



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



An Exploration of Research Hotspots and Trends in International Multi-modal Language Assessment: A Visual Analysis Based on CiteSpace

Jiayao Chen¹, Weijie Hu^{2*}

¹ MA candidate, 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: 10.10.2026 | Accepted: 15.10.2026 | Published: 19.10.2026

*Corresponding author: Weijie Hu

Abstract

Multimodal language assessment is an emerging direction in language assessment, yet existing reviews focus mainly on pedagogical dimensions, leaving systematic reviews of assessment largely unexplored. Based on 283 screened Web of Science papers (2022–2026), this study uses CiteSpace keyword clustering and burst analysis, along with publication trend statistics, to examine the research landscape, thematic structure, and hotspots evolution. Findings show that no stable core author group has emerged, collaborative networks remain loose, and interdisciplinary dialogue is still nascent. Thematically, research clusters into three areas: “Research on Multimodal Input, Cognitive Processing, and Language Acquisition,” “Research on Multimodal Output, Interactive Competence, and Identity Construction,” and “Research on Multimodal Assessment and Clinical Language.” Hotspots evolution has moved through three phases: fundamental exploration and theoretical expansion - framework construction and empirical application - technology-driven and in-depth analysis. Future research should conduct close readings of high-burst, high-centrality publications and employ qualitative methods (e.g., interviews, think-aloud protocols) to link macro patterns with micro mechanisms.

Keywords: International multimodal language assessment; CiteSpace; visualized analysis; research landscape; thematic structure; evolution of research hotspots

1. Introduction

The rise of multimodal language assessment is rooted in evolving contemporary communicative practices. With digital technologies, meaning construction now involves diverse symbolic resources—images, videos, gestures, and spatial arrangements. In their editorial review for the special issue on multimodal constructs in *Language Assessment Quarterly*, Yu and Clark (2025) explicitly noted that multimodality has become salient trends in language assessment; digital technologies have reshaped contemporary life and rendered many assessment frameworks inadequate for capturing emerging learning modes. Grapin (2022) further argued, from a social semiotic perspective, that multimodal assessment extends traditionally privileged linguistic paradigms to encompass multimodal communication theories, requiring assessors to re-examine linguistic constructs. Under this background, multimodal language assessment requires examinees to use both linguistic (text, speech) and non-linguistic modalities (images, videos, gestures, spatial arrangements) to complete tasks, simulating the collaborative operation of symbolic systems in daily communication.

Research on multimodal language studies traces back to Kress's (2009) social semiotic tradition and the New London Group's "Multiliteracy" concept; however, early work focused on the teaching and learning, with assessment explored later. Whithaus (2014) examined multimodal writing among first-language writers; Early et al. (2015) conceptualized multimodality as a form of social practice and cultural production within language education; Crawford Camiciottoli and Campoy-Cubillo (2018) explored how multimodality enhance English language assessment in higher education; Tan et al. (2020) first integrated evaluation as a core focus, concentrating on multimodal literacy assessment across language, science, and cross-cultural emotion; Yi (2020) investigated processes and products of multimodal writing tasks; and Peters and Munoz (2020) examined multimodal input's impact on second-language acquisition. Recently, assessment has shifted from margin to center. Palmour (2024) revealed a problem of "insufficient construct representativeness" in EAP oral presentation assessments: grading criteria and teacher assessors often overemphasized the linguistic aspects of oral performance while neglecting the multimodal communicative resources that students actually mobilize during their presentations. This highlights the core contradiction in multimodal oral assessment: when examinees naturally employ gestures, eye contact, and spatial layout, these elements remain excluded from the scoring criteria's construct framework.

Despite growing empirical research, reviews in multimodal language assessment remain scarce. Existing reviews focus on pedagogy or learning process: Lim, Toh, and Nguyen (2022) reviewed multimodal teaching methods in English classrooms; Han, McNaughton, and Zhang (2026) examined teacher involvement in digital multimodal writing; and Perez and Cecilia (2025) explored social media and multimodal content creation's impact on English learning motivation. These reviews share a focus on "how to teach" and "how to learn," rather than "how to assess." In terms of assessment, the most direct review is Yu and Clark's (2025) editorial, which systematically organized nine empirical studies featured in a special issue and proposed a core list of research questions for multimodal assessment—including whether multimodality constitutes a supplementary component or

an integral part of language assessment constructs, the cognitive processes underlying multimodal assessment tasks, the development and validation of assessment criteria, and the role of artificial intelligence in multimodal assessment. However, it is a special issue introduction, not a rigorous systematic review; it lacks systematic search and coding procedures and quantitative analysis of the research landscape and its evolution. This gap leaves researchers without a holistic understanding of the knowledge structure, hotspots, and trends in multimodal language assessment. Although Chinese scholar Zhang Delu (2018) has provided a systematic exposition of multimodal discourse analysis theory, he has likewise not conducted a quantitative review of the empirical research landscape in multimodal language testing.

