# MULTIMODAL DISCOURSE ANALYSIS IN TEXTBOOKS: A SCOPING REVIEW AND SEMI-BIBLIOMETRIC SYNTHESIS (1995-2025)

Tran Hong Lam[1,*], Lam Quang Dong[2]

*[1]Air Defense - Air Force Academy, Doai Phuong, Hanoi, Vietnam*
*[2]VNU University of Languages and International Studies, No.2 Pham Van Dong, Cau Giay, Hanoi, Vietnam*

**Abstract:** This article presents a scoping and semi-bibliometric review of 131 peer-reviewed studies on multimodal discourse analysis (MDA) in textbooks published between 1995 and 2025. The review identifies major research trends, theoretical orientations, methodological approaches, and research gaps by combining quantitative mapping and qualitative thematic coding. The findings reveal a significant expansion in research output since the early 2010s, with over 40% of the studies published between 2020 and 2024. Asia emerges as the dominant region, led by Indonesia and China, yet geographic imbalances persist, with limited contributions from Sub-Saharan Africa and Latin America. The review also shows a pronounced disciplinary concentration on English language teaching, accounting for nearly 60% of the corpus, while STEM and arts-related subjects remain underrepresented. Most of the studies adopt qualitative designs, relying heavily on content analysis and visual grammar frameworks, with 88.5% treating textbooks as isolated artefacts. Only 8.4% of the studies engage with textbook users such as teachers or students, signalling a lack of pedagogical contextualisation. Based on these findings, the article highlights five critical axes of discussion: regional asymmetry, disciplinary imbalance, methodological rigidity, theoretical convergence, and user exclusion. By synthesizing these insights, this review provides a comprehensive map of the field and directs future research toward more participatory, interdisciplinary, and digitally responsive studies that account for multimodality in situated educational practice.

*Keywords:* multimodality, textbook discourse, scoping review, bibliometric analysis, visual semiotics

---

* Corresponding author.
 Email address: lamth.1985@gmail.com

# PHÂN TÍCH DIỄN NGÔN ĐA PHƯƠNG THỨC TRONG SÁCH GIÁO KHOA: MỘT NGHIÊN CỨU TỔNG QUAN PHẠM VI VÀ TỔNG HỢP BÁN THƯ TRẮC (1995-2025)

Trần Hồng Lam[1], Lâm Quang Đông[2]

*[1]Học viện Phòng không - Không quân, Đoài Phương, Hà Nội, Việt Nam*
*[2]Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội, số 2 Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam*

**Tóm tắt:** Bài báo này thực hiện một nghiên cứu kết hợp tổng quan phạm vi và tổng hợp bán thư trắc đối với 131 công trình được bình duyệt về phân tích diễn ngôn đa phương thức (MDA) trong sách giáo khoa xuất bản trong giai đoạn 1995-2025. Tổng quan xác định các xu hướng nghiên cứu chủ yếu, các định hướng lý thuyết, các cách tiếp cận phương pháp luận và các khoảng trống nghiên cứu trong lĩnh vực thông qua việc kết hợp lập bản đồ định lượng với mã hóa chủ đề định tính. Kết quả nghiên cứu cho thấy sự gia tăng đáng kể về sản lượng nghiên cứu từ đầu thập niên 2010, với hơn 40% số nghiên cứu được công bố trong giai đoạn 2020-2024. Châu Á xuất hiện với vai trò vùng lãnh thổ chiếm ưu thế, với Indonesia và Trung Quốc dẫn đầu về số lượng, song sự mất cân bằng địa lý vẫn tiếp tục tồn tại với sự đóng góp có hạn từ khu vực châu Phi cận Sahara và Mỹ Latin. Tổng quan cũng chỉ ra sự tập trung rõ nét về mặt khoa học đối với lĩnh vực giảng dạy tiếng Anh, chiếm gần 60% tổng thể ngữ liệu, trong khi các môn học STEM và nghệ thuật vẫn chưa được đại diện tương xứng. Đa số các nghiên cứu sử dụng thiết kế định tính, chủ yếu dựa vào phân tích nội dung và các khung lý thuyết ngữ pháp thị giác, với 88,5% xem xét sách giáo khoa như những ấn phẩm biệt lập. Chỉ có 8,4% các nghiên cứu có sự tham gia của người sử dụng sách giáo khoa như giáo viên hay học sinh, điều này cho thấy sự thiếu hụt trong việc bối cảnh hóa sư phạm. Dựa trên các phát hiện này, nghiên cứu làm nổi bật năm trục thảo luận then chốt: bất cân xứng khu vực, mất cân bằng khoa học, cứng nhắc về phương pháp luận, sự đồng quy lý thuyết, và sự loại trừ đối tượng tiếp nhận. Trên cơ sở tổng hợp các kết quả này, bài báo không chỉ phác hoạ một cách toàn diện mà còn định hướng cho các nghiên cứu trong tương lai phát triển theo hướng tăng cường tính tham gia, liên ngành và khả năng thích ứng số, nhằm nắm bắt bản chất đa phương thức trong những bối cảnh giáo dục cụ thể.

*Từ khóa:* đa phương thức, diễn ngôn sách giáo khoa, tổng quan phạm vi, phân tích trắc lượng thư mục/phân tích thư trắc, ký hiệu học thị giác

## 1. Introduction

Over the past three decades, the growing use of multimodal resources in educational contexts has demanded analytical frameworks capable of interpreting meaning-making beyond the verbal mode. Multimodal Discourse Analysis (MDA), grounded in Systemic Functional Linguistics (SFL), has emerged as a key approach for examining the interplay of language, image, layout, and design in educational texts, particularly textbooks (Kress & van Leeuwen, 2021; Jewitt, 2014). Since O'Toole's (1994) pioneering work applying SFL to visual art, MDA has been increasingly adopted in applied linguistics, literacy studies, and textbook research.

As intentionally designed multimodal artefacts, textbooks play a pivotal role in mediating curricular content and ideological messaging. Textbook discourse has become a fertile ground for multimodal inquiry with the rise of image-rich layouts, digital platforms, and

multiliteracies pedagogy (The New London Group, 1996). Despite this, systematic syntheses of MDA applied to textbooks remain limited, with most reviews focusing on broader multimodal communication (Bateman, 2014) or specific pedagogical practices (Lv et al., 2024).

This review addresses this gap by analysing 131 peer-reviewed studies published between 1995 and 2025, employing a semi-bibliometric approach that combines descriptive statistics with qualitative thematic analysis (Zupic & Čater, 2014). To establish a clear scope, a textbook is defined as a formally published, curriculum-aligned core resource used in primary, secondary, or tertiary education, including both print and digital formats, as well as integrated supplementary materials such as workbooks. The selected timeframe captures a 30-year development arc, beginning with foundational theoretical shifts and extending into the contemporary landscape shaped by digital learning environments.

