Jupiter: Jurnal Pendidikan Terapan
Homepage : https://journal.diginus.id/jupiter/index
Email: jurnalpendidikanterapan@gmail.com

p-ISSN : 2985-3214; e-ISSN : 2964-3171
Jupiter, Vol. x, No. x, March 2025
Page 61-77

Optimizing Students' Language Skills Through a Multimodal Learning Model in Indonesian Language Learning in Elementary Schools: A Systematic Literature Review

Budi Febriyanto[1], Dadang Sunendar[2], Bachrudin Musthafa[3], Yuliawati[4], Agus Rofi'i[5]
[1,2,3,4]Universitas Pendidikan Indonesia, Bandung, Indonesia
[5]Universitas Majalengka, Majalengka, Indonesia

ARTICLE INFO

Article history:
Submitted: March 18, 2026
Final Revised: March 29, 2026
Accepted: March 31, 2026
Published: March 31, 2026

Keywords:
Elementary School; Indonesian Language Learning; Language Skills; Multimodal Learning; Systematic Literature Review

ABSTRACT

Background: Multimodal learning has gained increasing attention in language education because it enables learners to construct meaning through text, visuals, audio, gesture, space, and social interaction. However, the literature remains fragmented, and no integrated model has been clearly established for Indonesian language learning in elementary schools. **Purpose:** This study analyses the conceptual and pedagogical characteristics and design components of multimodal learning, its influence on students' language-skill development, and the conceptual, methodological, and assessment gaps in the literature. **Methods:** This study used a Systematic Literature Review (SLR) design. Articles indexed in Scopus were selected through the PRISMA flow. The search identified 277 records, and 44 reports were included in the final analysis. **Findings:** The integration of text, visuals, audio, gestures, social interaction, and meaning-making activities within structured instructional designs characterises multimodal learning. Across the reviewed studies, it tends to support reading, writing, speaking, listening, vocabulary development, and communicative competence. However, its effectiveness varies depending on instructional design, teacher readiness, student characteristics, and classroom context. The literature also remains conceptually, methodologically, and contextually fragmented, especially regarding Indonesian language learning in elementary schools. **Research implications:** The findings provide a conceptual foundation for developing Indonesian language instruction that is more contextual, participatory, and supportive of integrated language-skill development. They also offer guidance for designing more coherent instructional models, implementation strategies, and assessment systems for elementary school settings. **Conclusion:** Multimodal learning should be understood not merely as media variation, but as a design of meaning and learning experience. Future research needs to test integrated multimodal models directly in Indonesian elementary school language classrooms. **Originality:** This study systematically maps the conceptual foundations, pedagogical patterns, empirical trends, and research gaps in multimodal learning as a basis for developing Indonesian language-learning models in elementary schools. The review highlights that the existing literature remains fragmented and has yet to produce many fully integrated models for this context.

Doi: https://doi.org/10.61255/jupiter.v4i1.873

INTRODUCTION

The development of 21st-century literacy has shifted language learning beyond reliance on printed text alone toward communication practices that combine text, audio, visuals, gestures, space, and digital artefacts as resources for meaning-making. In this context, multimodality should not be understood simply as media variation, but as a pedagogical design that orchestrates relationships among modes to achieve measurable cognitive, communicative, affective, and social outcomes (Garcia, 2026; L. Li et al., 2025; Rahmanu & Molnár, 2024; Tong & An, 2026). This shift is especially relevant to Indonesian language learning in elementary schools, where listening, speaking, reading, and writing develop through concrete, contextual, and richly represented learning experiences.

Existing evidence shows that technology-based literacy instruction has a positive effect on elementary students' literacy achievement. However, the average effect remains relatively small, suggesting that instructional design quality is a decisive factor (Dahl-Leonard et al., 2024). Research on multimodal immersion also indicates that combining modes can strengthen vocabulary, reading, speaking, and writing and is positively associated with communicative competence (Boaventura et al., 2021; Rahmanu & Molnár, 2024). Tong & An (2026) further argue that the core of multimodal learning lies in the relationships among modes rather than in the number of media used, while Garcia (2026) shows that multimodal environments can simultaneously affect communicative performance, motivation, and cognitive control. Nguyen-Thi et al. (2025) add that multimodal experiences also shape learning emotions, including belonging, agency, self-confidence, and cognitive-emotional overload. Together, these findings suggest that Indonesian language learning in elementary schools requires a multimodal model that is not only rich in modes but also goal-aware, process-aware, and developmentally responsive.

The literature further shows that multimodal learning has developed into a pedagogy of production, design, interaction, and reflection. In digital multimodal composing, learning quality is measured not only by linguistic accuracy but also by how text, image, sound, and layout work together to reinforce meaning (Tong and An, 2026). At a more micro level, studies on multimodal reading, multimodal glossing, and reading-while-listening show that integrating modes can enrich attention, lexical representation, and phonological processing, thereby supporting reading and writing development (Malone et al., 2025; Ramezanali et al., 2021). From an instructional-design perspective, Yue et al. (2025) show that elementary students can act as artefact designers through design-based learning, while Hong and Kim (2025) report that real-product-based projects can improve literacy through active student engagement. Sheng et al. (2026) add that multimodal learning can be designed dialogically and progressively, with feedback operating from the word, sentence, to dialogue level, making it relevant to speaking development. Zhou et al. (2024) and (Heo et al., 2023) emphasise the role of locally relevant visual representations in fostering engagement and reasoning, while Hagerman et al. (2022) show that maker literacies create space for integrating the body, materials, images, and language in meaning-making. In the reflective domain, Umino (2023) and Farías and Véliz (2016) demonstrate that multimodality can also help map students' learning experiences and beliefs through narratives, images, and symbols. However, Meneses et al. (2023) caution that the benefits of multimodality depend on giving students sufficient opportunities for language production rather than maintaining teacher-dominated instruction. Ding et al. (2022) further stress that high-quality multimodal implementation requires teacher noticing, reflective practice, and sustained professional learning. These studies show that multimodal learning has broad pedagogical potential, but they also indicate that the field remains scattered across different contexts and purposes.