Based on the mentioned research background and existing gaps, this study focuses on the overall developmental trajectory of international multimodal language testing over the past five years, aiming to address three primary research questions:

- 1) What characteristics have defined the research landscape of international multimodal language testing in the past five years? What features do its core author clusters, author collaboration networks, and disciplinary distribution exhibit?
- 2) What logical connections exist among the research themes within this field?
- 3) What stage-specific characteristics and evolutionary trajectories have emerged in the evolution of research hotspots in international multimodal language assessment? What are the underlying causes of these developments?

Theoretically, this study applies a systematic review methodology to conduct a comprehensive knowledge graph analysis of multimodal language assessment, addressing the gap in quantitative review analysis in this field. Through CiteSpace clustering and burst detection, it reveals both the static knowledge structure and dynamic evolutionary trajectory, providing a reference knowledge graph for future research. Practically, the core themes and emerging trends identified can guide task design, scoring criteria development, and validity validation in multimodal language testing. Given the growing maturity of artificial intelligence and multimodal learning analytics, this research holds significant contemporary relevance and forward-looking value.

2. Research Design

The design of this study comprises: describing the data sources and screening procedures; providing the research tools and their application methods; and detailing the implementation of the specific research methods.

2.1 Data Sources and Screening

All literature samples for this study were retrieved from Web of Science. In line with the academic development trajectory of international multimodal language assessment research, this study limited the search period to the past five years—from 2022 to 2026. The search focus was centered on "multimodal language assessment"; therefore, the specific search query was formulated as "Topic= 'multimodal language assessment' OR 'multimodal language testing'".

To ensure that the analyzed samples were highly relevant to the research topic, strict filters were applied to both the type of

literature and its source during the literature search phase. This approach aimed to exclude non-research-oriented publications, low-relevance works, and duplicate literature that could potentially distort the research findings, adhering to three guiding principles: academic rigor, relevance, and uniqueness.

First, a uniqueness screening was conducted to exclude literature that had been previously published, was submitted to multiple journals simultaneously, or contained similar content; next, an academic screening was performed to filter out non-academic materials—such as book reviews, meeting minutes, and news reports—retaining only professional academic journal articles; finally, a relevance screening was carried out by verifying titles, reviewing abstracts, and sampling portions of the full texts to exclude literature from unrelated fields, thereby ensuring the validity of the selected literature.

After rigorous screening, 283 valid papers were ultimately selected to serve as the sample for this study.

2.2 Research Tools

This study employs CiteSpace6.4.R1 as the core tool for visual analysis, supplemented by Microsoft Excel for data processing, to construct knowledge graphs and conduct statistical analyses of the data, respectively.

CiteSpace is a specialized software for literature data analysis, enabling users to generate scientific knowledge graphs from extensive literature datasets. These graphs utilize visual representations to illustrate the knowledge structure, associative patterns, and evolutionary trends within a specific research field. CiteSpace offers significant advantages in knowledge graph construction, cluster analysis, and keyword burst detection, providing an intuitive visualization of research hotspots and developing trajectories in the target domain.

Microsoft Excel is primarily employed for statistical aggregation of basic data from sample literature, such as the distribution of publications across educational stages and the total number of publications.

2.3 Research Methods

This study employs three research methods: descriptive statistical analysis, keyword cluster analysis, and burst word detection analysis.

Descriptive statistical analysis is conducted using Microsoft Excel to perform quantitative analyses on 283 samples, with a particular focus on examining the distribution characteristics of active authors and their works. Simultaneously, the Price's Law is applied to calculate the threshold for core authors in this field, thereby clarifying the composition of the core research community and, consequently, delineating the overall research landscape of international multimodal language assessment research.

Keyword cluster analysis is conducted using the CiteSpace to analyze keyword cluster distributions; high-frequency function words lacking substantive significance are filtered out, and the node filtering criterion is set to “g-index” with “k=25”.

Using the burst word detection function of CiteSpace, this study identifies high-burst-intensity keywords in the field of international multimodal language assessment research from 2022 to 2026, thereby delineating the research frontiers and hotspots across different periods. Complemented by statistical analysis of publication counts over the past five years, conducted using Microsoft Excel, this approach clearly illustrates the emergence and evolutionary dynamics of research hotspots, while systematically mapping the developing trajectory and academic rationale of this field.

3. Research Findings

Based on the analysis of the knowledge maps generated by CiteSpace, this study addresses three key research questions by conducting quantitative analyses from three perspectives: the research landscape, the thematic structure, and the evolution of research hotspots.

3.1 Research Landscape

The research landscape of international multimodal language assessment research exhibits the following characteristics: limited concentration of authors, dispersed overall collaboration, incomplete core group of leading authors, and insufficient interdisciplinary and cross-domain dialogue; these trends are particularly evident in the author group features and disciplinary distributions.

This study employs Price's Law to analyze the author distribution characteristics and delineate the core author group. The law states that within the same subject area, half of the papers are authored by a group of highly productive authors, the number of whom is approximately equal to the square root of the total number of authors. The mathematical expression is: $m \approx 0.749\sqrt{N_{\max}}$, where “ $N_{\max}$ ” represents the number of papers authored by the most productive author, and “ m ” denotes the minimum number of papers authored by the core authors.