In doing so, this study aims to systematically chart and synthesise thirty years of research to identify prevailing trends, conceptual gaps, and underexplored areas in applying MDA to textbooks, while also mapping its evolution across geographic regions, disciplines, and methodologies, and identifying key blind spots and future research trajectories.

## 2. Literature Review

### 2.1. Theoretical Foundations

MDA has emerged as a significant theoretical and analytical paradigm in response to the increasing complexity of meaning-making across semiotic modes. Grounded in Halliday's (1985) SFL, MDA extends the metafunctional framework, ideational, interpersonal, and textual, beyond verbal language to include visual and spatial modalities (O'Halloran, 2004). A foundational contribution to this shift was O'Toole's (1994) *The Language of Displayed Art*, which applied SFL to visual art and marked a theoretical departure toward systematic multimodal inquiry.

This trajectory was further advanced by Kress and van Leeuwen (2021) in *Reading images: The grammar of visual design*, where they articulated visual grammar to explain how compositional features such as salience, framing, and modality contribute to meaning, creating a framework that has become central to textbook discourse studies, particularly in analysing image–text relations. Simultaneously, the concept of multiliteracies introduced by The New London Group (1996) expanded the definition of literacy to encompass multimodal communication, reinforcing the pedagogical relevance of MDA.

In textbook research, MDA has been widely employed to examine how visual-verbal configurations structure disciplinary knowledge, guide learner interpretation, and embed ideological content (Bezemer & Kress, 2008; Jewitt, 2014). For example, Martinec and Salway (2005) explored how image–text relations can facilitate or obstruct comprehension. In ELT contexts, studies have identified mismatches between visuals and accompanying language, while in STEM fields, the interplay of symbolic, visual, and textual complexity has been shown to impact learners' understanding (Neri & Retelsdorf, 2022). Furthermore, Cross-cultural analyses (e.g., Torres, 2015) have also demonstrated how textbook visuals reflect implicit ideologies concerning race, gender, and national identity.

### 2.2. Previous Reviews

Despite the growing volume of research on multimodality in educational contexts, literature reviews that synthesise MDA-based textbook studies remain limited. Vojíř and Rusek (2019), for instance, systematically reviewed science textbooks, focusing on methodological

tendencies and regional distribution without engaging deeply with multimodal theory. Similarly, Neri and Retelsdorf (2022) reviewed linguistic features in science and math comprehension, emphasising language complexity but not multimodal integration.

In broader pedagogical research, Lv, Ni, and Tan (2024) synthesised studies on multimodal classroom teaching and found consistent benefits for learner engagement and literacy acquisition. However, their review did not address how textbooks, as central multimodal artefacts, function within these instructional practices. Overall, most existing reviews either target specific disciplines or focus on narrow pedagogical strategies, thus failing to offer a cross-disciplinary synthesis of how MDA has been applied to textbooks over time and across contexts. These reviews are often fragmented and discipline-bound, offering no panoramic synthesis of how MDA has evolved across textbooks, subjects, and cultural contexts.

Given the increasing shift toward digitally mediated and visually complex educational resources, a comprehensive review of MDA in textbook research is timely and necessary. Critically, no existing study has offered a longitudinal synthesis that spans multiple subjects, educational levels, and geographic contexts. The present review directly addresses this gap by systematically analysing 131 peer-reviewed studies published between 1995 and 2025 to map MDA's conceptual and methodological evolution in textbook analysis. It further aims to highlight areas for future inquiry, particularly those related to layout design, digital affordances, and user-centred evaluation, thereby contributing to a more integrated understanding of multimodal literacy in education.

## 3. Methodology

### 3.1. Review Design

This study adopts a scoping review framework, enriched by a semi-bibliometric synthesis, to investigate research trends, theoretical orientations, and methodological approaches in applying Multimodal Discourse Analysis (MDA) to textbooks published between 1995 and 2025. The review follows the PRISMA 2020 guidelines (Page et al., 2021) to ensure transparency and replicability, with adaptations made for educational discourse research, where conceptual depth and interpretive rigour are often prioritised over citation metrics.

Rather than assessing the quality of individual studies, the primary aim of this review is to map the landscape of textbook-based MDA by identifying conceptual gaps, methodological developments, and thematic concentrations. Scoping reviews are especially suited for such broad and heterogeneous domains, where empirical designs vary widely, and formal appraisal tools may not apply uniformly (Arksey & O'Malley, 2005; Tricco et al., 2018).

To achieve this, a hybrid design was employed, which combines quantitative bibliometric mapping, including publication year trends, geographic distribution, keyword frequency, and co-occurrence network analysis, with qualitative thematic synthesis of theoretical frameworks, research questions, analytical strategies, and modal emphases (Zupic & Čater, 2014; Ramos-Rodríguez & Ruíz-Navarro, 2004).

The selected 30-year time frame (1995–2025) is grounded in the field's theoretical evolution. Specifically, it begins shortly after the publication of O'Toole's (1994) *The Language of Displayed Art*, widely recognised as a foundational application of SFL to visual and multimodal meaning-making. By extending to 2025, the review captures the most recent developments, including the digital transformation of textbooks, the rise of multiliteracies pedagogy, and increased scholarly interest in multimodal education.

In sum, this review provides a mapping of current conceptualisations of MDA in textbook research and, by adopting a scoping review approach, focuses on identifying areas for future empirical inquiry rather than evaluating specific pedagogical interventions.

### 3.2. Data Collection and Selection Criteria

Following this design, a total of 131 peer-reviewed studies were included in this review through a transparent, replicable process aligned with PRISMA 2020 guidelines (Page et al., 2021). The dataset in Appendix A was compiled through targeted manual searches across multiple platforms, including Google Scholar, ERIC, ResearchGate, Semantic Scholar, Academia.edu, and selected institutional repositories. To ensure comprehensive coverage, backwards and forward citation tracing from foundational works, such as Kress and van Leeuwen (2021) and Jewitt (2014), was also employed to enhance coverage.

The search strategy used Boolean combinations, e.g., "textbook" AND ("multimodal discourse analysis" OR "visual grammar" OR "systemic functional linguistics" OR "semiotics" OR "multiliteracies"). The inclusion window spanned January 1995 to March 2025, covering three decades of theoretical and empirical development. Only studies written in English or containing a full English abstract were eligible. Regarding document types, peer-reviewed journal articles, book chapters, conference proceedings, and scholarly monographs were all considered.

Studies were included if they: (1) applied MDA or related frameworks either empirically or theoretically, (2) focused on textbooks or derived materials (e.g., coursebooks, workbooks), and (3) were available in full text. Conversely, exclusion criteria ruled out studies focusing solely on classroom interaction, those addressing textbook policy without a multimodal lens, editorials, and dissertations without original analysis.

