



ISRG PUBLISHERS

Abbreviated Key Title: ISRG J Edu Humanit Lit

ISSN: 2584-2544 (Online)

Journal homepage: <https://isrgpublishers.com/isrgjehl/>

Volume – III Issue – V (September-October) 2026

Frequency: Bimonthly



Bibliometric Analysis of Research on Foreign Language Speaking Assessment in China

Juntong Wang¹, Weijie Hu^{2*}

¹ BA student, College of Foreign Languages, Zhejiang Normal University; No. 688 Yingbin Avenue, Jinhua City, Zhejiang 321000, China.

² Professor, College of Foreign Languages, Zhejiang Normal University; No. 688 Yingbin Avenue, Jinhua City, Zhejiang 321000, China.

| Received: 17.10.2026 | Accepted: 22.10.2026 | Published: 28.10.2026

*Corresponding author: Weijie Hu

Abstract

As China continues to deepen its opening-up policies and gradually improve the foreign language speaking assessment system, foreign language learning has shifted its focus from written proficiency alone to practical language use. Meanwhile, the body of research literature on foreign language speaking assessment in China has been growing, with increasingly diverse assessment perspectives. Based on CNKI literature published from 2005 to 2026 (literature retrieval cut-off: September 2026; the 2026 dataset does not include publications for the full calendar year), this paper adopts bibliometric methods to examine publication volume, patterns of publication changes, research hotspots and trends in domestic foreign language speaking assessment research. It further predicts directions for future research. Visual analysis of CNKI literature shows that existing studies mainly target university-level learners, with core themes covering speaking instruction, assessment and learners' oral production. Nevertheless, current research suffers from limitations including narrow research coverage, inconsistent long-term research efforts and insufficient cross-team collaboration among scholars.

Keywords: Research on foreign language speaking assessment in China, Bibliometrics, CiteSpace, Research hotspots, Research trends, Mapping analysis

1. Introduction

Research on foreign language speaking assessment in China closely aligns with national language planning policies and core goals of assessment reform (Zhou, 2019). With the wide application of AI, new assessment tools have been continuously developed.

The interdisciplinary integration promoted by the New Liberal Arts initiative has also been gradually adopted in this field, bringing new opportunities for research on foreign language speaking assessment in China (Li, 2026). Conducting bibliometric analysis in this area carries practical significance for improving China's

foreign language speaking assessment system and meeting the demands of foreign language education reform and opening-up in the new era.

A review of existing domestic studies reveals several major research lines of empirical work on speaking assessment. Some scholars focus on the internal features of learners' oral production, conducting quantitative analysis on dimensions such as speech complexity(Li, 2026), fluency(Shi, 2013), noun phrase use(Bao, 2025) and affective word performance(Bao & Chen, 2026) to explore how language proficiency shapes oral output indicators. Other studies build on the China Standards of English Language Ability (CSE)(Liu & Yang, 2021), exploring how to apply this framework in speaking test design, classroom teaching, self-assessment and peer assessment, and refining the construct of speaking ability and scoring criteria. Alongside advances in smart education, many studies discuss AI-integrated approaches for the integrated teaching, learning and assessment of speaking, and reflect on opportunities and ethical concerns of large-scale automated speaking tests(Li, 2026). In addition, some research adopts cognitive and psychological perspectives to examine how individual factors such as foreign language speaking anxiety(Xu et al., 2023), cognitive agency(Liao & Li, 2025) and critical thinking(Yin & Jin, 2022) influence test performance in speaking assessments.

However, notable gaps remain in existing research. Most empirical studies investigate specific topics as isolated case inquiries. Few efforts systematically map the overall landscape of foreign language speaking assessment research in China. Moreover, bibliometrics and visualisation tools have rarely been used to comprehensively summarize the field's hot topics, evolutionary paths and cutting-edge trends. This makes it difficult for researchers to grasp the big picture of the research field and identify weak points and potential directions for further exploration.

Against this background, this paper uses core journal literature from CNKI as raw data and applies bibliometric and visual analysis in CiteSpace. It systematically sorts out research hotspots, main research strands and evolutionary trends of foreign language speaking assessment in China, and analyses strengths and drawbacks of current studies. It aims to offer references for subsequent empirical research, optimisation of speaking assessment systems and reforms in oral English teaching.

