ2 by highlighting metacognitive development and learner engagement as outcomes and considerations associated with its use.

However, the findings also indicate that access to AI-generated feedback does not necessarily result in effective learning. Limgomolvilas and Sukserm (2025) raised concerns about learners' use of AI for translation and comprehension rather than engaging more deeply with the language produced. Wu et al. (2026), in contrast, highlighted the importance of active engagement with AI-supported activities in speaking development. These findings suggest that the usefulness of AI-generated feedback depends partly on how learners engage with the information provided and how the activity is designed. Teacher guidance may therefore be important in helping learners interpret feedback and apply it to subsequent speaking activities.

From a Self-Regulated Learning perspective, the findings suggest that AI-generated feedback may support learners' monitoring and regulation of their speaking performance. However, the reviewed evidence does not establish that AI feedback automatically produces metacognitive development. Rather, the reported improvements appear to be associated with particular AI-supported learning conditions and learner engagement. Future research should therefore examine how learners interpret and act on AI-generated feedback and whether such feedback leads to sustained improvements in speaking performance and self-regulated learning.

Conclusion

This systematic literature review synthesised 13 peer-reviewed studies on AI-mediated EFL speaking assessment and practice. The review identified five themes: AI chatbot-mediated speaking practice, automated speech scoring and analytics, affective factors in AI-mediated speaking, hybrid and collaborative approaches to AI-supported speaking, and feedback and metacognitive development. In relation to RQ1, the findings show that AI is being used in several ways to support EFL speaking assessment and practice. AI chatbots provide opportunities for repeated spoken interaction, while automated speech scoring and analytics provide information about aspects of learners' speaking performance. AI is also being incorporated into collaborative activities involving peers and teachers, and AI-generated feedback is used to support learners in monitoring their performance. These findings indicate that AI-mediated speaking assessment and practice involve different approaches rather than a single form of technology use.

For RQ2, the findings highlight several challenges and considerations in the use of AI for EFL speaking assessment and practice. Learners may experience technology-related difficulties and speaking anxiety, while differences in digital literacy may influence their experiences with AI-supported activities. Some learners may also rely on AI for translation and comprehension when engaging with speaking activities. Automated scoring may lead learners to focus more on assessment performance than on broader communicative ability. The reviewed studies also reported positive outcomes of the use of AI for EFL speaking assessment and practice, including improvements in speaking performance, enjoyment, perceived communicative

competence, engagement, self-regulation, and metacognitive development. However, AI-supported feedback may contribute to metacognitive development only when accompanied by active learner engagement and appropriate guidance. The studies also highlighted learner engagement, task design, and teacher involvement as important considerations in AI-supported speaking activities.

Although the findings confirm that the use of AI can extend opportunities for EFL speaking practice, assessment, and feedback, its use does not eliminate the need for teachers or human interaction. AI-mediated approaches to EFL speaking assessment and practice should therefore be guided by appropriate pedagogical design and teacher involvement. Future research should explore the use of AI for EFL speaking assessment and practice across different EFL contexts and learner groups, particularly whether AI-supported activities significantly impact communication with human interlocutors.

Acknowledgements:	The authors sincerely valued the insights and technical support provided by those involved in the study. Appreciation is also extended to the reviewers for their constructive feedback. Special thanks to the editorial board of IJMOE for the guidance and publication opportunity.
Funding Statement:	This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Conflict of Interest Statement:	The authors declare no conflict of interest in the publication of this study.
Ethics Statement:	The study strictly adhered to the established standards of academic integrity and ethical publishing. It did not involve human participants, animals, or sensitive data requiring ethical approval.
Author Contribution Statement:	All authors contributed to the study by developing the research design, collecting and analysing the data, interpreting the findings, and preparing, reviewing, and approving the manuscript for publication.

References

- Abouzahra, M., Tan, J., & Ghasemaghahi, M. (2020). Exceptional health information seeking and the role of internet-based technology: A systematic literature review. *International Journal of Medical Informatics*, 139, 104149. <https://doi.org/10.1016/j.ijmedinf.2020.104149>
- Boland, A., Cherry, M. G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd ed.). SAGE Publications.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(7), e12667. <https://doi.org/10.1111/exsy.12667>
- García-Peñalvo, F. J. (2022). Developing the "State of the Art" of a Topic: Editorial. *Education in the Knowledge Society (EKS)*, 23, Article e28640. <https://doi.org/10.14201/eks.28640>
- Gong, K. (2023). Challenges and opportunities for spoken English learning and instruction brought by automated speech scoring in large-scale speaking tests: A mixed-method investigation into the washback of SpeechRater in TOEFL iBT. *Language Testing in Asia*, 13(1), 1–25. <https://doi.org/10.1186/s40468-023-00234-x>
- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd ed.). SAGE Publications.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). John Wiley & Sons.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Keele University and University of Durham.
- Li, Z. (2025). Bridging pedagogy and technology: A generative AI and IoT approach to transformative English language education. *Humanities and Social Sciences Communications*, 12(1), 1–15. <https://doi.org/10.1057/s41599-025-06151-6>
- Limgomolvilas, S., & Sukserm, P. (2025). Beyond teacher-led approaches: Student-initiated translanguaging with artificial intelligence tools in foreign language acquisition. *Journal of Language and Education*. [Advance online publication].
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(n71). <https://doi.org/10.1136/bmj.n71>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Susoy, Z. (2026). Reducing anxiety and enhancing performance: The impact of AI chatbots versus human facilitation on EFL speaking assessment outcomes. *Frontiers in Psychology*, 16, 1745942. <https://doi.org/10.3389/fpsyg.2025.1745942>

- Wang, C., & Zou, B. (2025). Enhancing EFL learners' speaking performance and affective experiences through AI-supported language learning. *IRAL - International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2024-0310>
- Wang, J., Wang, C., Xiao, T., & Zhang, X. (2025). Enhancing English language learning through moral dilemmas: A comparative study of GPT and human-written stories. *Language Teaching Research*. <https://doi.org/10.1177/13621688251391412>
- Wu, T.-T., Hapsari, I. P., & Huang, Y.-M. (2025). Effects of incorporating AI chatbots into think-pair-share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*. [Advance online publication].
- Wu, T.-T., Hapsari, I. P., & Huang, Y.-M. (2026). Intelligent chatbot-supported collaborative learning: Impact on student engagement and English speaking skills. *ReCALL*. [Advance online publication].
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723394>
- Yan, X., Zhang, C., & Fan, J. J. (2023). Sixty years of language assessment: A bibliometric analysis of *Language Testing and Language Assessment Quarterly*. *Language Testing*, 40(2), 488–514. <https://doi.org/10.1177/02655322221124403>
- Yu, J., & Qiu, B. (2026). Big data and AI in English language learning: Predicting student performance and identifying learning gaps. *Journal of Applied Science and Engineering*, 29(6), 1459–1470. [https://doi.org/10.6180/jase.202606_29\(6\).0013](https://doi.org/10.6180/jase.202606_29(6).0013)