**Figure 1**

*PRISMA Flow Diagram for Study Selection*
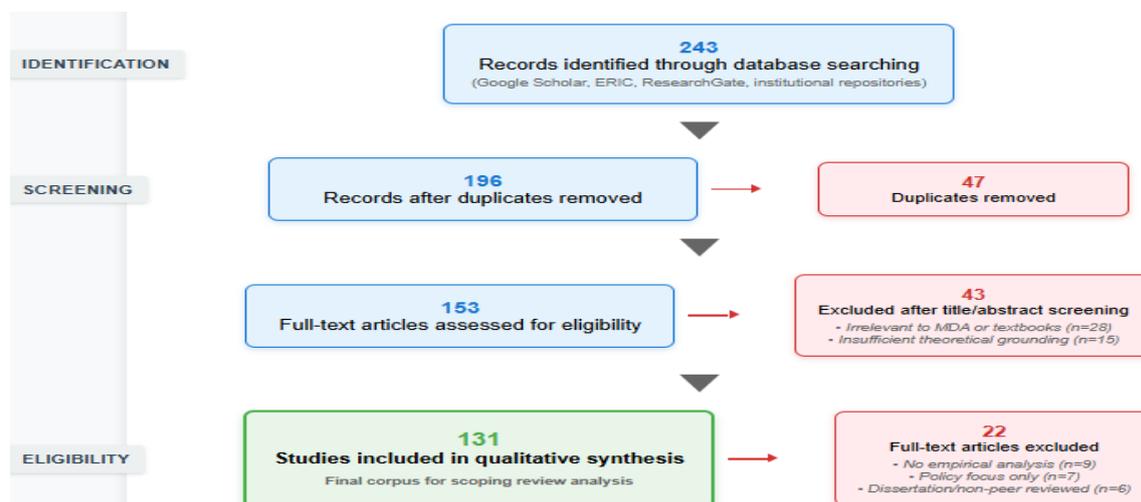


As shown in Figure 1, the selection process began with 243 initial records, and 47 duplicates were removed. After title and abstract screening (n = 196), 43 were excluded for irrelevance or insufficient grounding. Subsequently, of the 153 full-text articles assessed, 22 were excluded, resulting in 131 studies in the final synthesis.

### 3.3. Coding and Data Management

Following their selection, all 131 studies included in the final synthesis were read in full

and systematically coded using a structured spreadsheet comprising 22 metadata variables. These variables were organised into four categories: (1) bibliographic information (e.g., author, year, journal, country, indexation status); (2) research features (e.g., methodological design, sample size, education level, subject domain); (3) multimodal features (e.g., image, layout, typography, theoretical framework); and (4) user engagement (i.e., whether the study included data from teachers, learners, or classroom observations).

The coding scheme was grounded in established frameworks in multimodal discourse research. In particular, SFL (Halliday, 1994), particularly the metafunctional model, was the principal basis for analysing ideational and interpersonal meanings across modes. Simultaneously, Kress and van Leeuwen's (2021) visual grammar guided the identification of compositional features such as salience, gaze, modality, and framing, which are key for analysing text–image relationships and layout strategies.

Additionally, several studies employed hybrid theoretical approaches, combining SFL and social semiotics (Jewitt, 2014), intersemiotic complementarity (Royce, 2007), or critical multimodal discourse analysis (Machin, 2013) to explore ideological positioning, identity, and evaluative stance in textbooks. These frameworks informed sub-coding dimensions related to representation, learner roles, and cultural framing.

To further map the conceptual structure of the field, a corpus of 642 unique keywords was extracted from article metadata and full texts. Co-occurrence networks were generated using VOSviewer 1.6.20 (threshold: $\geq 5$). Using the LinLog/modularity algorithm for clustering, this process allowed for the identification of dominant themes, emergent concepts, and evolving methodological orientations across the 30-year review period.

### 3.4. Analytical Procedure

The analysis followed a two-stage approach. In the first, quantitative stage, methods were used to chart publication trends across time, regions, disciplines, and education levels. To visualize these patterns, bar charts, line graphs, and word frequency clouds were created to reveal longitudinal and geographic patterns. In the subsequent qualitative stage, thematic synthesis was conducted to group studies according to shared conceptual foundations, modal emphases, and research intentions.

This synthesis enabled the classification of studies into analytical typologies, such as design-focused analyses that prioritise layout, colour, or text–image integration; ideology-oriented studies exploring themes of representation, power, or identity; and pedagogically oriented research investigating how multimodal resources affect learning outcomes and instructional design. Overall, this dual approach allows for both a macro-level overview and a micro-level interpretive account of how multimodality has been conceptualised and applied in textbook research.

## 4. Results

The present section outlines the findings of a systematic and semi-bibliometric review of 131 studies on MDA in textbooks published between 1995 and 2025. Drawing from the coded dataset and statistical synthesis, the results are organised into eight analytic dimensions: (1) year of publication, (2) national and regional distribution, (3) educational level, (4) subject area, (5) type of materials analysed, (6) inclusion of textbook users, (7) publication types and scholarly visibility, and (8) keywords of the studies.

This multi-layered analysis not only enables us to trace the historical evolution of the field but also identifies disciplinary and geographic imbalances, as well as reveals emerging

trends and persistent gaps. Each dimension is presented with supporting tables, figures, and brief interpretive commentary. While descriptive statistics are prioritised in this section, more profound theoretical reflections will follow in the discussion.

### 4.1. Temporal and Geographical Patterns

The evolution of MDA in textbook research over the past three decades shows both an apparent temporal acceleration and a pronounced geographical concentration. Among the 131 studies reviewed, the earliest was published in 1997. From 1995 to 2004, research activity remained minimal, with only 10 studies published over the entire decade. A period of modest growth continued between 2005 and 2012, with a few intermittent spikes, such as five studies in 2010.

**Figure 2**

*Annual Number of Studies on MDA in Textbooks (1995–2025)*
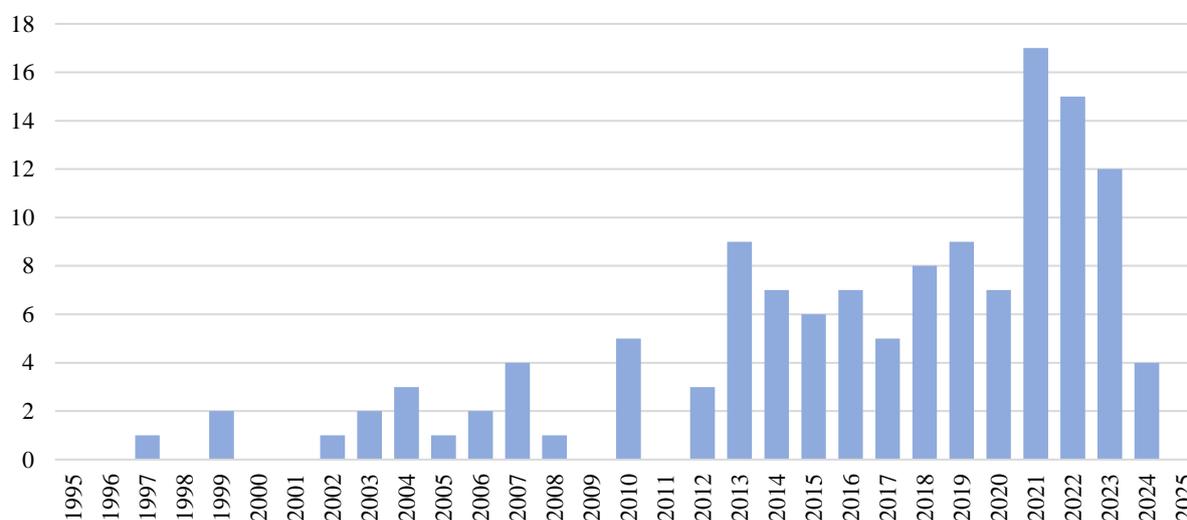