2. Data Sources

This study adopts CiteSpace to conduct visual bibliometric analysis of literature data. All data are retrieved from the China National Knowledge Infrastructure (CNKI). Advanced search is performed using subject keywords including "foreign language", "speaking", "English" and "assessment". Source types are restricted to core journals, namely Peking University Core Journals, CSSCI and AMI. The time span ranges from January 2005 to September 2026. After manual screening, a total of 201 valid papers are obtained. The literature dataset is then imported into CiteSpace to generate visual maps for analysing research hotspots, research directions, trends and other features of domestic research on foreign language speaking assessment.

3. Specific Research Analysis

3.1 Temporal Analysis of Publication Output and Research Trends

The annual publication volume of studies on foreign language speaking assessment in China shows obvious fluctuations. As illustrated in Figure 1, the number of relevant papers kept rising from 2005 to 2008. Research in this period mainly addressed fundamental issues concerning foreign language speaking assessment. The publication count hit its peak for this whole period in 2008. From 2008 to 2011, publication output declined and reached its lowest point in 2011. Publications rebounded rapidly in 2012, forming the second research hotspot peak.

After 2012, the overall publication volume in this field experienced a slow downward trend with minor intermittent recoveries. Short research peaks appeared in 2017, 2019 and 2023, yet the corresponding publication numbers were far lower than those in 2008 and 2012. Publications continued to drop after 2023. Such fluctuations in overall output are likely associated with policy adjustments for foreign language assessment and the application of emerging information technologies in teaching and assessment. Viewed from the long-term trajectory, this research field enjoyed high popularity in its early stage, followed by gradual cooling of research interest and shrinking paper output.

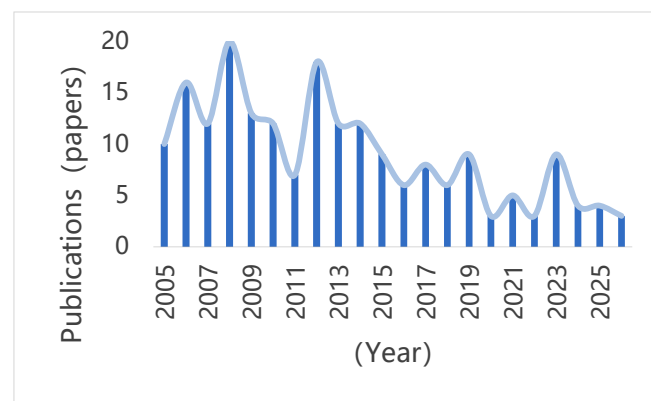


Figure 1 Annual Publication Volume of Research on Foreign Language Speaking Assessment in China (2005–2026) (Literature retrieval cutoff: September 2026. The 2026 dataset is incomplete and for reference only.)

3.2 Co-occurrence and Cluster Analysis of Core Authors

The author collaboration network (see Figure 2) is analysed, and scholars with no fewer than two publications are selected. According to the structure of the collaboration network, several major research clusters centred on leading scholars such as Bao G., Jin Y., X J.F. and Y H.Z. have taken shape in the field of foreign language speaking assessment in China. Nevertheless, the collaboration density between different research groups remains low.

Most researchers work individually or collaborate within small teams. Large-scale cross-team research is rare, and a tightly connected research community has not yet been established. Further textual review of representative works by these core authors reveals that their research mainly focuses on validating foreign language speaking tests (Qi et al., 2024) and developing intelligent speaking assessment systems (Xu et al., 2026). Their studies generally prioritise the optimisation of assessment instruments (Zou, 2023) and the improvement of test validity (Jin & Jie, 2017).

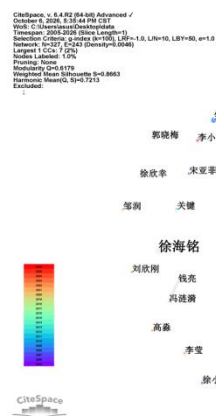


Figure 2 Core Author Co-occurrence Map

4. Analysis of Research Hotspots and Trends

4.1 Co-cited Literature and Keyword Co-occurrence Analysis

Analysis of highly cited literature helps identify core research achievements in this field and further reveals the evolutionary path of research hotspots and the structure of knowledge(Bai & Wang, 2026). Table 1 lists the top six highly cited papers in studies on foreign language speaking assessment in China. The analysis shows that highly cited literature mainly falls into two major strands. The first strand focuses on oral English classrooms, exploring the design and implementation of speaking instruction. The second strand focuses on speaking assessment, discussing existing problems and corresponding solutions for the development of speaking tests in China.