In these samples, the authors with the highest number of publications are Clark, Tony; Balaman, Ufuk and Jiang, Lianjiang, each having a maximum publication count of $N_{\max}=3$. After calculation, $m = 0.749 \times \sqrt{3} \approx 1.297$, and upon rounding up, $m=2$. This indicates that in this research field, authors who have published at least two papers can be defined as core authors. Statistical analysis reveals that there are 24 core authors, accounting for 17.3% of the total number of authors (138); their combined publication count is 69, representing 24.3% of the total publication count (283). From these data, it can be concluded that a core author group has not yet formed in this field, the research efforts are relatively dispersed, and the 50% threshold of Price's Law has not been met.

Meanwhile, the author co-occurrence network constructed using CiteSpace (Figure 1) reveals that the network consists of $N=138$ nodes, $E=54$ edges, and a network density of 0.0057. The figure shows only a few small collaborative clusters—such as Clark and Tony, and Yu and Guoxing—serving as central research groups, indicating that a broad, tightly-knit academic collaboration network has not yet emerged in this field; instead, the landscape is characterized by small collaborative clusters and overall relative dispersion.

Figure 1 : Author Co-occurrence Network



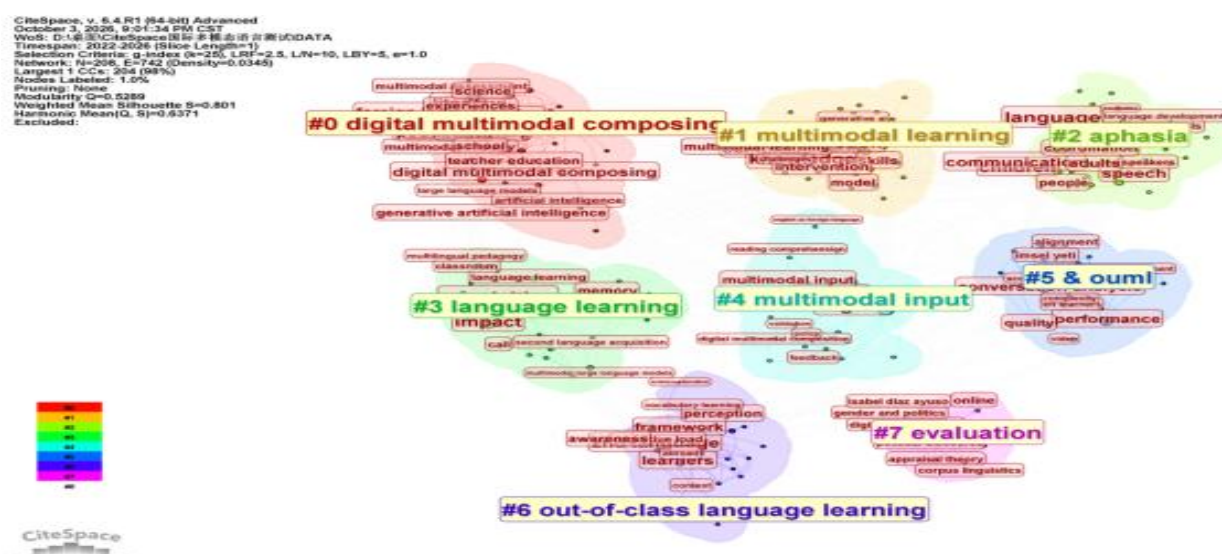
In terms of disciplinary distribution, the 283 selected papers are predominantly centered on linguistic research, with educational research serving as a key focus, while psychology, computer science, and neuroscience exhibit collaborative interdisciplinary intersections. Linguistic research accounts for the largest share at approximately 72%, primarily addressing topics such as multilingual education, language neurocognition, discourse and pragmatics, and phonology and phonemic variation. Educational research follows with approximately 68% of the total, frequently co-occurring with linguistic research; their overlapping areas include multilingual education, science education, autonomous learning, and reading ability cultivation. Additionally, the shares of studies in psychology, neuroscience, and computer science are approximately 7%, 3.1%, and 1.7%, respectively. Overall, the international multimodal language assessment remains primarily driven by linguistic and educational research, whereas studies from other applied disciplines remain relatively scarce.

3.2 Thematic Structure

In this study, we employed CiteSpace to conduct a cluster analysis of the keywords from the samples (Figure 2), utilizing the LLR algorithm (Log-Likelihood Ratio) and combining modularity with the silhouette coefficient to validate the quality of the clusters. The results showed a cluster network modularity of $Q=0.5289$ and a weighted mean silhouette of $S=0.801$, indicating a robust cluster structure, high intra-cluster homogeneity, and reasonably reliable clustering outcomes.

Through keyword clustering analysis, research in this field can be categorized into eight core themes: digital multimodal composing, multimodal learning, aphasia, language learning, multimodal input, ouml, out-of-class language learning, and evaluation. Among these, the theme of digital multimodal composing has the most extensive body of research findings and constitutes the core focus of current research in this domain; the other themes provide complementary perspectives from diverse dimensions.

Figure 2 : Keyword Clustering Map



Based on the inherent logic and interrelationships of the cluster labels, the eight clusters mentioned above can be distilled into three major themes, each encompassing its corresponding cluster modules, thereby establishing a clear cluster architecture.