A clear turning point occurred in 2013, when the number of publications tripled compared to the previous year. This surge marked the beginning of a steady rise in scholarly output. The most prolific period was 2020–2024, accounting for 54 studies, or over 41% of the total corpus. For example, 2021 alone saw 17 publications, followed by 15 in 2022 and 12 in 2023. Although the counts for 2024 (4) and 2025 (0 as of August 2025) are lower, these figures likely reflect data-collection cutoffs rather than a reversal of the trend.

Geographically, the research is highly unevenly distributed. Although studies span 37 countries across six continents, more than 60% originate from Asia. Indonesia leads with 21 studies, followed by China (17), Iran (10), Vietnam (6), South Korea (5), Singapore (4), Iraq (3), and Japan (3). Countries with fewer than three studies are aggregated as "Other". Remarkably, these eight countries alone account for more than half of all reviewed studies. In contrast, European contributions are led by Greece (7 studies) and the United Kingdom (5), with limited representation from Western and Southern Europe. Australia accounts for four studies in the Australasian region, while Latin America and Sub-Saharan Africa remain markedly underrepresented. Additionally, only six studies in the dataset employed cross-national comparative analysis, highlighting a missed opportunity to explore how multimodal practices differ across cultures and educational systems. To illustrate, Indonesia's dominance is exemplified by studies such as ID#31, ID#68, and ID#122, which focus on textbook gender representation and EFL materials.

**Figure 3**

*Top Contributing Countries in MDA Textbook Research (1995-2025)*
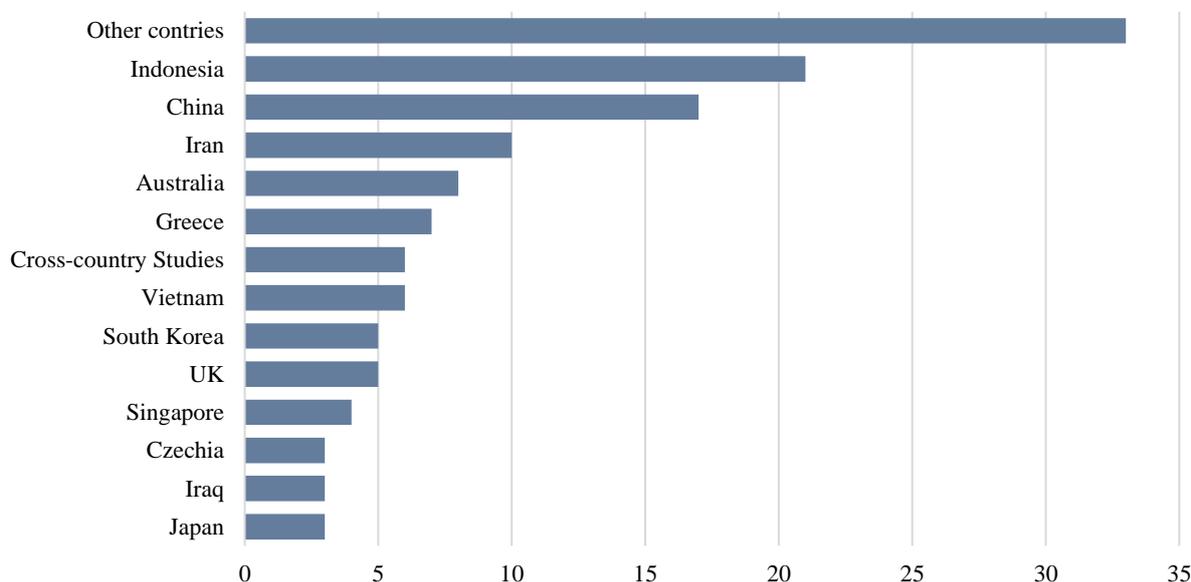


In summary, the past three decades have steadily accelerated scholarly interest in multimodal textbook analysis, particularly in East and Southeast Asia. These combined spatio-temporal trends reflect shifting research priorities and, at the same time, reveal systemic disparities in global academic participation, which continues to shape the visibility and scope of MDA research.

### 4.2. Educational Levels and Subject Focus

Building on the temporal and geographical patterns, this section examines how MDA has been applied across different educational levels and subject domains. Among the 131 studies reviewed, the majority, 66 studies (50.4%), focused on secondary education, particularly middle and high school textbooks. Primary education was addressed in 23 studies (17.6%), whereas tertiary-level materials, such as university coursebooks, were the subject of 19 studies (14.5%). A smaller subset of research (10 studies, 7.6%) spanned multiple educational levels, and 13 studies (9.9%) did not specify the level or focused on early childhood materials.

The predominance of secondary-level research aligns with the widespread use of textbooks in formal curriculum delivery and the density of disciplinary content at this stage. On the other hand, primary and tertiary education remain less explored, a notable gap given the prominence of visual storytelling in early childhood materials and the increasing use of ESP (English for Specific Purposes) coursebooks in higher education contexts.

Regarding subject focus, English-related disciplines clearly dominate the field. A total of 75 studies (57.3%) concentrated on English as a Foreign Language (EFL), English Language Teaching (ELT), or ESP. Science education followed distantly with 21 studies, while biology and geography appeared in 6 studies each. Other subjects, including mathematics, history, civics, and religious education, were sporadically represented, typically as single-case analyses or components of broader curriculum studies. Studies such as ID#59 and ID#113 typify the ELT focus, whereas only a handful, like ID#19, address science textbooks.

**Figure 4**

*Distribution of Studies by Educational Level*
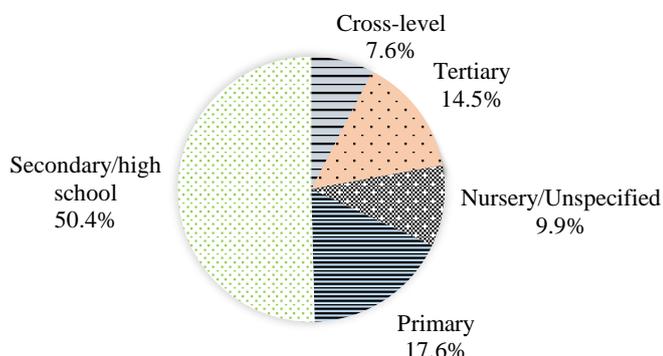


This disciplinary skew largely reflects the origins of MDA within applied linguistics and visual semiotics, where analytical tools such as SFL and visual grammar are particularly suited for deconstructing verbal-visual relations in language textbooks. Despite the high visual complexity of STEM and humanities subjects, few studies applied multimodal frameworks to these domains, thereby revealing an uneven application of MDA across curricular fields. Consequently, the scope of multimodal textbook research thus appears to be shaped more by disciplinary methodology than by pedagogical need.