Among these, Zhang (2017) investigating the effectiveness of Production-oriented Approach (POA) has the highest citation count of 1017 and ranks first in download volume, making it an

influential representative work in this field. The top four cited publications all revolve around classroom teaching(Cheng et al., 2007; Qi et al., 2016; Zhang, 2007; Zhang, 2017), while the fifth and sixth shift to scale construction and test research(Zhong et al., 2013; Cai, 2005). Notably, there is a certain mismatch between highly cited papers and core authors in this domain. This phenomenon may arise because some highly cited works were published at an earlier stage. Their authors mostly did not carry out continuous research on foreign language speaking, resulting in a lack of sustained in-depth exploration. This explains the discrepancy between authors of highly cited papers and the core author group.

The statistical results of high-frequency keyword co-occurrence in Table 2 corroborate the two major research themes presented by the highly cited literature. Highly cited papers reflect two main lines: oral classroom instruction and speaking assessment, which is also mirrored in keyword statistics. “College English” and “oral teaching” rank high in frequency and centrality, corresponding to numerous highly cited studies in Table 1 that conduct classroom-based oral teaching research. This indicates that classroom teaching practice has long attracted scholarly attention in this field. Meanwhile, keywords such as “oral test”, “validity” and “reliability” also have high frequency and centrality, matching the orientation of later highly cited papers in Table 1 concerning test development, validity verification and scale research.

It can be seen that whether from highly cited literature or keyword co-occurrence results of the full dataset, inquiry into teaching practice and assessment quality constitute two parallel pillars of this field. On this basis, keywords such as “communicative competence” and “fluency” further specify micro-level research focuses. Within the broad framework of teaching and assessment, scholars also examine concrete issues including features of learners’ oral production and ability constructs, enriching the research content of this field.

Table 1 Top 6 Highly Cited Papers of Research on Foreign Language Speaking Assessment in China

Author	Title	Journal	Year	Citations	Downloads
Zhang	<i>On the Teaching Effectiveness of the Production-oriented Approach</i>	<i>Modern Foreign Languages</i>	2017	1017	15370
Cheng, et al.	<i>A Study on the Influence of English Classroom Anxiety on College Students' Oral English Expression</i>	<i>Language Teaching and Linguistic Studies</i>	2007	267	4036
Qi, et al.	<i>A Study on the Design and Effects of an English Audio-Visual-Speaking Course Based on POA</i>	<i>Education Academic Monthly</i>	2016	265	4790
Zhang	<i>On the Organizational Strategies of Cooperative Learning in Oral English Classrooms</i>	<i>Foreign Languages and Their Teaching</i>	2007	256	9270
Zhong, et al.	<i>A Pilot Study on the Construction of a Self-assessment Scale for Intercultural Communicative Competence Among Chinese College Students</i>	<i>Foreign Language World</i>	2013	239	7353
Cai	<i>A Study on the Validity, Reliability and Operability of the Computer-based CET Spoken English Test (CET-SET)</i>	<i>Foreign Language World</i>	2005	185	11946

The first occurrence year of keywords reveals that early keywords already covered assessment-related concepts such as communicative competence, oral tests and validity, consistent with the early emphasis on constructing speaking assessment systems. Later, keywords related to teaching models gradually emerged, and research focus shifted towards classroom practice. This aligns with the thematic evolution of highly cited papers and outlines the overall developmental trajectory of research on foreign language speaking assessment in China.

Table 2 Keyword Frequency and Centrality of Research on Foreign Language Speaking Assessment in China (Ranked by Frequency)

Frequency	Centrality	Year	Keywords
19	0.31	2006	College English
16	0.18	2006	Oral Teaching
11	0.18	2005	Communicative Competence
10	0.13	2005	Oral Test
10	0.2	2005	Validity
8	0.16	2007	Spoken English
8	0.02	2005	Reliability
6	0.08	2008	English Teaching
5	0.07	2016	English Proficiency
4	0.1	2015	Fluency
4	0.03	2008	Countermeasures
4	0.01	2007	Oral Language
4	0.02	2012	Teaching Model
4	0.02	2010	Retelling

4.2 Keyword Cluster Mapping Analysis

Based on Figure 3 and Table 3, current research on foreign language speaking assessment in China demonstrates the following characteristics.

First, research themes are closely interconnected, with overlapping keywords across different clusters. Co-occurrence keywords from the “College English” and “English teaching” clusters largely

intersect with many other clusters, indicating that research on foreign language speaking assessment in China focuses on classroom teaching practice. Judging by the citation years and thematic distribution of clusters, research contexts are predominantly situated at the tertiary level. Research on speaking assessment for primary and secondary education is relatively scarce. This may stem from the long-standing insufficient emphasis on oral English tests within the college entrance examination evaluation system, leading to limited practical experience and research accumulation of speaking assessment at primary and secondary levels.