Several gaps remain clear in the literature. Conceptually, most studies are situated in foreign language learning, higher education, digital literacy, AI literacy, data literacy, or computational domains. As a result, they have not developed a direct multimodal pedagogical model for Indonesian language learning in elementary schools, with integrated targets for listening, speaking, reading, and writing. Methodologically, many studies still focus on a single skill, a single task type, or a single mode, while studies examining authentic classroom models that target the four language skills simultaneously remain limited (Alsubaie, 2022; Bai & Lei, 2025; Chung et al., 2024; Engman, 2021). In their assessment, Hadad et al. (2023) show that students' self-appraisals do not always align with actual performance, suggesting the need for performance-based evidence in multimodal learning. Tong & An (2026) also warn that, without rubrics capable of capturing relationships among modes, the quality of multimodal products may be inaccurately judged. In addition, continuance intention, task-technology fit, teacher efficacy, classroom emotional climate, and the quality of literacy opportunities all influence implementation outcomes (Condie & Pomerantz, 2020; Huang et al., 2024; Jensen et al., 2025). On a broader level, the literature has not yet fully integrated affective, social, critical, inclusive, and cultural-local dimensions into a single multimodal model for elementary school learners (Carter & Abbott, 2024; Goo et al., 2020; Melo-Pfeifer & Chik, 2022). These gaps highlight the need for a synthesis that maps the characteristics of multimodal models, explains their effectiveness, and identifies theoretical and practical directions for future model development.

Based on these conditions, this study aims to systematically analyse the conceptual and pedagogical characteristics and design components of multimodal learning, synthesise empirical findings on its influence on language-skill development, and identify conceptual, methodological, and assessment gaps together with directions for model development recommended in the literature. A Systematic Literature Review was chosen because it is well-suited to mapping a fragmented field, comparing findings across studies, and formulating an evidence-based agenda for model development. Within this framework, multimodality is treated as a promising approach whose effectiveness depends on designs that are mode-aware, context-aware, process-aware, and assessment-aware. The research questions are as follows: (1) What conceptual and pedagogical characteristics, and what design components, of multimodal learning are reported in the literature to support the development of language skills? (2) How do empirical findings

explain the influence of multimodal learning on language-skill development, and what conditions support or limit its effectiveness?; and (3) What conceptual, methodological, and assessment gaps remain in multimodal learning research, and what directions for model development are recommended in the literature? By addressing these questions, this study seeks not only to summarise previous research but also to consolidate and contextualise scattered findings into a scientific foundation relevant to Indonesian language learning in elementary schools. In this way, the study is expected to provide a stronger academic basis for developing an integrative, participatory, contextual, and evidence-based multimodal learning model for improving students' language skills.

METHOD

This study used a Systematic Literature Review (SLR) design to map the conceptual and pedagogical characteristics, instructional designs, empirical findings, and research gaps related to multimodal learning for the development of language skills in elementary school students. SLR was selected because it allows the review process to be conducted systematically, transparently, and replicably through explicit selection criteria, operational definitions, and reporting procedures. This design is also appropriate for research that aims to identify patterns in the literature, analyse gaps, and provide a foundation for future model development.

The data sources were journal articles retrieved from the Scopus database using keywords related to multimodal language learning and elementary student literacy. The review used Scopus as the sole database because it provides broad coverage of peer-reviewed international journals and robust metadata for indexing. To improve reproducibility, the exact search string used in Scopus was as follows: TITLE ((multimodal AND language AND learning) AND (elementary AND student AND literacy))AND PUBYEAR > 2019 AND PUBYEAR < 2027

Table 1. Inclusion and exclusion criteria

Aspect	Inclusion criteria	Exclusion criteria
Source type	Peer-reviewed journal articles indexed in Scopus Q1.	Books, book chapters, conference papers, editorials, notes, dissertations, and non-journal outputs.
Language	Articles published in English.	Publications in languages other than English.
Publication period	Studies published from 2020 to 2026.	Studies published before 2020 or after the search window.
Topical relevance	Studies addressing multimodal learning, multimodal instruction, multimodal reading or composing, or related multimodal pedagogies linked to language, literacy, or language-skill development.	Studies unrelated to instructional contexts.
Participants and context	Studies involving elementary or primary learners, teachers, teacher candidates, or adjacent educational contexts that provided direct conceptual or pedagogical evidence relevant to elementary language learning.	Studies with no clear educational relevance or with contexts too distant from the review purpose.
Outcomes	Studies reporting language, literacy, communicative, affective, or pedagogically relevant implementation outcomes.	Studies lacking analyzable outcomes or failing to report findings relevant to the review questions.
Data sufficiency	Articles with retrievable full texts for appraisal and extraction.	Records without abstracts for screening or full texts that could not be retrieved.

Quality appraisal was conducted after full-text screening using an adapted approach for different study designs. Empirical qualitative, quantitative, and mixed-methods studies were assessed based on the clarity of research questions, design appropriateness, sampling, data collection, and alignment between evidence and conclusions. Reviews and meta-analyses were additionally assessed for search transparency, eligibility criteria, synthesis procedures, and reporting of limitations, while computational or design-oriented studies were evaluated for dataset transparency, metrics, and reproducibility. Each study was rated as high, moderate, or low quality. Quality appraisal informed the interpretation of findings and research gaps, but studies were excluded only when the report lacked sufficient information for appraisal.