The first research theme, titled “Research on Multimodal Input, Cognitive Processing, and Language Acquisition,” comprises Cluster 1 (multimodal learning), Cluster 3 (language learning), Cluster 4 (multimodal input), and Cluster 6 (out-of-class language learning), with a total number of nodes equal to 100. Among these clusters, Cluster 1 has the largest number of nodes, yet its silhouette coefficient is only 0.69, which is relatively lower compared to the other clusters within the same network. Clusters 3, 4, and 6 encompass language learning contexts both within and outside the classroom, collectively forming the logical chain of “input-processing-acquisition”. The literature in Cluster 1 provides empirical support for both the learning process and instructional intervention, thereby establishing a comprehensive research framework that spans from input design to cognitive processing and ultimately to learning outcomes. Overall, the published literature in this theme offers cognitive mechanistic explanations for multimodal language teaching, advancing the field’s evolution from a focus on “material presentation” to an emphasis on “learner processing.” Centering on key constructs—multimodal input, cognitive load, memory, vocabulary acquisition, and reading comprehension—this theme examines how multimodal input influences the comprehension, memory retention, and acquisition outcomes of second-language learners. It also incorporates methodological considerations such as mixed methods, feedback mechanisms, and validation procedures, thereby laying a solid theoretical foundation for future research on task design and cognitive validity in multimodal languageassessment.

The second research theme, titled “Research on Multimodal Output, Interactive Competence, and Identity Construction,” comprises Cluster 0 (Digital Multimodal Composing) and Cluster 5 (Conversation Analysis), with a total number of nodes equal to 62. Among these, Cluster 0—Digital Multimodal Composing—features the largest number of nodes and a high silhouette coefficient, serving as the core anchor point within the multimodal writing research domain; Cluster 5, on the other hand, focuses on conversation analysis and interactive competence. Together, these two clusters constitute the “Output-Interaction” framework, acting as critical links between language learning and assessment. In terms of scholarly contributions, this theme has provided methodological pathways for evaluating multimodal language output and interactive competence, thereby facilitating the evolution of this field from “monomodal writing” toward

“multimodal meaning construction.” The theme centers on key research topics—including digital multimodal writing, generative artificial intelligence, learner identity, peer interaction, teacher education, conversation analysis, interactive competence, negotiation, and alliance formation—focusing on learner identity construction and the mobilization of interactive resources. It also encompasses methodological approaches such as multimodal conversation analysis and peer feedback, offering direct methodological references for the subsequent design of multimodal language assessment tasks and the development of scoring criteria for interactive competence.

The third theme, titled “Research on Multimodal Assessment and Clinical Language,” comprises Cluster 2 (Aphasia) and Cluster 7 (Evaluation), with a total number of nodes equal to 38. Among these, Cluster 2 (Aphasia) features a larger number of nodes and a higher silhouette coefficient, whereas Cluster 7 focuses on broad categories of evaluative behaviors, such as political discourse and digital impoliteness. Together, these two clusters constitute a dual “clinical-social” assessment dimension, serving as crucial support for expanding the boundaries of multimodal evaluation research. In terms of content, the literature within this thematic category provides a dual perspective—combining clinical and social analysis—for multimodal assessment, thereby facilitating the evolution of this field from “classroom language testing” to “broad-based language performance assessment.” Centered on core topics including aphasia, language development, communication, coordination, assessment, political discourse, digital impoliteness, and corpus linguistics, this thematic category examines multimodal manifestations in language disorders and discourse evaluation. Simultaneously, it encompasses methodological considerations from both clinical linguistics and corpus linguistics, offering a cross-disciplinary reference framework for subsequent validity verification and the development of scoring criteria in multimodal language assessment.

3.3 Hotspots Evolution

Building upon the keyword temporal distribution maps and burst keyword detection capabilities of CiteSpace 6.4.R1, this study conducted a temporal analysis of core literature from international multimodal language assessment research over the past five years, identifying 13 keywords with high burst intensity (Figure 3). In conjunction with statistics on the number of publications (Figure 4), the research hotspots in this field can be divided into three distinct phases, collectively exhibiting a developmental trajectory of “fundamental exploration and theoretical expansion - framework construction and empirical application - technology-driven and in-depth analysis.”

Figure 3 : Keyword Bursting Map

Top 13 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2022 - 2026
science	2022	1.81	2022	2023	
communication	2022	1.3	2022	2023	
learners	2022	1.24	2022	2023	
adults	2022	1.09	2022	2023	
aphasia	2022	0.9	2022	2023	
awareness	2022	0.62	2022	2024	
speech	2022	1.7	2023	2024	
impact	2023	1.33	2023	2024	
framework	2023	0.93	2023	2024	
abroad	2023	0.86	2023	2024	
multimodal input	2024	1.28	2024	2026	
skills	2024	1.28	2024	2026	
multimodal learning analytics	2024	1.02	2024	2026	

Figure 4 : Publication Volume Statistics Chart



The first phase was the foundational exploration and theoretical expansion period (2022-2023). This phase marked the initial stage of international multimodal language assessment research, with the number of published articles rising from 25 in 2022 to 28 in 2023—a slow increase within a relatively small overall research base. During this phase, certain keywords emerged prominently, albeit with relatively low frequency, short duration, and limited scope; these keywords primarily included “science,” “communication,” “learners,” “adults,” “aphasia,” and “awareness.” The research perspectives were diverse, with the primary objective being to establish the existence and recognition of “multimodality” in specific populations (adults) or within basic science education. The prominent keywords in this period (e.g., “aphasia,” “speech”) also corroborated the early activity of Cluster 2 (clinical language disorders) in the cluster map.