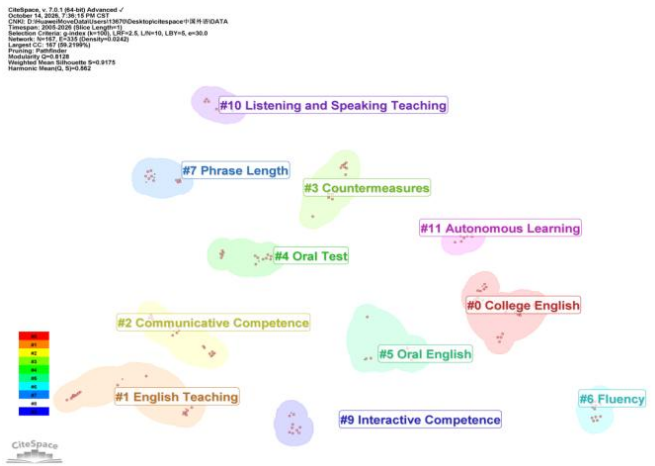


Figure 3 Keyword Clustering Map of Research on Foreign Language Speaking Assessment in China

Second, research perspectives cover both classroom instruction and learners’ autonomous learning pathways. Keywords including “mobile learning”, “oral output”, “interaction pattern”, “cooperative learning” and “autonomous learning” appear in multiple clusters. This shows that scholars have begun to explore diverse ways of oral language learning and learner engagement beyond classrooms. Nevertheless, an overview of cluster themes reveals that discussions on autonomous and interactive learning are mostly embedded within classroom contexts. This indirectly reflects the insufficient supply of in-class oral foreign language practice. Researchers therefore attempt to adopt diverse learning modes to remedy this deficiency in classroom oral practice.

Table 3 Keyword Clusters and Co-occurring Keywords within Each Cluster for Research on Foreign Language Speaking Assessment in China

Cluster ID	Cluster Name	Cluster Size	Citation Year	Co-occurring Keywords
0	College English	30	2013	College English, oral teaching, diversification, computer-based test, human-computer collaboration
1	English Teaching	21	2012	higher vocational colleges, guiding documents, centralized examination, action orientation
2	Communicative Competence	19	2007	validity, reliability, test validity, operability
3	Countermeasures	16	2011	foreign language output, research status, cultivation system, oral acquisition
4	Oral Test	14	2009	teaching, networking, middle school English, CET

5	Spoken English	13	2010	mobile learning, systematization, assessment model, social software
6	Fluency	13	2016	task conditions, disfluency, oral production, achievement level
7	Phrase Length	13	2019	fluency, noun phrases, complexity of oral production, English proficiency
9	Interactive Competence	10	2016	validity verification, interaction mode, exam discourse, turn-taking mechanism
10	Listening and Speaking Teaching	10	2013	junior high school English, language teaching, listening and speaking skills, listening practice
11	Autonomous Learning	8	2009	cooperative learning, second classroom, supervision mechanism, College English

Third, research topics form a dual structure of “assessment quality validation” and “oral production exploration”. Studies in this field can be divided into two themes. The first takes assessment instruments themselves as research objects, with the core task of verifying their effectiveness, feasibility and scientific nature. The second focuses on learners’ speaking performance to uncover inherent patterns of oral production. The “communicative competence” cluster contains keywords such as validity, reliability and test validity, representing the first theme. The “oral test” and “spoken

English” clusters further extend this theme by exploring assessment implementation pathways and evaluation models (Qi et al., 2024; Jin & Jie, 2020). In contrast, clusters of “fluency”, “phrase length” and “interactional competence” belong to the second theme, which mainly investigates micro-level features of learners’ oral production under task conditions (Bao et al., 2025; Zou, 2023). These two themes have distinct priorities and complement each other in research on foreign language speaking assessment.

Overall, current research on foreign language speaking assessment in China forms a research system centred on classroom teaching, alongside examination of speaking assessment quality and exploration of oral production. Although research contexts are mostly confined to higher education and teaching scenarios, which are relatively homogeneous, scholars are gradually exploring new learning modes such as autonomous learning and mobile learning to compensate for inadequate classroom oral practice. In light of these features, future research can further expand speaking assessment studies in primary and secondary contexts and continuously enrich cross-context and multimodal practices of foreign language speaking assessment.

4.3 Keyword Timeline Analysis

The keyword timeline map (see Figure 4) intuitively demonstrates the evolutionary patterns of research themes in spoken English studies, which show four main features.