Table 2. Article Summary

ID	Authors	Year	Title	Context	Method	Participants	Key Findings
1	Peiru Tong & Irene Shidong An	2026	Synaesthesia in Digital Multimodal Composing: The Case of a Mobile-Assisted Task for Learning Chinese as a Foreign Language	University in Australia (The University of Sydney)	Qualitative analysis of student artifacts and interviews in a WeChat-based task for learners of Chinese as a…	CFL university learners	The study identified three main mechanisms of intermodal interaction: synaesthetic representation, interaction, and composition
2	Manuel B. Garcia	2026	Multilingual Language Learning in a Multimodal Metaverse: A Multidimensional Study of Communicative, Affective, and Cognitive Development	A technology university in the capital of the Philippines	Quasi-experimental, cluster-assigned pretest-posttest control group design with 80 students in an…	80 higher-ed students	The metaverse group showed greater gains in code-switching accuracy, spoken fluency, motivational engagement, and cognitive control than the…
3	Hongwei Sheng, Xin Shen, Heming Du, Xin Yu	2026	Mobile Auslan: A Multimodal Dialogue-Centered Sign Language Learning System	University of Queensland, Brisbane, Australia	System development combined with technical evaluation and a user study	Sign-language learners / user-study participants	The system supports sign-language learning through multimodal input from multiple viewpoints and feedback at the word, sentence, and dialogue levels
4	Jonathan Malone, Bronson Hui, Nick Pandža, Tetiana Tytko	2025	Eye Movements, Item Modality, and Multimodal Second Language Vocabulary Learning: Processing and Outcomes	University of Maryland, United States	Experiment with 119 learners of English comparing reading only versus reading while listening, supported by…	119 English learners	Multimodal reading was slightly slower, but it produced better phonological learning without harming orthographic learning
5	Miao Yue, Morris Siu-Yung Jong, Yun Dai, Wilfred Wing-Fat Lau	2025	Students as AI-Literate Designers: A Pedagogical Framework for Learning and Teaching AI Literacy in Elementary Education	Hong Kong	Mixed methods; intervention with Grade 5 students using pretest-posttest measures and focus group interviews	Grade 5 students	SAILD had a positive impact on AI literacy, especially skills, ethics, and attitudes, and showed how real-world problem-based task design can foster…
6	Li Mengyu, Wang Yuxing, Zan Ziqing, Liu Lizhu, You Lili	2025	Physical Literacy among Chinese Elementary School Students: The Mediating Role of Physical Knowledge and Physical Competency	China: Hebei, Sichuan, Qinghai, and Shenzhen	Cross-sectional survey with multistage cluster sampling of 3,091 elementary students; mediation analysis	3,091 elementary students	Physical competency mediated the relationship between physical motivation and participation by 56.40%, and physical literacy levels declined in…
7	Yulan Bai & Songhua Lei	2025	Cross-Language Dissemination of Chinese Classical Literature Using Multimodal Deep Learning and Artificial Intelligence	China and Japan (authors affiliated with Nanchang, China, and…	Development of the TMNMT model and comparative testing against baseline systems	Benchmark datasets	The TMNMT model outperformed baselines on BLEU and METEOR scores and improved both multimodal and text-only translation performance
8	Lu Li, Xiuxiu Bai, Junxiu Xu, Dingkang Wang, Tao Jiang	2025	Multimodal Learning Audio-Visual Detection for Obtaining Object-Level Sound Sources in Japanese-Language Teaching Rooms	Xi'an, China	Development of the AVDor model and benchmark evaluation	Benchmark datasets	The study demonstrates the feasibility of audio-visual detection of object-level sound sources in Japanese-language classrooms
9	Ying Yang, Yan-Qiu Yang, Gang Ren, Ben-Gong Yu	2025	Hierarchically Trusted Evidential Fusion Method with Consistency Learning for Multimodal Language Understanding	Hefei, China	Development of the HTEF model and testing on the MIntRec and MELD datasets	Benchmark datasets	The model outperformed the state of the art in accuracy, generalization, and reliability for intent and emotion recognition
10	Minh-Hong Nguyen-Thi, Khai-Xuan Tran, Thien-Vu Giang	2025	Exploring the Emotional Experience in Learning Chinese as a Second Language of Students from the Multimodal Affective Perspective: A Case Study in Vietnam	A public secondary school in Vietnam	Qualitative case study of 15 students over five months; data from interviews, classroom observations, and…	15 CSL learners	Four major themes were found: joy/ownership, cognitive-emotional overload, peer-mediated regulation, and empowerment/pride
11	I Wayan Eka Dian Rahmanu; Gyöngyvér Molnár	2024	Multimodal Immersion in English Language Learning in Higher Education: A Systematic Review	Literature study (higher education context across studies)	Systematic review	Prior studies (review corpus)	The study synthesizes how multimodal immersion is used in English language learning in higher education and maps the direction of the field
12	Eniafe Festus Ayetiran; Özlem Özgöbek	2024	A Review of Deep Learning Techniques for Multimodal Fake	Computational literature review; multimodal dataset and benchmark…	Review/survey of deep learning techniques for multimodal content detection	Prior studies / benchmark papers	The article maps techniques, datasets, challenges, and research agendas for multimodal