**Figure 5**

*Distribution of Studies by Subject Area*
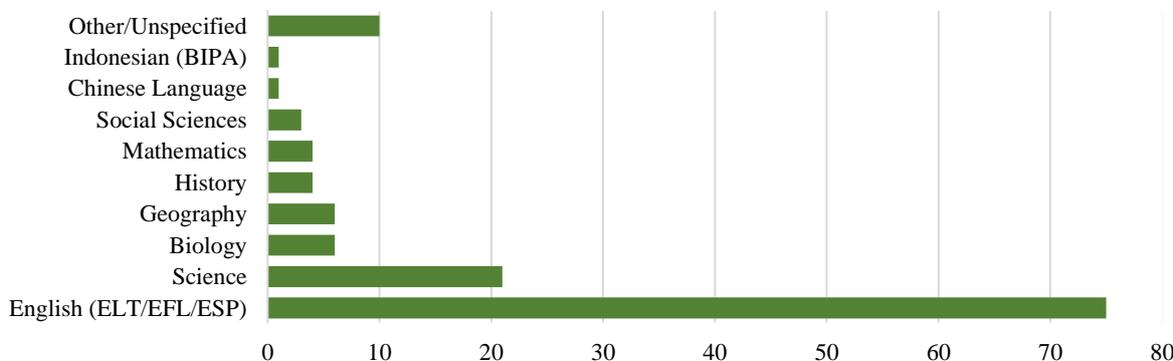


### 4.3. Publication Types, Scholarly Visibility, and User Engagement

The publication landscape of multimodal textbook research reflects both the disciplinary consolidation of the field and its uneven integration into global scholarly communication. Among the 131 studies reviewed, 104 (79.4%) were published as peer-reviewed journal articles, demonstrating the field's growing academic legitimacy. The remainder consisted of 14 book chapters (10.7%), 10 conference proceedings (7.6%), and three monographs (2.3%). This dominance of journal articles marks a shift from early theoretical explorations in monographs toward empirically grounded research disseminated through discipline-specific venues in applied linguistics, literacy, and education.

Within the 104 journal articles, 49 studies (47.1%) appeared in Scopus-indexed journals, of which 28 were published in Q1-ranked outlets. These Q1 outlets include titles such as *Multimodal Communication*, *Visual Communication*, and *Systemic Functional Linguistics*,

reflecting the theoretical maturation of the field, especially within visual semiotics and SFL-based frameworks. However, 55 articles were published in non-indexed journals, many of them regional or language-specific, especially in Asia and Latin America. While such outlets often provide culturally contextualised insights, their limited indexing may restrict international reach and citation impact.

**Figure 6**

*Visibility of Journal Articles by Indexing and Ranking*
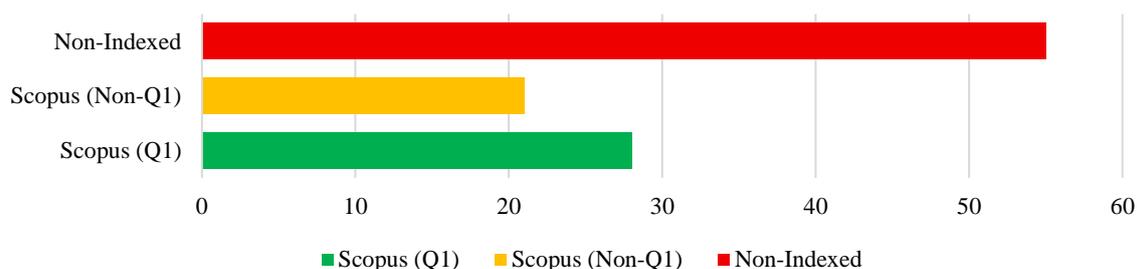


Despite its semiotic orientation, multimodal textbook research has largely treated textbooks as static, decontextualised objects. A clear pattern emerges: 116 studies (88.5%) examined textbooks alone, without extending analysis to accompanying materials or classroom use. Only four studies (3.1%) analysed supplementary items such as teacher guides or policy documents, and just 11 studies (8.4%) incorporated real user data through interviews, surveys, or classroom observations. For instance, only a few studies, like ID#10, directly incorporated user perspectives through classroom observation or learner interviews. Moreover, these user-inclusive studies remain methodologically narrow, with few employing multimodal interaction analysis, eye-tracking, or ethnographic observation.

Even within textbook-centred research, the material scope varied. For example, some studies analysed entire volumes, while others selected units, visuals, or exercises. A related issue is that digital multimodality remains underrepresented; only eight studies explicitly addressed interactivity features such as hyperlinks, audio, or animations. As textbooks increasingly shift toward hybrid formats, this limited engagement with digital modes reflects a methodological lag in capturing emerging media practices.

Taken together, these findings suggest an academically established research tradition that is still evolving in responsiveness to user perspectives and digital pedagogical realities. Greater openness to ethnographic and participatory methods could, therefore, help reposition textbooks not merely as objects of analysis but as multimodal interfaces embedded in teaching and learning.

### 4.4. Analytical Frameworks, Methods, and Keyword Trends

The methodological landscape of multimodal textbook research reveals a strong preference for qualitative and interpretive approaches, reflecting the field's theoretical foundations in discourse analysis and social semiotics. Among the 131 studies reviewed, 96 (73.3%) employed qualitative designs, followed by 25 mixed-methods studies (19.1%) and 10 quantitative studies (7.6%). Although qualitative work dominates, recent years have seen a gradual increase in mixed approaches, especially those incorporating frequency-based content coding, layout mapping, or software-aided visualisation.

Analytically, multiple coding was allowed; percentages are calculated over 131 studies and may exceed 100%. For readability, Figure 7 displays only methods that appear in ≥7 studies

(methods below this threshold are not shown). Within this scope, content analysis was the most frequently used method (64 studies, 48.9%), encompassing both qualitative and quantitative variants. Following this, MDA (49 studies, 37.4%) and social semiotics (36 studies, 27.5%) were also prevalent, often employed in conjunction with SFL (e.g., ID#10, ID#16 adopt the Hallidayan framework), critical discourse analysis, or intersemiotic models. In total, 15 distinct methods were identified across the corpus, signalling methodological diversity. Tools such as NVivo, UAM CorpusTool, and transitivity analysis further supported more systematic multimodal coding.