First, research themes have evolved from broad topics to specific investigations of students’ oral proficiency. Early studies around

2005 mostly focused on general teaching directions such as college English, English teaching and spoken English (Gao, 2008). As research advanced, studies became more refined, gradually branching into specific research directions including communicative competence, oral tests, speaking fluency, phrase length, classroom interactional competence, listening and speaking instruction, and autonomous learning. Research focus is no longer limited to classroom teaching but extends to learners’ actual oral performance and individual ability development.

Second, research on oral evaluation criteria has been continuously refined. As the foundation of evaluation criteria, reliability and validity became key research topics in oral test studies after 2005. Meanwhile, subsequent research is not restricted to oral studies alone. Scholars began to examine various detailed factors affecting oral quality. Indicators that directly reflect speaking proficiency such as oral fluency and interactional competence have remained stable research priorities.

Third, research priorities feature stable core themes and shifting peripheral branches. College English and English teaching have been core research topics over the past two decades. In terms of subfields, from 2005 to 2015, scholars mainly studied spoken English, communicative competence and oral tests. After 2015, research priorities shifted towards learners’ oral performance such as speaking fluency, phrase length and interactional competence (Bao & Wang, 2023; Liu, 2021; Bao et al., 2025). The research focus transformed from “whether oral assessment can be implemented” to “how to evaluate students’ oral quality scientifically and accurately” (Zou, 2023).

Fourth, research perspectives integrate teacher instruction and student autonomous learning, achieving greater comprehensiveness. In recent years, studies on listening-speaking instruction and autonomous learning have emerged, breaking the previous single framework of classroom teaching and examinations. Research scope has expanded to students’ autonomous learning strategies and integrated listening and speaking instruction, reflecting dual research perspectives of teaching and learning (Zhou, 2012).