ID	Authors	Year	Title	Context	Method	Participants	Key Findings
			News and Harmful Languages Detection				fake-news and harmful-language detection
13	Eniafe Festus Ayetiran; Özlem Özgöbek	2024	An Inter-Modal Attention-Based Deep Learning Framework Using Unified Modality for Multimodal Fake News, Hate Speech and Offensive Language Detection	Computational study evaluated on public benchmark datasets	Development of a deep-learning model tested on benchmark datasets, including ablation experiments	Benchmark datasets	The framework achieved strong performance on multimodal detection tasks compared with baselines through inter-modality attention and unified modality
14	Yan Huang; Wei Xu; Paisan Sukjairungwattan; Zhonggen Yu	2024	Learners' Continuance Intention in Multimodal Language Learning Education: An Innovative Multiple Linear Regression Model	China (N=165)	Survey/questionnaire combined with multiple linear regression	Language learners (survey respondents)	The findings show that acceptance/fit and personal investment factors contribute to continuance intention in multimodal language learning
15	Aruna Gladys A.; Vetriselvi V.	2024	Sentiment Analysis on a Low-Resource Language Dataset Using Multimodal Representation Learning and Cross-Lingual Transfer Learning	India (authors' affiliations); multimodal dataset and low-resource…	Computational experiments on several datasets (e.g., MOSI/MOSEI/MELD/MSAT) and model evaluation	Multimodal sentiment datasets	The study shows strong performance for multimodal sentiment analysis, although cross-lingual transfer may introduce overfitting issues in certain…
16	Maria Therese Jensen; Oddny Judith Solheim; Espen Olsen	2024	Leader Support in Relation to Teacher Self-Efficacy, Classroom Emotional Climate and Students' Literacy Skills in Elementary School	Norway; 5,810 first-grade students from 300 classes	Teacher and student surveys plus literacy tests analyzed using Structural Equation Modeling	Teachers and elementary students	Leader support was linked to teacher self-efficacy for discipline and motivation; discipline-related self-efficacy predicted classroom emotional…
17	Katlynn Dahl-Leonard; Colby Hall; Delanie Peacott	2024	A Meta-Analysis of Technology-Delivered Literacy Instruction for Elementary Students	Meta-analysis (K-5 across studies)	Meta-analysis of 53 studies using random-effects and moderator analysis	53 primary-literacy studies	The overall effect was small but positive (Hedges g≈0.24), and moderator analyses did not identify consistent determinants
18	JiYeon Hong; Kwihoon Kim	2024	Impact of AIoT Education Program on Digital and AI Literacy of Elementary School Students	Korea (Seoul and Gyeonggi); Grade 6 students (n=24)	Program development and effectiveness testing using a pre-post paired t-test	Grade 6 elementary students	There was a significant improvement in all aspects of digital literacy and AI literacy after the intervention
19	LeighAnn R. Bedwell; Jackie M. Lodato; Joshua D. Danish	2024	Using Network Visualizations to Engage Elementary Students in Locally Relevant Data Literacy	Midwestern United States; Grade 5–6 students (n=22)	A three-week Net.Create intervention with pre/post-tests and interaction analysis	Grade 5–6 students	Students improved their data literacy understanding, and they were motivated when the data were locally relevant and visualized
20	Kimin Chung, Soohwan Kim, Yeonju Jang, Seongyune Choi, dan Hyeoncheol Kim.	2024	Developing an AI Literacy Diagnostic Tool for Elementary School Students	Republic of Korea (Seoul, South Korea), as indicated by the authors'…	Research and development / instrument development and validation study	15 experts; 287 and 293 Grade 6 students	The diagnostic tool developed in the study was valid and reliable for measuring AI literacy in elementary students
21	Kristen D. Schwartz; Mary Abbott; Jessica L. Courtney	2023	Literacy Teachers in the Making: A Look at Teacher Candidates' Experiences as They Tutor Elementary Students	Teacher education program in the western United States; tutoring…	Single case study with qualitative analysis of teacher candidates' reflections	Teacher candidates	There was a shift from focusing on "doing activities" to focusing on students' needs, and reflection helped teacher candidates organize their…
22	Hadad et al.	2023	Digital Literacy among Arab Minority Students: Comparing Self-Assessment and Digital Task Performance	Israel (Arab minority elementary and middle school students)	Comparison of self-appraisal and digital-task performance using correlation and group analyses	Elementary and middle school students	Correlations between perception and performance were generally weak, and actual performance was low, especially in social-emotional literacy
23	Tae Umino	2023	Using Multimodal Language Learning Histories (MLLHs) to Understand Learning Experiences and Beliefs of L2 Learners in Japan	Japan (university students; Tokyo University of Foreign Studies)	Qualitative study using visual-content analysis and written narratives, classification of visual elements,…	21 L2 learners in Japan	Five visual categories were identified (language, place, person, resource, and analysis) along with four MLLH patterns (person-, resource-,…
24	Meneses, Uccelli, & Valeri	2023	Teacher Talk and Literacy Gains in Chilean Elementary Students (Teacher Participation, Lexical Diversity, Instructional Non-Present Talk)	Chile (Pre-K to Grade 2 classrooms)	Lesson observation and coding in 16 classrooms, with baseline and end-of-year literacy tests for 343…	343 elementary students; 16 classrooms	A higher ratio of teacher talk predicted lower literacy gains

ID	Authors	Year	Title	Context	Method	Participants	Key Findings
25	Yoon, Choi, & Choi	2023	Multimedia Analysis of Robustly Optimized Multimodal Transformer Based on Vision and Language Co-Learning	Global dataset context (tweets and images from 2017 disasters across…	Computational experiments comparing unimodal and multimodal models (SF/FF/multimodal BERT) on CrisisMMD	CrisisMMD and related datasets	RoBERTaMFT outperformed unimodal and multimodal baselines in accuracy and F1 across tasks
26	Heo, Kang, & Seo	2023	Natural-Language-Driven Multimodal Representation Learning for Audio-Visual Scene-Aware Dialog System	South Korea (authors' affiliations); evaluated using audio-visual…	Model development with quantitative and qualitative evaluation, achieving state-of-the-art results on…	Audio-visual dialogue datasets	The model addressed two problems: underuse of audio and limited interpretability, and it performed strongly in three-modality settings
27	Al Otaiba et al.	2023	What We Know and Need to Know about Literacy Interventions for Elementary Students with Reading Difficulties and Disabilities, including Dyslexia	United States (synthesis of studies from 2010–2020)	Synthesis of meta-analyses and systematic reviews, a review of reviews	Prior reviews and meta-analyses	There is converging evidence for the effectiveness of reading interventions, and there is also evidence that writing can support reading through…
28	Minna Maijala	2023	Multimodal Postcards to Future Selves: Exploring Pre-Service Language Teachers' Process of Transformative Learning during a One-Year Teacher Education Programme	University of Turku, Finland	Qualitative study of visual and written data collected at the beginning (N=51) and end (N=43) of the program	Pre-service language teachers	Disorienting dilemmas were evident at the beginning, especially around self-confidence, classroom work, and subject knowledge
29	Mary R. Hermes, Mel M. Engman, Meixi, & J. McKenzie	2023	Relationality and Ojibwemowin in Forest Walks: Learning from Multimodal Interaction about Land and Language	Ojibwe lands / Mille Lacs Band of Ojibwe reservation, Minnesota,…	Qualitative micro-interactional analysis of selected episodes from a corpus of forest walks recorded with…	Ojibwe Elder-child interactions	Language is understood as something living and tied to the body and the land
30	Hagerman et al.	2022	Literacies in the Making: Digital-Physical Literacy Practices while Making Musical Instruments from Recycled Materials	Canada (Ottawa; French-language school; urban community)	Qualitative study of a maker project using process observation and students' multimodal artifacts	Elementary students in a maker project	Students used many modes—images, color, text, hyperlinks, and more—to communicate their process, but transferring sensory/verbal modes into writing…
31	Ding, Glazewski, & Pawan	2022	Language Teachers and Multimodal Instructional Reflections during Video-Based Online Learning Tasks	United States (affiliations; video-based online reflective learning…	Case study (n=5 language teachers) in a video-based online learning environment	5 language teachers	Teachers evaluated their teaching through multimodal interactions, both verbal and nonverbal, such as pauses, gaze, and body language, and developed…
32	Li et al.	2022	Dual Coding or Cognitive Load? Effects of Multimodal Input on EFL Vocabulary Learning	China (Guangzhou; junior secondary EFL students)	Mixed methods: quasi-experiment comparing a multimodal experimental group and a monomodal control group, with…	EFL learners	The experimental group performed better on the immediate posttest but worse on the delayed posttest
33	Ana Pellicer-Sánchez	2022	Multimodal Reading and Second Language Learning	Not specific (review article; L2 context)	Literature review with emphasis on eye-tracking evidence and multimodal-processing research	Prior L2 multimodal reading studies	The review summarizes findings showing that images attract attention, L2 readers still spend much time on text, audio can increase image processing…
34	Merfat Ayesh Alsubaie	2022	Distance Education and the Social Literacy of Elementary School Students during the Covid-19 Pandemic	Saudi Arabia (Grade 3 elementary teachers; public schools)	Qualitative phenomenological study using in-depth interviews with six elementary teachers	6 elementary-school teachers	The main findings concern factors affecting social literacy, the positive and negative effects of distance education, and the challenges of teaching…
35	Jiexin Lin, Haomin Zhang, & Xiaoyu Lin	2022	Prosodic Transfer in English Literacy Skills among Chinese Elementary-Age Students: Controlling for Non-Verbal Intelligence	Grade 4 Mandarin-speaking learners of English in China (authors…	Quantitative study using mediated multivariate analyses with moderation on 224 fourth-grade students; it…	224 fourth-grade students	There was a positive relationship between first-language tone awareness and English reading/spelling, mediated by stress awareness, and this effect…
36	Maria Filomena Caldeira dkk. (tercantum banyak afiliasi)	2021	Promoting Ocean Literacy in Elementary Students	Portugal (Lisbon/Cascais) with collaboration from the United Kingdom…	Mixed methods: pre-post test, platform-data analysis, and content analysis of products/exhibitions	Elementary students in an ocean-literacy program	Knowledge increased from pretest to posttest, and students also demonstrated ICT competencies and public-communication products such as posters,…