**Figure 7**

*Most Frequently Used Analytical Methods in MDA Textbook Research*
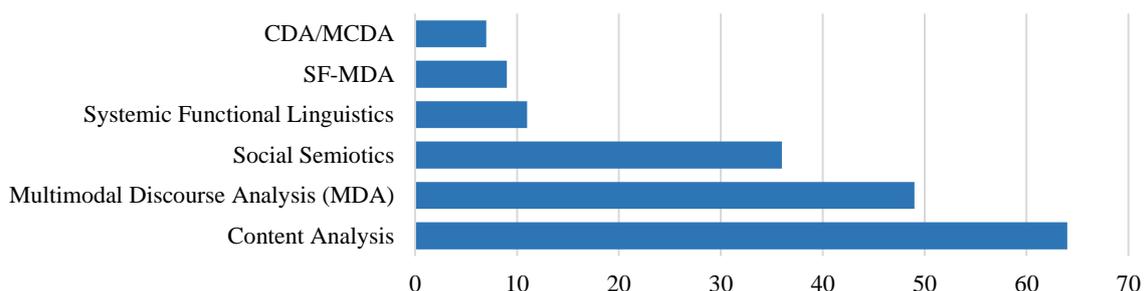


Triangulation was standard: 65.6% of the studies used two or more analytical techniques. Pairings such as Content Analysis + MDA (n = 18) or MDA + Social Semiotics (n = 15) were particularly frequent, which enabled researchers to combine structural analysis with interpretive depth. For example, ID#7 and ID#8 exemplify the use of social semiotics and visual grammar in traditional MDA, while ID#6 uses spatial layout analysis in geography textbooks. Despite this, only a few studies employed interaction-sensitive techniques (e.g., eye-tracking, classroom discourse), again indicating a continued focus on the textbook as a static object of analysis.

The scope of methodological units also varied. Specifically, while 76 studies analysed single cases or selected textbook chapters, 32 conducted small-scale analyses (<100 items), and 23 studies reached medium or large-scale thresholds, often through corpus-driven approaches. Chronologically, methodological proportions remained relatively stable, although quantitative and mixed-methods studies showed modest growth in the 2020s.

**Figure 8**

*Distribution of Keyword Groups in Multimodal Textbook Research*
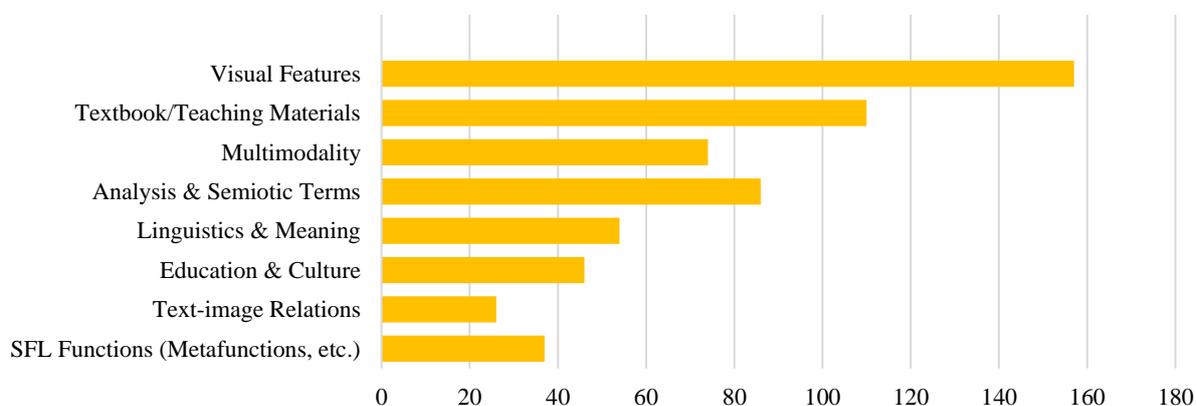
In Figure 8, only high-frequency keywords are displayed; low-frequency items are omitted for readability. Keyword analysis further clarifies the theoretical structure of the field. From 642 keywords extracted across all studies, visual-related terms (e.g., image, diagram, illustration) appeared 157 times, accounting for 24.5% of the total. This finding underscores the centrality of image–text relations and frameworks such as visual grammar and intersemiotic theory. Terms like *textbook*, *curriculum*, and *material* were also standard (17.1%), confirming the object-based focus of most research. A related issue is that while keyword frequency sheds light on the dominant conceptual and analytical emphases, it does not capture subject-specific content such as science, mathematics, or history. These disciplinary distinctions have already been systematically analysed in the dedicated section on subject distribution (Section 4.2), and are therefore not reiterated in the keyword analysis. Instead, the keyword data are interpreted as indicators of theoretical orientation, methodological preferences, and neglected analytical domains, particularly in evaluative meaning, social identity, and user interaction.

Methodological terms such as *content analysis*, *meaning*, *discourse*, and *semiotics* appeared in 13.4% of keywords, while those linked to linguistics (*SFL*, *metafunction*, *transitivity*) appeared 8.4% of the time. Pedagogical and sociocultural themes, *education*, *culture*, *literacy*, were referenced in 46 instances. By contrast, critical concepts such as *appraisal*, *gender*, *layout*, or *multiliteracies* were notably scarce, each appearing fewer than six times, thereby suggesting potential directions for future expansion.

Overall, these patterns reveal a field grounded in interpretive and linguistically-informed frameworks, with strong interest in visual resources but underexplored attention to design structure, evaluative meaning, and user interaction. Thus, methodological innovation and theoretical broadening remain important frontiers for advancing multimodal textbook research.

While the results provide a comprehensive overview of multimodal textbook research across multiple dimensions, several limitations must be acknowledged. First, the data are primarily descriptive, with limited use of inferential statistical methods to explore potential relationships among variables. Second, although efforts were made to include the most recent studies, the apparent drop in publications in 2024–2025 may reflect incomplete indexing rather than actual trends. Lastly, variations in how studies defined or sampled "textbooks" pose challenges to direct comparability, particularly in cross-disciplinary and digital contexts. Consequently, future analyses could benefit from more systematic sampling criteria, longitudinal comparisons, and inferential analytics to strengthen claims.