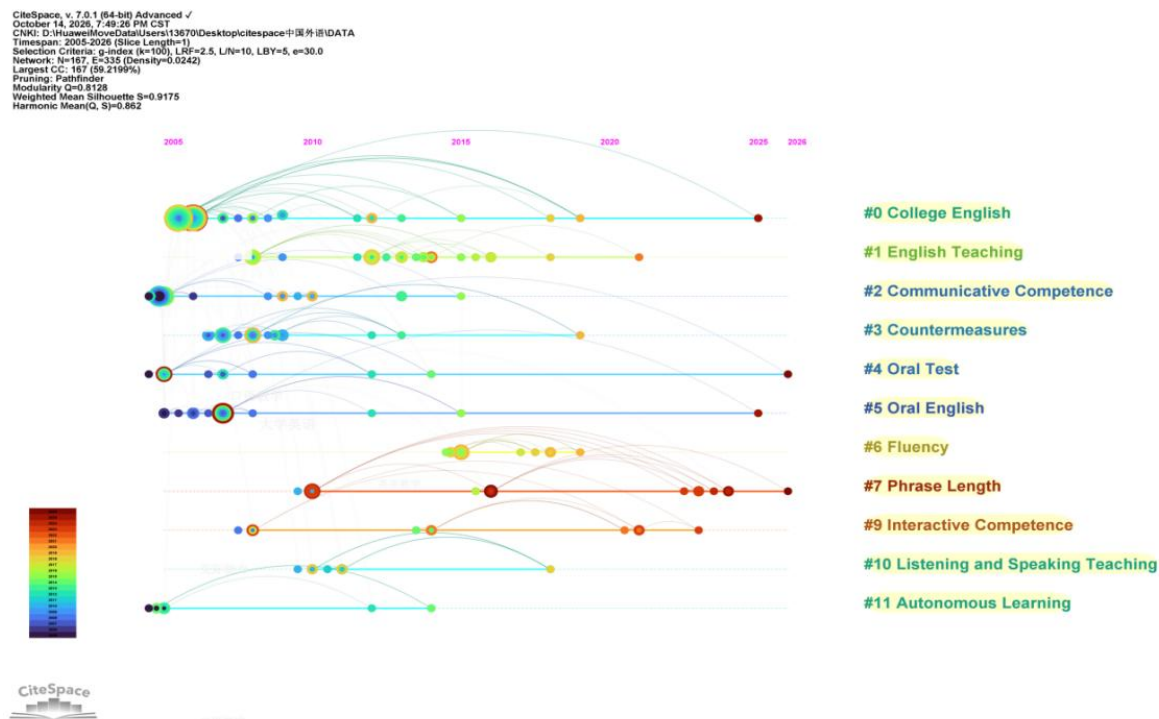


Figure 4 Keyword Timeline Map

According to burst keyword mapping data (see Figure 5), the number of burst keywords retrieved in this study is small and the overall dataset size is limited. The staged changes of research hotspots in this field are relatively straightforward without frequent and complicated hotspot replacements. Existing data shows that reliability and validity were burst keywords in 2005, the earliest research hotspots in this field. Studies at that time focused on the scientificity and rationality of spoken English test instruments, laying a foundation for the construction of spoken English assessment systems. Starting from 2016, “English proficiency” became a burst keyword and its burst period continued until 2026 (the literature collection of this study only ends in September 2026). It has remained a popular research direction in recent years, demonstrating long-standing scholarly interest in the relationship between learners’ overall English proficiency and oral performance.

The burst keywords in 2005 concentrated on basic evaluation criteria for oral tests. Reliability and validity are foundational elements of oral test research, which experienced a research boom only in 2005 and belong to foundational early-stage research. After

2016, research interest in “English proficiency” kept rising, with a burst strength of 2.28 and burst time spanning 2023–2026. It is the core research hotspot at present. Relevant studies mainly explore how learners’ overall English proficiency affects their oral ability.

In general, the developmental trajectory of spoken English research is clear. Research has gradually shifted from verifying the feasibility and scientificity of oral test instruments, to exploring various detailed features of oral production, and finally to investigating multiple influencing factors such as students’ comprehensive English ability and autonomous learning. Due to the small sample of burst keywords in this study, it cannot fully capture the rich changes of research hotspots in each stage. Nevertheless, the field presents a clear research trajectory and definite development trends. Based on available data, it can be inferred that future research will pay more attention to the correlation between students’ English proficiency, oral fluency and interactional competence. Meanwhile, scholars will further explore how autonomous learning can improve learners’ spoken English ability within listening and speaking teaching contexts.

Keywords	Year	Strength	Begin	End	2005 - 2026
Reliability	2005	2.43	2005	2005	■ ————
Validity	2005	2.34	2005	2005	■ ————
English Proficiency	2016	2.28	2023	2026	————— ■■■■

Figure 5 Keyword Burst Map of Research on Foreign Language Speaking Assessment in China

5. Analysis of Research Trends

Based on the overall dataset, research on foreign language speaking assessment in China demonstrates clear periodic fluctuations. Its developmental pattern can be summarized as follows: steady development and accumulation of research findings

in the early stage, reduction in publication quantity with improved research quality and refined research content in the middle stage, and transition to exploration of emerging directions in the later stage. The publication volume changes, research hotspots and priorities in each phase are closely associated with educational policies, industrial progress and technological updates of corresponding periods.

2005–2015 marks the high-yield phase of this field, with relatively high annual publication volume and represents the most active period for studies on speaking assessment. According to the temporal shifts of keywords, research topics in this phase were relatively fundamental and unified. Scholars mainly explored basic issues including oral English instruction, speaking tests, and reliability and validity verification, closely targeting traditional assessment formats such as CET Spoken English Test (CET-SET)(He & Zhang, 2008). Meanwhile, a group of core scholars gradually emerged in this period, laying the preliminary foundational framework for research on foreign language speaking assessment.

During 2016–2019, the overall publication output declined and research popularity cooled down, while research quality improved steadily. Alongside continuous refinement of China's domestic foreign language assessment standards and updated industrial norms, scholarly research priorities shifted accordingly. Instead of merely examining basic testing instruments, researchers began to pay attention to learners' actual oral production and individual proficiency levels. Research perspectives gradually moved from pure assessment studies to refined exploration of language competence, greatly enhancing the practicality and pertinence of relevant studies.

2020–2026 witnesses a further drop in publication scale, and research officially enters a stage of transformation and innovation (the literature retrieval of this study ends only in September 2026). Against the backdrop of educational reform and the development of intelligent technologies in the new era, research in this field breaks the traditional framework of speaking assessment and carries out innovative studies integrated with intelligent technologies and modern educational philosophies(Xu et al., 2026). Judging from cutting-edge keyword hotspots, recent research focuses more on diverse dimensions such as learners' comprehensive English proficiency, autonomous learning and oral interactional competence. It achieves upgrades and expansion of research content and further optimises the overall research landscape.

Taken together with the full dataset, research on foreign language speaking assessment in China consistently keeps pace with national policies and technological advances, featuring a clear and progressive research trajectory. The field has evolved from extensive, fundamental preliminary explorations in the early years to refined, diversified and intelligent in-depth investigations. Nevertheless, scholars in this field mostly conduct research individually or in small groups, and tightly connected research teams have not yet been formed. Future efforts should strengthen academic exchanges and cross-team collaboration to further promote innovative development of foreign language speaking assessment research.