ID	Authors	Year	Title	Context	Method	Participants	Key Findings
37	F. Liu dkk.	2021	DiMBERT: Learning Vision-Language Grounded Representations	Computational research using datasets such as MSCOCO, RefCOCO+, and…	Computational experiments involving pretraining and evaluation across several vision-language tasks	Vision-language benchmark datasets	DiMBERT improved performance in captioning, storytelling, and referring-expression tasks, and ablation analysis showed the contribution of the DiM…
38	Amy L. Ferrell	2021	Exploring Critical Literacy for Elementary Students with Disabilities	Special-education context in the United States; the researcher worked…	Qualitative/case study; participant-observer, one-on-one work over three months using mnemonics, pictures,…	3 elementary students with disabilities	Students produced nuanced responses about power, injustice, truth, vulnerability, and self-worth, although discussions were still largely…
39	Mel M. Engman	2021	A Worksheet, a Whiteboard, a Teacher-Learner: Leveraging Materials and Colonial Language Frames for Multimodal Indigenous Language Learning	A kindergarten classroom in a K-12 tribal school on Ojibwe land in…	Qualitative study using linguistic ethnography, critical discourse analysis, and multimodal social-semiotic…	Kindergarten Ojibwe teacher-learners	The teacher transformed worksheets and the whiteboard, which were initially English-oriented, into a space for Ojibwe languaging; materials and…
40	Nasrin Ramezanali; Takumi Uchihara; Farahnaz Faez	2020	Efficacy of Multimodal Glossing on Second Language Vocabulary Learning: A Meta-Analysis	Meta-analysis (across studies)	Meta-analysis of 22 studies (26 effect sizes) with moderator analysis	Review/meta-analysis corpus	The added effect of multimodal glossing was moderate on the immediate posttest (g≈0.46) and small on the delayed posttest (g≈0.28)
41	Minkowan Goo; Diane Myers; Adela L. Maurer; Robert Serwetz	2020	Effects of Using an iPad to Teach Early Literacy Skills to Elementary Students with Intellectual Disability	Suburban public elementary school, South Central United States	Single-subject study using a multiple-probe design across students	3 students with intellectual disability	The iPad intervention with visual supports effectively improved phoneme-segmentation fluency in three students with intellectual disability
42	Cami Condie & Francesca Pomerantz	2020	Elementary Students' Literacy Opportunities in an Age of Accountability and Standards: Implications for Teacher Educators	Elementary classrooms in Massachusetts, United States, within an…	Qualitative study observing two literacy lessons across 14 preservice and in-service teachers, supported by…	Observed elementary classrooms and students	Students' literacy opportunities varied greatly; some classes had little or no reading, little writing, and low-level speaking/listening
43	Julie-Ann Scott	2020	(Re)directing a University Storytelling Troupe for At-Risk Elementary Students for Course Credit: A Story of Embodied Empathy, Literacy, and Personal Transformation	A University of North Carolina Wilmington program performing in Title…	Qualitative-reflective essay / performance ethnographic account of the redesign of one storytelling program	At-risk elementary students; university storytelling troupe	Embodied adaptations of children's literature performances can foster human connection, empathy, reading interest, critical reasoning, and student…
44	Sílvia Melo-Pfeifer & Alice Chik	2020	Multimodal Linguistic Biographies of Prospective Foreign Language Teachers in Germany: Reconstructing Beliefs about Languages and Multilingual Language Learning in Initial Teacher Education	University of Hamburg, Germany; prospective Spanish-language teachers	Qualitative analysis of 33 visual linguistic biographies/drawings	33 prospective language teachers	Participants tended to view multilingualism chronologically and as separated by language or country, yet still showed multimodal translanguaging…

The instruments used in this review were a search protocol, an inclusion-exclusion selection sheet, and a data-extraction matrix. The extraction matrix contained at least the title, research location, author, year of publication, theoretical foundation, method, main findings, and study limitations, following the structured format prepared for this review (Wahyudi, 2024).