The second phase encompasses the framework construction and empirical application period (2023-2024). This phase marks an accelerated stage of this field, during which the number of publications in 2024 surged to 40—representing an approximately 43% increase over the previous year—and the publication trajectory began to rise. In this phase, there was a marked shift toward concepts that are more operational and empirically grounded, with increased research intensity; key focus areas included “speech,” “impact,” “framework,” and “abroad.” Researchers no longer limited themselves to merely “discovering” multimodal phenomena; instead, they began to explore its implications and endeavor to establish theoretical frameworks. Concurrently, the emergence of the term “abroad” signaled a growing focus on multimodal applications in cross-cultural or overseas learning environments. This phase serves as a bridge connecting early clinical and foundational research with later technology-driven research.

The third phase, the technology-driven and in-depth analysis period (2024-2026), is characterized by an exponential surge in the number of publications: the number of publications surge to 84 in 2025 (doubling from 2024) and reach 106 in 2026. During this phase, emerging keywords exhibit strong technical and cutting-edge characteristics, and their emergence persists continuously until 2026, indicating that the current research landscape is in a vibrant phase of development, with primary focus areas including

multimodal input, skills, and multimodal learning analytics. Research in this phase not only emphasizes outcomes but also places greater emphasis on the acquisition of procedural skills. This trend aligns closely with the high activity levels observed in clusters such as Cluster 0 and Cluster 4 within the cluster map.

Based on the burst keyword detection results, the prominence of topics such as multimodal input, skills, and multimodal learning analytics has significantly increased in recent years. Combined with statistics on the number of publications per year, these topics are emerging as new frontiers of research in this field. Given current research gaps, further in-depth exploration in these directions can be pursued in the future.

4. Research Analysis

Building upon the quantitative analysis in the preceding section—which examines the research landscape, thematic structure, and evolving hotspots of international multimodal language assessment over the past five years—this analysis section systematically addresses the three research questions posed in the introduction. Its aim is to elucidate the intrinsic logic and underlying drivers of academic development in this field, thereby addressing the current gap in existing research regarding the patterns of its evolution and the intrinsic interconnections within its knowledge framework.

4.1 Deconstruction and Formative Logic of the Research Landscape

Based on an analysis of authors’ co-occurrence networks and disciplinary distribution patterns, the research landscape of international multimodal language testing over the past five years has exhibited the characteristics of “concentration of authors within limited circles and dispersed overall collaboration,” while interdisciplinary and cross-domain dialogue has yet to mature.

Existing bibliometric studies have shown that the level of international cooperation in language assessment research remains relatively low, at only 12.44%, and that research efforts exhibit a fragmented pattern (Nursipah Hidayah et al., 2026). This landscape is closely linked to the interdisciplinary nature of this field: researchers are dispersed across various traditional disciplinary domains—including language testing, second language acquisition, educational technology, conversation analysis, and clinical

linguistics—each employing distinct theoretical frameworks and methodological approaches, yet a unified core group of authors has yet to emerge.

From a theoretical perspective, the academic foundations of this field span multiple disciplinary traditions. The cognitive psychology tradition, exemplified by Mayer’s cognitive theory of multimedia learning, provides an explanation of the processing mechanisms underlying multimodal input design; the social semiotics and systemic functional linguistics traditions, centered on Kress’s (2009) multimodal theory, emphasize the equal status of visual and auditory symbols—beyond language—in meaning construction (Yu, Guoxing & Clark, 2025); and the conversation analysis tradition offers methodological tools for assessing interactive competence (Dai & David Wei, 2025). In the editorial note for the multimodal special issue of *Language Assessment Quarterly*, Yu and Clark (2025) point out that the core theoretical tension in multimodal language assessment lies in whether multimodal elements constitute “supplementary support components” or “indivisible components” of language assessment constructs—a definitional issue that has yet to reach consensus within the academic community, thereby objectively hindering theoretical dialogue and collaboration among researchers.

The fragmentation at the methodological level further exacerbates the fragmentation of collaborative efforts. Research in multimodal language testing employs diverse technical approaches—including eye-tracking, speech analysis, behavioral coding, conversation analysis, and corpus-based methods—where researchers are often deeply rooted in their respective methodological traditions. This characteristic is also evident in research within the domain of multimodal learning analysis: a systematic review of computer-supported collaborative learning contexts revealed that 89% of empirical studies were published after 2018, with researchers utilizing heterogeneous data types (e.g., audio, video, eye-tracking data, and physiological sensors) and employing diverse analytical methods (e.g., time-series analysis, process mining, and deep learning). The “lack of standardized models and frameworks” is regarded as one of the core challenges in multimodal learning analysis research (Ouyang, Fan & Xianping Bai 2026). Although multimodal language assessment shares technical overlaps with multimodal learning analysis, the two fields belong to distinct academic communities, differ in their journal publication platforms and academic conferences, and thus face limited avenues for cross-community dialogue.