**Figure 9**

*Keyword Co-Occurrence Network Generated Using VOSviewer (≥ 5 Occurrences)*
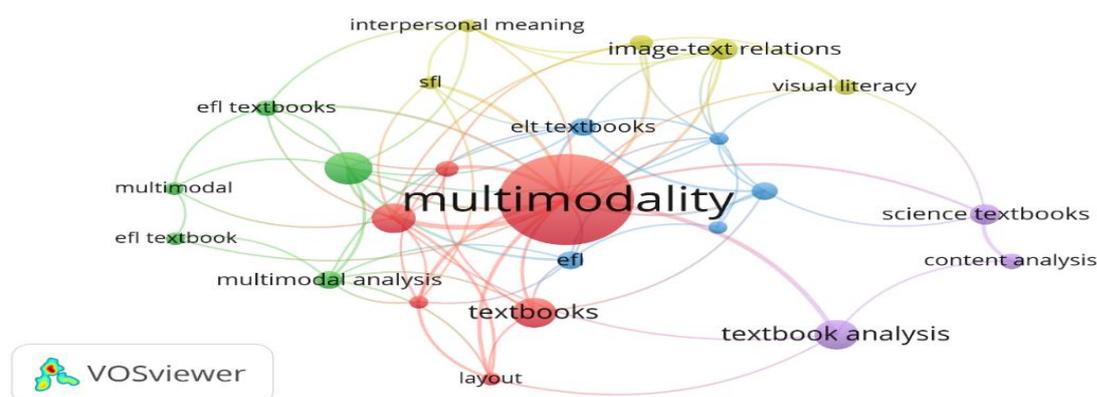
In short, these findings establish a strong empirical foundation for understanding the contours of multimodal textbook research. However, numbers alone cannot capture the field's conceptual tensions or neglected areas. The following section moves beyond description to critically interpret these patterns, foregrounding their implications for theory, pedagogy, and future research agendas.

## 5. Discussion

This section critically interprets the major patterns identified in the reviewed literature, moving beyond descriptive trends to examine MDA's epistemological, methodological, and theoretical underpinnings in textbook studies. Organised around five interrelated themes, the discussion evaluates the field's strengths, limitations, and future directions based on systematic evidence and established frameworks.

### 5.1. Global Expansion and Epistemic Inequality

The steep growth in publications, particularly from 2013 onward, reflects the increasing legitimacy of MDA within applied linguistics and literacy education (Jewitt, 2014; Kress, 2010). However, this growth has not been geographically uniform. The concentration of studies in Asia, especially Indonesia, China, and Iran, suggests regional engagement driven by national reforms in textbook policy and the rise of graduate-level research infrastructures. Yet, the underrepresentation of Sub-Saharan Africa and Latin America poses significant questions about epistemic inclusion and global knowledge equity (Canagarajah, 2002). This pattern is exemplified by high-output studies such as ID#63 (Vietnam), ID#39 (Malaysia), and ID#34 (Indonesia), while no study in the corpus originated from Sub-Saharan Africa.

The geographic concentration of research is not a neutral fact; it has profound implications for the field. First, it reveals a systemic epistemic inequality: the current knowledge base on textbook multimodality is predominantly shaped by specific cultural, educational, and policy contexts of a few high-output countries. Consequently, our understanding may be incomplete and biased towards certain pedagogical traditions and visual rhetorics, limiting the generalizability of findings. Second, the paucity of cross-national comparative studies (fewer than 5% of the corpus) hinders the development of theories that account for how multimodal communication varies across different linguistic and curricular environments.

These implications underscore the need to reorient research priorities. In line with recent calls for southernising educational research (Tikly, 2019), future work must actively address this geographic imbalance by fostering more inclusive, dialogic, and transnational approaches. This entails not only encouraging scholarship from underrepresented regions but also designing studies that explicitly compare multimodal practices across diverse contexts to decenter dominant paradigms. Thus, the geographic analysis serves as a crucial diagnostic tool, highlighting both the contours of current scholarship and the pathways toward a more globally representative field.

### 5.2. Disciplinary Boundaries and Missed Opportunities

Beyond geographical constraints, the field faces significant disciplinary limitations. The overwhelming dominance of English language education, 57.3% of all studies, indicates a disciplinary overreliance on applied linguistics and ELT frameworks (Unsworth, 2006; Royce, 2007). While this provides analytical clarity and access to well-established models (e.g., SFL, visual grammar), it unintentionally marginalises other domains such as STEM, history, civics,

and visual arts, where multimodal representations are equally salient but less studied. For instance, studies like ID#25, ID#55, and ID#86 analyse multimodality in ELT textbooks, while only a few, such as ID#71 (chemistry) or ID#20 (physics), explore STEM content.

This narrow focus has concrete consequences. Notably, mathematics and science textbooks, which heavily use symbolic and graphical resources, have received minimal attention despite their clear relevance to intersemiotic analysis (Lemke, 2002; Tang et al., 2022). This disciplinary imbalance risks narrowing the scope of multimodal research and overlooking domain-specific challenges in designing, interpreting, and teaching from multimodal texts. Therefore, a broader curricular engagement is essential to uncover how multimodality functions differently across epistemological traditions.

### 5.3. Methodological Rigidity and Innovation

Complementing these thematic and disciplinary patterns, the methodological profile of the field reveals both consistency and constraints. It is characterised by a dominance of qualitative research (73.3%), particularly content analysis and visual grammar-based MDA. This reflects the interpretive orientation of early multimodal studies (Kress & van Leeuwen, 2021; O'Halloran, 2008). For example, *ID#16* (Chen, 2022) employs interpersonal meaning analysis with multimodal annotations, while *ID#28* applies digital layout mapping in a science textbook.

However, this established paradigm may also indicate a degree of rigidity. Only 7.6% of the studies employed quantitative techniques, and few explored multimodal interaction in real-world settings. While triangulated designs (used in 65.6% of studies) indicate a growing awareness of complexity, very few studies utilised technological tools such as eye-tracking, clickstream analysis, or multimodal corpora, methods that are increasingly common in digital literacy research (Jewitt & Bezemer, 2015).

Consequently, there is a clear opportunity for methodological renewal. Future research could combine semiotic depth with behavioural insight and statistical modelling, particularly in studies examining user engagement or digital learning environments. Such innovation is crucial for capturing the dynamic and interactive nature of multimodality in contemporary educational practice.

In summary, this discussion has highlighted three interconnected axes shaping the field of MDA in textbook research: pronounced geographic and epistemic inequalities, a narrow disciplinary focus on language education, and a methodological leaning towards qualitative interpretation at the expense of innovative, user-centred, and quantitative approaches. These patterns collectively underscore a research domain that, while established and growing, requires a conscious expansion in its geographic, disciplinary, and methodological horizons to fully capture the complexities of multimodal meaning-making in global educational contexts.

## 6. Conclusion

This scoping and semi-bibliometric review has synthesized three decades of research to map the evolving landscape of Multimodal Discourse Analysis (MDA) in textbook studies. The synthesis reveals a field marked by distinct and interrelated trends. Geographically, research production is highly concentrated, with scholarship predominantly emerging from specific Asian contexts, while vast regions, including Sub-Saharan Africa and Latin America, remain on the margins of the academic conversation. Disciplinarily, the field exhibits a pronounced leaning, with the majority of analytical attention devoted to English language teaching materials, leaving the multimodal complexity of STEM and humanities textbooks largely

underexplored. Methodologically, a dominant paradigm is evident, characterized by qualitative, text-immanent analyses that often treat textbooks as static artefacts, with research rarely extending to the pedagogical contexts of their actual use or to the perspectives of teachers and learners. These converging trends outline a research domain that, despite demonstrating robust growth and consolidation, faces critical challenges regarding its global representativeness, curricular scope, and pedagogical relevance.