6. Conclusion

This paper adopts CiteSpace to conduct visual mapping analysis of relevant research on foreign language speaking assessment in China. It systematically sorts out and analyses literature in this field from 2005 to 2026 and draws the following conclusions:

- 1) Research on foreign language speaking assessment in China mainly falls into two domains: English teaching and speaking assessment. Over the past two decades, research content in this field has been continuously enriched and the overall research system gradually improved. A cohort of core scholars with long-term dedication and continuous outputs has emerged, providing stable academic support for disciplinary development. However, academic collaboration remains weak on the whole. Most scholars work independently or within small teams; links between different research groups are limited, and an interconnected academic community has not been established.
- 2) The core themes of foreign language speaking assessment focus on English teaching and assessment, while subtopics of branch studies go through three phases. 2005–2015 is the high-output period, focusing on fundamental content of teaching, testing and evaluation, which laid the foundation for domestic research on foreign language speaking tests. Although publication numbers decreased from 2016 to 2019, research shifted toward refined investigation of learners' actual oral competence. 2020–2026 enters the transformation and innovation stage (data collection of this paper ends in September 2026). Combined with ongoing teaching reform and the advancement of artificial intelligence technology, research perspectives extend to diverse areas underpinned by technology and educational reform, such as autonomous learning, technology-empowered assessment and oral interaction.

Overall, research on foreign language speaking assessment in China has evolved from exploration of basic domains toward intelligent, diversified and refined inquiries. Future development can promote intelligent speaking assessment and strengthen academic network collaboration to facilitate continuous in-depth progress of this field.

7. Prospects

In summary, future research directions identified in this paper are as follows:

First, refine research subjects and broaden coverage. Current research on foreign language speaking assessment in China largely concentrates on university students. Future studies may gradually extend to primary and secondary school learners covering students across different age groups. Meanwhile, scholars can explore disparities in oral performance among learners at different English proficiency levels, with a focus on analyzing fine-grained features such as speaking fluency and interactive output. In this way, speaking assessment research can better fit teaching practices at different educational stages.

Second, strengthen collaboration and communication among academic teams. Author collaboration data shows that most scholars in this field carry out independent or small-scale research

with insufficient cross-team connections, and an integrated research system is yet to be built. Future work should boost academic exchanges and joint research between different teams, integrate research resources, break the limitations of single research perspectives, and further improve the systematicness and depth of studies on foreign language speaking assessment.