The dispersion across academic disciplines also constitutes a structural factor constraining the deepening of collaborative efforts. An examination of the journal distribution in existing literature reveals that research on this field is scattered across journals with diverse orientations, such as *Language Assessment Quarterly*, *Language Learning & Technology*, *Journal of Second Language Writing*, and *Computers and Composition*. Although the papers indexed by *Language Assessment Quarterly* in 2025 focused on the theme of multimodal assessment, their research contexts spanned classroom assessment, high-stakes testing, and digital multimodal writing—indicating a lack of an integrative framework for multimodal assessment research that bridges diverse contexts and skills. Consequently, researchers have struggled to establish sustained collaborative networks.

4.2 Intrinsic Interrelationships and Constructive Drivers of the Thematic Framework

In this study, through keyword cluster analysis, research on international multimodal language assessment conducted over the past five years is distilled into three major themes: “Research on Multimodal Input, Cognitive Processing, and Language Acquisition,” “Research on Multimodal Output, Interactive Competence, and Identity Construction,” and “Research on Multimodal Assessment and Clinical Language.” These three clusters exhibit a progressive relationship, and the high stability ($Q=0.5289$) and high homogeneity ($S=0.801$) of the cluster results collectively validate the scientific rigor of this knowledge framework.

The intrinsic interrelationships within the thematic structure can be understood through three hierarchical dimensions: “cognitive processing—interactive construction—evaluative extension.” The three major thematic modules distilled from the cluster map constitute a progressive knowledge system centered on the questions: “How do learners process multimodal data? How do they utilize it? And how is it evaluated?” The rationale for constructing this framework stems from the multimodal turn in second language acquisition theory, the expanding needs for conceptualizing language assessment constructs, and the reshaping of language use patterns driven by digital technologies.

The cognitive processing module constitutes the foundational layer of this domain, with its core focus being: when linguistic input is simultaneously presented across multiple modalities—such as text, images, audio, and video—how learners allocate their attention, how they integrate this information, and how such information integration influences ultimate acquisition outcomes. The theoretical foundation of this module primarily draws upon Mayer’s Cognitive Theory of Multimedia Learning (CTML) and Paivio’s Dual Coding Theory (DCT), both of which emphasize that when information is presented synchronously via visual and auditory channels, memory and comprehension outcomes are superior to those achieved with single-channel input (Pu, Chang, and Wang, 2024). This theoretical framework has been consistently validated in empirical research; for instance, Webb and Chang (2022) found that the “reading–listening” paradigm significantly outperformed pure reading or pure listening in the acquisition of collocations, owing to the fact that audio cues enable collocations to be perceived as integrated units rather than being processed word-by-word. However, cognitive load theory highlights a boundary: the amount of multimodal input is not necessarily proportional to its benefits; redundant modalities—such as the simultaneous presentation of identical information in both text and subtitles—may increase cognitive load and, conversely, hinder learning (Sweller, 2005). Consequently, the rationale for developing this module is twofold: on one hand, the trend toward “multimodality” driven by technological advancements; on the other hand, the constraints imposed by cognitive science regarding the optimal choice of modalities.

This section on interactive construction bridges cognition and assessment, focusing on the core question: how learners mobilize both linguistic and non-linguistic resources through multimodal output and interaction to construct meaning and demonstrate competence. Digital Multimodal Composing (DMC) research, grounded in social semiotics and multi-modal literacy theory, redefines writing as a cross-modal meaning-design process rather than mere textual encoding (Jiang, L. & Hafner, C. 2024). Its primary impetus stems from evolving communicative practices in the digital era: as meaning on screens is increasingly co-

constructed through diverse modalities—such as images, layout, hyperlinks, and short videos—evaluating writing solely based on textual quality can no longer adequately reflect authentic communicative competence (Hafner & Ho, 2020). Concurrently, research on L2 Interactive Competence (L2 IC) employs multimodal conversation analysis as its methodological framework to reveal how learners, in authentic interactions, leverage resources—including turn-taking, eye contact, and gestures—to co-construct intersubjectivity. The rationale for developing this section lies in a reconceptualization of the notion of “competence”: shifting from individual cognitive abilities to distributed competencies within interactive contexts, and from a language-centered approach to a synergistic integration of language, body, and technology.

The evaluation of extended assessment domains represents both the frontier of current development and an inevitable outcome of applying multimodal language assessment in practical settings. Recent research on functional communication assessment for aphasia has explicitly established “multimodality,” “interactivity,” and “contextualization” as three core criteria for effective assessment—aligning with ongoing discourse within the field of language testing regarding the “multimodal construct” (Doedens & Meteyard, 2020). However, both approaches face the same fundamental challenge: when the assessment target expands from a single language skill to multimodal interactive performance, how can assessment criteria maintain both reliability and construct validity? The debate in a special issue of *Language Assessment Quarterly* stems from internal reflections within language testing scholarship regarding the tradition of “construct validity”: if authentic communication itself is inherently multimodal, can assessments focusing solely on linguistic modalities still possess ecological validity (Yu, Guoxing & Clark, 2025)?