In interpreting these findings, several limitations of the present review must be acknowledged. Its scope is defined by a curated sample and manual analysis, which, while enabling depth of engagement, may not capture the entire universe of relevant research. Additionally, an English-language bias likely excluded relevant studies published in other languages. Variations in how individual studies defined "textbooks" and the single-coder approach to data analysis, despite rigorous cross-checking, also pose constraints on generalizability.

Notwithstanding these limitations, the analysis provides a clear, evidence-based foundation for advancing the field. Looking ahead, the identified trends point to several critical and interconnected directions for future research. Theoretically, the strong reliance on SFL and social semiotics suggests room for expansion into complementary frameworks, such as affect theory or critical race visual analysis, that could more fully capture the experiential and sociopolitical dimensions of textbooks. Methodologically and pedagogically, the predominant focus on the text itself underscores an urgent need to develop user-centred and classroom-situated research designs that investigate how multimodality functions in practice. Finally, the stark geographic concentration of scholarship calls for a deliberate reorientation towards more equitable and comparative approaches that can decentralize dominant perspectives and incorporate a wider range of cultural and educational contexts. By addressing these priorities, theoretical diversification, methodological innovation, and geographic inclusion, MDA research can evolve to offer a more robust, equitable, and practically grounded understanding of multimodality in global education.

## References

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bateman, J. A. (2014). Using multimodal corpora for empirical research. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 238–252). Routledge. https://www.routledge.com/The-Routledge-Handbook-of-Multimodal-Analysis/Jewitt/p/book/9781138245198

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195. https://doi.org/10.1177/0741088307313177

Canagarajah, A. S. (2002). *A geopolitics of academic writing*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt5hjn6c

Halliday, M. A. K. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Deakin University Press. https://archive.org/details/languagecontextt0000hall/page/n1/mode/2up

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold. https://archive.org/details/introductiontofu0000hall

Jewitt, C. (Ed.). (2014). *The Routledge handbook of multimodal analysis* (2nd ed.). Routledge. https://www.routledge.com/The-Routledge-Handbook-of-Multimodal-Analysis/Jewitt/p/book/9781138245198

Jewitt, C., & Bezemer, J. (2015). *Multimodality, learning and communication: A social semiotic frame*. Routledge. https://www.routledge.com/Multimodality-Learning-and-Communication-A-social-semiotic-frame/Bezemer-Kress/p/book/9780415709620

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge. https://www.routledge.com/Multimodality-A-Social-Semiotic-Approach-to-Contemporary-Communication/Kress/p/book/9780415320610

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design (3rd ed.)*. Routledge. https://doi.org/10.4324/9781003099857

Lemke, J. L. (2002). Mathematics in the middle: Measure, picture, gesture, sign, and word. In M. Anderson, A. Saenz-Ludlow, S. Zellweger, & V. Cifarelli (Eds.), *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing* (pp. 215–234). Legas Publishing. https://www.academia.edu/3033680/Mathematics_in_the_middle_Measure_picture_gesture_sign_and_word

Lv, Y., Ni, W., & Tan, Y. (2024). An overview of research on multimodal teaching and learning. *Journal of Education and Educational Research, 8*(3), 492–495. https://doi.org/10.54097/qp4jg551

Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies, 10*(4), 347–355. https://doi.org/10.1080/17405904.2013.813770

Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication, 4*(3), 337–371. https://csis.pace.edu/~marchese/TextImage/image-text-allmedia..pdf

Neri, N. C., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: A systematic review and desiderata for future research. *Educational Research Review, 36*, 100460. https://doi.org/10.1016/j.edurev.2022.100460

O'Halloran, K. L. (2004). *Multimodal discourse analysis: Systemic-functional perspectives*. Continuum. https://www.amazon.com/Multimodal-Discourse-Analysis-Perspectives-Linguistics/dp/0826472567

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication, 7*(4), 443–475. https://doi.org/10.1177/1470357208096210

O'Toole, M. (1994). *The language of displayed art*. Fairleigh Dickinson University Press. https://www.routledge.com/The-Language-of-Displayed-Art/OToole/p/book/9780415595278

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., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Ramos-Rodríguez, A.-R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the *Strategic Management Journal*, 1980–2000. *Strategic Management Journal, 25*(10), 981–1004. https://doi.org/10.1002/smj.397

Royce, T. D. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce & W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 63–109). Lawrence Erlbaum. https://doi.org/10.4324/9780203357774

Tang, K. S., Park, J., & Chang, J. (2022). Multimodal genre of science classroom discourse: Mutual contextualisation between genre and representation construction. *Research in Science Education, 52*, 755–772. https://doi.org/10.1007/s11165-021-09999-1

The New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Tikly, L. (2019). *Education for sustainable development in the postcolonial world: Towards a transformative agenda for Africa* (1st ed.). Routledge. https://doi.org/10.4324/9781315211343

Torres, G. (2015). 'Reading' *World Link*: A visual social semiotic analysis of an EFL textbook. *International Journal of English Language Education, 3*(1), 239–253. https://doi.org/10.5296/ijele.v3i1.7200

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique, 5*(1), 55–76. https://eric.ed.gov/?id=EJ843820

Vojíř, K., & Rusek, M. (2019). Science education textbook research trends: A systematic literature review. *International Journal of Science Education, 41*(11), 1496–1516. https://doi.org/10.1080/09500693.2019.1613584

Zupic, I., & Čater, T. (2014). Bibliometric Methods in Management and Organisation. *Organisational Research Methods, 18*(3), 429-472. https://doi.org/10.1177/1094428114562629

## APPENDIX

### List of Included Studies

This appendix presents a sample of the 131 studies included in the review. Due to space constraints, the full list of references with complete bibliographic details, DOIs, and access links is available online at: https://github.com/lamth1985/Appendix-A.-List-of-Included-Studies

ID#1.    Ahmadi, H., Yazdani, S., Babasalari, Z., & Rabi, A. (2020). Representation meaning of multimodal discourse in textbooks (A case study of Persian language-learning textbook). *Journal of Advanced Pharmacy Education & Research, 10*(2), 114–122. http://www.japer.in

ID#2.    Alshwaikh, J., & Morgan, C. (2013). Analysing the Palestinian school mathematics textbooks: A multimodal (multisemiotic) perspective. *Proceedings of the British Society for Research into Learning Mathematics, 33*(2), 70–75. http://www.bsrlm.org.uk/wp-content/uploads/2016/02/BSRLM-IP-33-2-13.pdf

ID#130. Yu, Z., & Doran, Y. (2024). Chemical formalisms: Toward a semiotic typology. *Semiotica, 2024*(259), 31–59. https://doi.org/10.1515/sem-2022-0110

ID#131. Zhu, Y., & Yang, N. (2023). Visual images and image-text relations in ELT textbooks for young learners. *Language Teaching for Young Learners, 5*(2), 196–216. https://doi.org/10.1075/ltyl.00035.zhu