The PRISMA numbers were clarified as follows. The database search identified 277 records. Before screening, 158 records were removed because they fell outside the review protocol, for example, because of publication type, language, year range, or insufficient bibliographic information, leaving 119 records for title and abstract screening. At the screening stage, 44 records were excluded for not aligning with the topic focus or the review scope, leaving 75 reports for retrieval. Of these, 31 reports could not be retrieved in full text. The remaining 44 reports were assessed for eligibility, and all 44 satisfied the inclusion criteria and were therefore included in the final synthesis. Thus, the apparent difference between the PRISMA numbers reflects the distinction between records screened, reports sought for retrieval, reports not retrieved, and reports finally included. The data-collection procedure was conducted in several stages: article identification, removal of records that did not meet the criteria, title and abstract screening, full-text retrieval, eligibility assessment, and data extraction into a synthesis matrix.

The data were then analysed at two levels. First, descriptive analysis was used to map publication trends, research designs, contexts, and participant profiles. Second, a thematic-comparative analysis was used to answer RQ1, RQ2,

and RQ3 by coding intermodal orchestration, effects on language skills, supporting and limiting conditions, and conceptual, methodological, and assessment gaps.

This analytical approach is consistent with recommendations in the literature that multimodal studies should capture not only learning outcomes, but also the quality of instructional design, the context of implementation, and evidence of learning across modes more measurably and transparently (Dahl-Leonard et al., 2024; Hadad et al., 2023; Hong & Kim, 2025; Tong & An, 2026; Zhou et al., 2024).

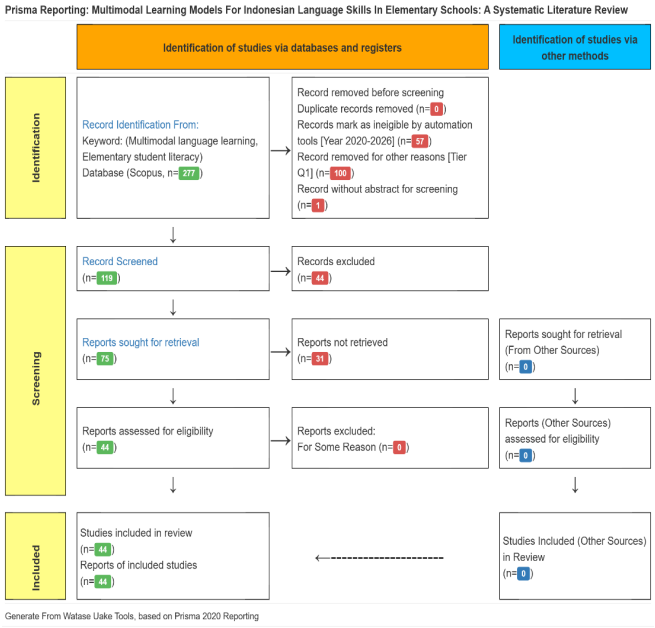

Figure 1. PRISMA 2020 Graphic Report

RESULTS

A. Study identification results and publication trends

The study identification process followed the PRISMA framework, beginning with the retrieval of records from the Scopus database using predefined title-based keywords: "multimodal language learning" and "elementary student literacy." At the identification stage, the keyword search produced a broader pool of records, as reflected in the raw search counts shown in the system output. However, these initial records did not directly represent the final dataset of the review, as duplicate handling, relevance screening, eligibility assessment, and application of the predefined inclusion and exclusion criteria were subsequently undertaken. After the full screening process had been completed, 44 studies were retained for inclusion in the final review.

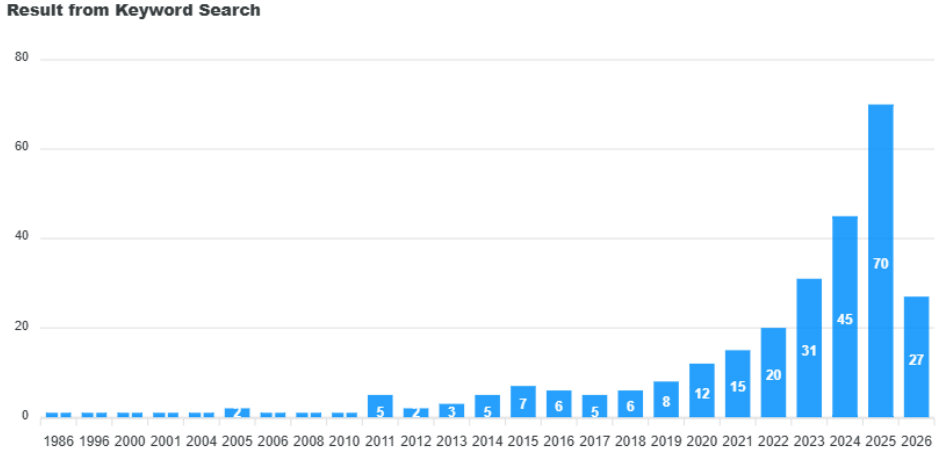

Figure 2. The annual distribution of Scopus search results was identified at the initial search stage, before screening and eligibility assessment

To ensure transparency in reporting, it is important to distinguish between the initial search output and the final included studies. The publication trend graph presented in Figure 2 reflects the annual distribution of the broader search results retrieved from Scopus rather than the yearly distribution of the 44 studies included in the final synthesis. This interpretation is evident because the graph includes records published before 2020. In contrast, the review itself was restricted to studies published between 2020 and 2026, and the cumulative number represented in the graph exceeds the final number of included studies. Accordingly, the graph should be interpreted as illustrating the broader development of publications on the search topic in Scopus, not the temporal profile of the final included sample alone.

The graph nevertheless provides a useful overview of the increasing scholarly attention to the topic. As shown in Figure 2, publications on multimodal language learning and elementary student literacy remained limited in earlier years. Still increased substantially after 2020, with a marked rise from 2022 onward and a peak in 2025. This pattern suggests growing academic interest in the field, particularly in recent years. The lower count for 2026 should be interpreted cautiously, as it may reflect incomplete indexing or an unfinished publication year rather than an actual decline in research output.