The intrinsic connections among the three thematic modules can be summarized as a logical chain of “condition - process - outcome”: the cognitive processing module addresses “how multimodal input is processed” (condition), the interactive construction module addresses “how multimodal competence is applied” (process), and the assessment and extension module addresses “how multimodal performance is evaluated” (outcome). The rationale for constructing this logical chain does not stem from any single theoretical tradition, but rather from the convergence of cognitive and sociological perspectives in second language acquisition within the context of technological transformation.

4.3 Evolutionary and Driving Characteristics of Research Hotspots

Based on burst keyword detection, sequential analysis, and publication volume statistics, this study delineates the evolutionary trajectory of international multimodal language assessment research into three progressive phases: the foundational exploration and theoretical expansion phase (2022-2023), the framework construction and empirical application phase (2023-2024), and the technology-driven and in-depth analysis phase (2024-2026). Across these three phases, the number of published studies has grown steadily, reflecting a continuous expansion in both the breadth and depth of research efforts.

One of the core drivers behind this evolution is the growing need for deeper theoretical construction. Systemic Functional Linguistics (SFL) and social semiotics provide foundational theoretical frameworks for multimodal research; Kress’s (2009)

discourse on language as a social semiotic system has been widely cited in DMC studies. Simultaneously, cognitive load theory and multimedia learning theory offer cognitive mechanistic explanations for multimodal task design, thereby accelerating the development of the bursting word “multimodal input” from 2024 to 2026.

Technological empowerment constitutes another crucial driving dimension. The exponential growth in the volume of scholarly publications has occurred in close temporal alignment with the maturation of generative artificial intelligence and multimodal learning analysis technologies. Research by Chen et al. (2026) demonstrates that EFL oral language learning driven by multimodal GAI can reveal learners’ strategic differences through behavioral clustering; “resource-driven strategists” exhibit stronger multimodal capabilities and greater learning autonomy. From a methodological standpoint, multimodal learning analysis (MLA)—which integrates eye-tracking data, physiological sensor data, and interaction logs—is reshaping the precision and multidimensional nature of language proficiency assessment. Eye-tracking studies further reveal significant individual differences in learners’ “nonverbal visual sensitivity”, which directly influences their performance in multimodal listening tests (Liu & Elola, 2025).

5. Summary

In this study, we conducted bibliometric and knowledge graph analysis of literature from Web of Science to systematically map the knowledge structure and academic trajectory of international multimodal language assessment research over the past five years. The findings indicate that, in terms of the research landscape of this field, a stable core academic community has yet to emerge; collaborative networks among authors remain relatively loose, and interdisciplinary dialogue is insufficient. Regarding thematic structure, research can be categorized into three major themes: “Research on Multimodal Input, Cognitive Processing, and Language Acquisition,” “Research on Multimodal Output, Interactive Competence, and Identity Construction,” and “Research on Multimodal Assessment and Clinical Language.” In terms of evolving research hotspots, the field has progressed through distinct phases: the foundational exploration and theoretical expansion phase (2022-2023), the framework construction and empirical application phase (2023-2024), and the technology-driven and in-depth analysis phase (2024-2026)—a trajectory shaped by the combined effects of theoretical deepening and technological advancement.

This study holds a distinct position among reviews. Existing reviews on multimodality and language education focus on pedagogical dimensions, while systematic reviews of multimodal language assessment remain scarce. This study systematically examines the research landscape, thematic structure, and hotspot evolution in multimodal language assessment, partially addressing this gap. Methodologically, it integrates CiteSpace cluster analysis with burst detection, balancing static knowledge structures with dynamic trend tracking. Temporally, it spans 2022-2026, covering both initial and rapid-expansion phases, thus offering a comprehensive account of the field’s evolution.

However, this study has limitations. In sample coverage, including only Web of Science journal articles—excluding conference proceedings and monographs—may hinder comprehensive presentation of exploratory achievements and systematic findings outside journal publications. Methodologically, CiteSpace captures

macro-level development patterns and evolutionary trajectories, but cannot replace close reading and qualitative analysis, thus failing to provide nuanced interpretation of theoretical connotations underlying key research questions.

Given these findings and limitations, future research can advance in two directions. First, expand data sources and comparative perspectives by incorporating databases such as Scopus, ERIC, and Chinese core journals, and conducting cross-linguistic and cross-regional comparisons to test the robustness of this study's conclusions. Second, achieve complementarity between quantitative and qualitative analysis. Building on CiteSpace's macro-level insights, future research could conduct close readings of high-burst, high-centrality publications to examine the operationalization of multimodal constructs, cognitive mechanisms in task design, and validity of scoring criteria; alternatively, qualitative methods—interviews, think-aloud protocols, and discourse analysis—could explore test developers', scorers', and test-takers' perspectives and practices, linking macro patterns with micro mechanisms.