References

- Bai, X. H., & Wang, K. (2026). Research hotspots and frontier trends of China Europe Railway Express: A CiteSpace visual analysis. *Logistics Research*, (9), 11 16.
- Bao, G. (2025). Research on the use of different types of noun phrases in English learners' oral speech. *Foreign Language Teaching*, 46(3), 19 25. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2025.03.00>
- Bao, G., & Chen, C. L. (2026). A study on the relationship between English learners' proficiency, oral affective performance and lexical complexity. *Foreign Languages and Literature*, 42(2), 178 191. <https://doi.org/10.26922/j.cnki.waiguoyuwen.2026.02.003>
- Bao, G., Lu, A. Q., & Li, L. (2025). A study of noun phrase length in oral production by EFL learners and native speakers. *Foreign Language Research*, 42(5), 78 84. <https://doi.org/10.13978/j.cnki.wywj.2025.05.007>
- Bao, G., & Wang, Y. Q. (2023). A factor structure study of disfluency in English learners' story retelling. *Foreign Languages Research*, 40(6), 60–65, 77. <https://doi.org/10.13978/j.cnki.wywj.2023.06.010>
- Cai, J. G. (2005). A study on validity, reliability and operability of computer based CET SET. *Foreign Language World*, (4), 66 75.
- Cheng, Y. P., He, Q. G., & Han, X. L. (2007). Impacts of English classroom anxiety on college students' oral production. *Language Teaching and Linguistic Studies*, (1), 89 95.
- Gao, B. L. (2008). A review of research hotspots in college oral English testing. *Foreign Language Teaching and Theoretical Practice*, (1), 92 96.
- He, L. Z., & Zhang, J. (2008). A reliability study of CET SET under the many facet Rasch model. *Modern Foreign Languages*, (4), 388 398+437.
- Jin, Y., & Jie, W. (2017). Principles and methods for developing the Speaking Scale of China's Standards of English. *Foreign Language World*, (2), 10–19. <https://doi.org/10.26971/j.cnki.flw.1004-5112.2017.02.002>
- Jin, Y., & Jie, W. (2020). Research on the application and washback effect of language proficiency scales: Taking the Speaking Scale of China's Standards of English as an example. *Foreign Language World*, (3), 52 60. <https://doi.org/10.26971/j.cnki.flw.1004-5112.2020.03.008>
- Li, S. (2026). Logical framework and realization path of digital intelligence empowered teaching learning assessment integration. *Journal of Nanjing Normal University (Social Science Edition)*, (5), 57 66.
- Li, X. (2026). A study on the influence of task complexity on college students' oral English acquisition. *Journal of Jiamusi University (Social Science Edition)*, 44(6), 196 200.
- Li, X. H. (2026). Research on the theoretical system construction and internal correlation of English core literacy cultivation under the background of new liberal arts. *Journal of Hubei Open Vocational College*, 39(17), 53 54+66.
- Liu, J. D., & Yang, M. Z. (2021). The application of China's Standards of English in language assessment. *Foreign Language Testing and Teaching*, (2), 1 11. <https://doi.org/10.26970/j.cnki.fltt.1004-5112.2021.02.001>
- Liu, L. D. (2021). Turn taking behaviours of Chinese EFL learners in paired oral tests. *Foreign Language Research Journal*, (3), 58 64. <https://doi.org/10.16263/j.cnki.23-1071/h.2021.03.010>
- Liao, H. J., & Li, L. W. (2025). Framework construction and empirical validation of college English intercultural oral communication MOOC from the perspective of cognitive agency development. *Computer Assisted Foreign Language Education*, (6), 34 39+105. <https://doi.org/10.20139/j.issn.1001-5795.20250605>
- Qi, L. Y., Di, A. Y., Liu, X. G., & Zhou, N. (2024). A study on the development and validation of an oral English test based on China's Standards of English. *China Examinations*, (5), 79 89. <https://doi.org/10.19360/j.cnki.11-3303/g4.2024.05.008>
- Qi, P., & Shi, X. C. (2016). Curriculum design and effect research of English audio visual and speaking course based on POA. *Academic Monthly of Education*, (8), 106 111. <https://doi.org/10.16477/j.cnki.issn1674-2311.2016.08.015>
- Shi, T. H. (2013). Quantitative assessment of college oral English fluency and teaching suggestions. *Journal of Panzhuhua University*, 30(5), 86 88.
- Xu, C. Y., Zhu, L., Wang, Y. P., & Liu, C. H. (2023). Effects of foreign language speaking anxiety on verbal interaction quality: An fNIRS hyperscanning study. *Acta Psychologica Sinica*, 55(12), 1949 1965.
- Xu, Y., Zhang, B. X., & Li, X. D. (2026). Practice and reflection on artificial intelligence in large scale oral English testing. *Foreign Language Research Journal*, (1), 70 76. <https://doi.org/10.16263/j.cnki.23-1071/h.2026.01.011>
- Yin, S. K., & Jin, Y. (2022). The influence of critical thinking on integrated oral test performance. *Foreign Languages and Their Teaching*, (6), 66 77+146 147. <https://doi.org/10.13458/j.cnki.flatt.004905>
- Zhang, L. L. (2017). Research on the teaching effectiveness of Production Oriented Approach. *Modern Foreign Languages*, 40(3), 369 376+438.
- Zhang, X. (2007). On organizational strategies for cooperative learning in oral English classrooms. *Foreign Languages and Their Teaching*, (8), 38 41. <https://doi.org/10.13458/j.cnki.flatt.003588>
- Zhong, H., Bai, Q. H., & Fan, W. W. (2013). A pilot study on constructing a self rating scale for Chinese college students' intercultural communicative competence. *Foreign Language World*, (3), 47 56. <https://doi.org/10.26971/j.cnki.flw.1004-5112.2013.03.007>
- Zhou, Q. S. (2019). Seventy years of research on China's language policy. *Journal of Xinjiang Normal University*

(Philosophy and Social Sciences Edition), 40(6),60 71+2.
<https://doi.org/10.14100/j.cnki.65-1039/g4.20190713.001>

28. Zhou, Y. (2012). Design and practice of a micro blog based mixed teaching model for college English oral presentations. *E Education Research*, 33(6), 100 105+116. <https://doi.org/10.13811/j.cnki.eer.2012.06.002>
29. Zou, R. (2023). A validity study of the paired oral test rating scale based on the Rasch model. *Journal of PLA University of Foreign Languages*, 46(5), 20 29.