B. General profile of the reviewed studies

The reviewed studies vary in context, participants, research design, and outcome focus. In terms of context, the studies come from higher education, elementary and secondary schools, teacher education programs, and computational research using multimodal datasets. At the elementary-school level, several studies explicitly involved Grade 5 students, Grade 6 students, Grade 5–6 students, Grade 1 students, students with intellectual disabilities, and third-grade elementary school teachers.

The reviewed studies also show strong methodological variation. The corpus includes systematic reviews, meta-analyses, quasi-experimental pretest–posttest studies, mixed-methods studies, case studies, single-subject designs, surveys using regression and SEM, and computational experiments. Examples include cluster-assigned pretest–posttest control group designs in Garcia's study, mixed-methods designs in Yue et al., single-subject multiple-probe designs in Goo et al., meta-analyses in Dahl-Leonard et al. and Ramezanali et al., and qualitative case studies in Nguyen-Thi et al. and Ding et al.

The reported outcomes are equally diverse. They include language-skill scores, vocabulary, phonological processing, digital literacy, data literacy, learning engagement, emotions, and continuance intention. Some studies focus directly on pedagogy, while others focus on computational integration of text, images, audio, and video through fusion, attention, and representation learning. Overall, the evidence shows that multimodal learning research is cross-level, cross-approach, and cross-form.

C. Conceptual, pedagogical, and design characteristics of multimodal learning

The synthesis for RQ1 shows that the reviewed studies define multimodal learning as the purposeful integration of multiple representational modes. At the conceptual level, Tong and An identify three mechanisms of intermodal interaction: synaesthetic representation, interaction, and composition. Rahmanu and Molnár describe multimodal immersion as a combination of visual, audio, textual, action-based, gestural, and environmental inputs.

At the pedagogical level, several recurring characteristics appear across the reviewed studies. These include the production of multimodal documents, transformation across modes, dialogue and social interaction, gradual scaffolding, and reflection on learning experiences. The most frequently reported design forms include digital multimodal composing, multimodal reading, glossing, reading while listening, project-based production, design-based learning, maker activities, network visualisation, and dialogue-centred learning.

More specific design patterns also appear in individual studies. Yue et al. position elementary students as designers of solutions and artefacts, while Hong and Kim report product-based activities in an AIoT program. Sheng et al. design a learning system with feedback at the word, sentence, and dialogue levels. Garcia structures a multimodal environment through language zones, interactive objects, visual cues, and spatial scenarios. Pellicer-Sánchez, Malone et al., and Ramezanali et al. describe multimodal reading designs that combine text with images, audio, or digital glosses. Hagerman et al. and Farías and Véliz show that instructional design can also involve images, colour, narrative, text, the body, and physical materials as part of meaning-making communication. Across the reviewed literature, the most frequently reported design components are text, visuals, audio, gestures, space, digital or physical artefacts, production tasks, and the assessment of multimodal outcomes.

D. The effects of multimodal learning on language skills and related variables

The synthesis for RQ2 shows that multimodal learning influences both language-learning outcomes and related instructional outcomes. Rahmanu and Molnár report that multimodal approaches have been used to improve vocabulary, reading, speaking, writing, and communicative competence. Dahl-Leonard et al. report a small overall

positive effect of technology-based literacy instruction among elementary school students, with a Hedges' g of around 0.24. Ramezanali et al. report a moderate immediate posttest effect for multimodal glossing, with g of around 0.46, and a small delayed posttest effect, with g of around 0.28.

Several individual studies support these positive trends. Malone et al. find that the reading-while-listening condition improves phonological learning without disadvantaging orthographic learning, though it slightly slows reading. Garcia reports that the metaverse group showed greater improvement in code-switching accuracy, spoken fluency, motivational engagement, and cognitive control than the traditional learning group. Nguyen-Thi et al. identify four affective themes in multimodal learning: joy/ownership, cognitive-emotional overload, peer-mediated regulation, and empowerment/pride. At the elementary-school level, Yue et al. report positive effects of SAILD on skills, ethics, and attitudes; Hong and Kim report significant improvement across all aspects of digital literacy and AI literacy; Bedwell et al. report gains in data-literacy understanding; and Goo et al. report improved phoneme-segmentation fluency among three elementary students with intellectual disability.

However, the findings are not uniformly positive. Li et al. report that the multimodal group outperformed the comparison group on the immediate posttest but scored lower on the delayed posttest. Alsubaie finds that distance learning improved some social skills, such as listening and participation, but reduced others because of limited physical interaction. Huang et al. report that acceptance, fit, and personal investment contributed to continuance intention in multimodal language learning. Jensen et al. note relationships among leader support, teacher self-efficacy, classroom emotional climate, and literacy ability in elementary school students. Overall, the reviewed evidence suggests that multimodal learning tends to support language development. Still, its effects depend on the target skill, the task structure, the learning environment, and the quality of instructional design.

E. Conceptual, methodological, and assessment gaps

The synthesis for RQ3 shows that the most frequently reported gaps appear at the conceptual, methodological, and assessment levels. At the conceptual level, the focus of existing studies remains fragmented across reading, composing, glossing, digital literacy, AI literacy, data literacy, affective learning, and multimodal computation. These areas have not yet been fully integrated into a single comprehensive pedagogical model for language learning.

At the contextual level, most reviewed studies are not situated directly in Indonesian language learning in elementary schools. Instead, they come from higher education, L2/CFL contexts, computational fields, or non-language elementary school settings. At the methodological level, the review shows a strong presence of descriptive qualitative studies, along with fewer experimental, review, and survey studies. Studies that test authentic classroom models targeting the integrated four language skills remain scarce.

At the assessment level, Hadad et al. report that the correlation between self-appraisal and actual performance tends to be weak. Tong and An note the need for rubrics that can capture relationships among modes. Several computational studies also report issues such as the heterogeneity gap, semantic gap, attention conflict, and the need for multi-evidence integration across text, image, audio, and video. The literature also shows inconsistent retention findings, as seen in Li et al. and the meta-analysis by Ramezanali et al.

In terms of implementation, the reviewed studies mention variables such as teacher noticing, teacher reflection, teacher self-efficacy, classroom emotional climate, and continuance intention, but they usually examine these variables separately. In terms of student differentiation, Goo et al., Al Otaiba et al., and other studies on social and affective literacy indicate the need for explicit support, accessibility, and inclusive conditions. However, these findings have not yet been consolidated into a single model design. Overall, the review shows a lack of studies that directly combine intermodal relationships, the four language skills, learning emotions, social interaction, performative assessment, retention, and the Indonesian elementary-school context within one research model.