Bibliography

- Carilla Pérez, M. C., & Ruiz Cecilia, R. (2025). Creativity and social networks in English language learning: A systematic literature review. *Onomázein*, (NE XVI), 209–240. <https://doi.org/10.7764/onomazein.ne16.10>
- Chen, Y., Jong, M. S.-Y., Jiang, M. Y.-C., & Li, M. (2026). Learning analytics on multimodal GAI-driven EFL oral learning: uncovering learning behavior clusters with motivation and performance dynamics. *Computer Assisted Language Learning*, 1–37. <https://doi.org/10.1080/09588221.2026.2664202>
- Crawford Camiciottoli, B., & Campoy-Cubillo, M. C. (Eds.). (2018). Introduction: The nexus of multimodality, multimodal literacy, and English language teaching in research and practice in higher education settings. *System*, 77, 1–9. <https://doi.org/10.1016/j.system.2018.03.005>
- Dai, David Wei. (2025). Assessment of Interactional Competence. In *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle. Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal20786>
- Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: A critical review. *Aphasiology*, 34(4), 492–514. <https://doi.org/10.1080/02687038.2019.1702848>
- Early, M., Kendrick, M., & Potts, D. (2015). Multimodality: Out from the margins of English language teaching. *TESOL Quarterly*, 49(3), 447–460. <https://doi.org/10.1002/TESQ.246>
- Grapin, S. E. (2023). Assessment of English Learners and Their Peers in the Content Areas: Expanding What “Counts” as Evidence of Content Learning. *Language Assessment Quarterly*, 20(2), 215–234. <https://doi.org/10.1080/15434303.2022.2147072>
- Hafner, C. A., & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing*, 47, Article 100710. <https://doi.org/10.1016/j.jslw.2020.100710>
- Han, H. C., McNaughton, S., & Zhang, L. J. (2026). A systematic review of teacher engagement in digital multimodal composing: Bridging the “last-mile” towards learner outcomes. *Teaching and Teacher Education*, 181, Article 105722. <https://doi.org/10.1016/j.tate.2026.105722>
- Hidayah Nursipah, et al. (2026). Bibliometric Analysis of Global Research Trends on Language Proficiency Tests (2018–2025). *Indonesian Journal of Instruction* 7, 48–59. <https://doi.org/10.23887/iji.v7i1.108833>
- Jiang, L. (George), & Hafner, C. (2024). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching*, 58(4), 528–546. <https://doi.org/10.1017/s0261444824000107>
- Kress, G. (2009). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge. <https://doi.org/10.4324/9780203970034>
- Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, 69, Article 101048. <https://doi.org/10.1016/j.linged.2022.101048>
- Liu, Y. T., & Elola, I. (2025). Seeing beyond the words: Modality preferences, nonverbal sensitivity, and the role of visual dynamic richness in L2 listening. *System*, 134, Article 103834. <https://doi.org/10.1016/j.system.2025.103834>
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Ouyang, Fan, and Xianping Bai. (2026). A Systematic Review of Multimodal Learning Analytics in Computer-Supported Collaborative Learning. *Computers & Education* 247: 105574. <https://doi.org/10.1016/j.compedu.2026.105574>
- Palmour, L. (2024). Assessing speaking through multimodal oral presentations: The case of construct underrepresentation in EAP contexts. *Language Testing*, 41(1), 9–34. <https://doi.org/10.1177/02655322231183077>
- Peters, E., & Muñoz, C. (2020). Introduction to the special issue: Language learning from multimodal input. *Studies in Second Language Acquisition*, 42(3), 489–497. <https://doi.org/10.1017/S0272263120000212>
- Pu, Pu, Daniel Yu-Sheng Chang, and Shuang Wang. (2024). Incidental Learning of Collocations through Different Multimodal Input: The Role of Learners' Initial L2 Proficiency. *System* 125: 103416. <https://doi.org/10.1016/j.system.2024.103416>
- Sweller, J. (2005). The Redundancy Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 159–167). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.011>
- Tan, L., Zammit, K., D'warte, J., & Gearsides, A. (2020). Assessing multimodal literacies in practice: a critical review of its implementations in educational settings. *Language and Education*, 34(2), 97–114. <https://doi.org/10.1080/09500782.2019.1708926>
- Webb, S., & Chang, A. C. S. (2022). How does mode of input affect the incidental learning of collocations?

- Studies in Second Language Acquisition*, 44(1), 35–56.
<https://doi.org/10.1017/s0272263120000297>
23. Whithaus, C. (Ed.). (2014). Multimodal assessment [Special issue]. *Computers and Composition*, 31, 1–86.
[https://doi.org/10.1016/S8755-4615\(14\)00005-X](https://doi.org/10.1016/S8755-4615(14)00005-X)
24. Yi, Y., Shin, DS., Cimasko, T. (2020). Special Issue: Multimodal composing in multilingual learning and teaching contexts. *Journal of Second Language Writing*, 47(2), 100717.
<https://doi.org/10.1016/j.jslw.2020.100717>
25. Yu, Guoxing and Tony Clark. (2025). Multimodality in Second Language Assessment. *Language Assessment Quarterly* 22, 383-399.
<https://doi.org/10.1080/15434303.2025.2608744>
26. 张德禄, 黄立鹤 (主编). (2018). 《多模态与外语教育研究》. 上海: 同济大学出版社. ISBN: 978-7-5608-8142-3