DISCUSSION

The reviewed literature positions multimodal learning not simply as the use of multiple media, but as the deliberate orchestration of semiotic modes to construct meaning, direct attention, and expand students' ways of expressing understanding. In this view, multimodality involves the functional relationships among text, visuals, audio, gestures, space, and social interaction. Therefore, learning quality depends more on how these modes support one another than on the number of modes used. This interpretation aligns with Tong and An (2026), who place intermodal relations at the centre of multimodal composition quality, and with (W. Li et al., 2022), who shows that multimodal environments can function as learning ecologies that shape communicative, affective, and cognitive processes.

This synthesis also shows that multimodality in language learning works at two connected levels. At the semiotic level, it organises relationships among representations. At the psychological level, it shapes how students select, organise, and integrate information across modes. Rahmanu & Molnár (2024) emphasise that multimodal pedagogies

have been used to improve vocabulary, reading, speaking, and writing, while Pellicer-Sánchez (2022), Malone et al. (2025), and Ramezanali et al. (2021) show that combinations of text, images, and audio affect attention allocation, lexical representation, and phonological encoding. Taken together, these findings shift the discussion from simple media variation to the design of language experience. For Indonesian language learning in elementary schools, this means that listening, speaking, reading, and writing should be designed as interconnected activities supported by multiple modes, rather than as isolated skills taught through single-channel instruction (Hermes et al., 2023; Scott, 2020).

At the pedagogical level, the synthesis reveals a stable pattern across the reviewed studies. Effective multimodal learning repeatedly involves artefact production, intermodal transformation, social interaction, gradual scaffolding, and reflection. Yue et al. (2025) and Lin et al. (2022) show that elementary students can act as artefact designers through design-based learning. Hong & Kim (2025) show that product-based activities can improve technological literacy in elementary settings. Hagerman et al. (2022) further highlight the value of maker literacies and embodied literacies that combine text, images, the body, colour, and physical materials. Zhou et al. (2024) add that locally relevant visual representations can strengthen engagement and reasoning. Sheng et al. (2026) and Yang et al. (2025) show that multimodal learning can also be organised progressively from word to sentence to dialogue with structured feedback. These findings suggest that Indonesian language instruction in elementary schools should move beyond decorative media use. A stronger model would integrate multimodal reading tasks, multimodal composition or production, dialogue or performance activities, locally grounded projects, teacher and student reflection, and alignment among technology, tasks, and language goals. The value of this synthesis lies in its ability to organize scattered findings into a more stable pedagogical pattern that can support later model development.

The effects of multimodal learning on language skills are generally positive, though not uniform. Rahmanu and Molnár (2024) show broad support for vocabulary, reading, speaking, writing, and communicative competence. Ramezanali et al. (2021) show benefits for vocabulary learning through multimodal glossing. Malone et al. (2025) show that reading while listening can strengthen phonological learning without harming orthographic learning. Pellicer-Sánchez (2022) shows that texts combined with images and or audio can support mental-model construction and inference when the visuals serve meaningful functions rather than decorative ones. Garcia (2026) and Sheng et al. (2026) also suggest that multimodal learning can strengthen spoken performance, engagement, and communicative participation. On the affective side, Nguyen-Thi et al. (2025) and Ayetiran & Özgöbek (2024) show that multimodal tasks can foster ownership, agency, and empowerment. However, cognitive-emotional overload when task demands are too high. These findings are relevant to Indonesian language learning in elementary schools because they show that language instruction need not remain fragmented and monomodal. The literature instead supports a model that integrates multimodal input, meaning processing, and language output. However, the evidence also shows that effectiveness is conditional. Tong & An (2026) and (Yoon et al., 2023) show that multimodality can remain superficial when intermodal relations are weak. Li et al. (2022) show that immediate gains may not persist in delayed posttests when cognitive load, distraction, or redundancy increases. Huang et al. (2024) show that continuance intention depends on task-technology fit, user acceptance, and personal investment. Meneses et al. (2023) show that multimodality is less effective when opportunities for student language production are limited by teacher-dominated discourse. Condie & Pomerantz (2020) further show that classroom literacy opportunities, leadership support, teacher self-efficacy, and emotional climate influence outcomes.

For this reason, the most useful question is no longer whether multimodality is effective in general, but for whom, for which tasks, under which conditions, and through what kind of design. This review, therefore, highlights the need for coherence among modes, language goals, students' developmental levels, classroom interaction, teacher competence, and assessment systems. In the Indonesian elementary school context, this point is especially important because multimodality should support language development, not distract from it or reduce learning to a technological display.

The review also reveals a clear conceptual gap. Existing studies remain dispersed across multiple orientations, including reading, digital composition, social interaction, digital and AI literacy, data literacy, and computational representation. As a result, the literature still lacks a model that integrates semiotic, cognitive, affective, social, and pedagogical dimensions into a single framework for basic language learning. Most of the reviewed studies also do not directly develop a multimodal pedagogical model for Indonesian language learning in elementary schools, with integrated targets in listening, speaking, reading, and writing. Even when studies are conducted in elementary settings, they often focus on one limited area, such as AI literacy, digital literacy, early phonological skills, or classroom emotional climate.

A further gap appears between computational studies and pedagogical studies. Computational work contributes useful concepts such as fusion, attention, consistency, and conflict control, but it rarely provides direct classroom

guidance (Ayetiran & Özgöbek, 2024a, 2024b; Gladys & Vetriselvi, 2024). Pedagogical studies, in contrast, often provide concrete activities but do not always explain intermodal mechanisms clearly. Based on this synthesis, a future Indonesian-language multimodal learning model for elementary schools should treat modes not merely as channels, but as functional components within a clear instructional syntax. Such a model should integrate multimodal reading, artefact production, dialogue or performance, reflection, and portfolio assessment, while ensuring that each mode serves an explicit linguistic function. The main novelty would therefore lie not in the use of multimedia itself, but in the operationalisation of an integrated, contextually grounded, and language-functionally grounded multimodal pedagogy.

From a methodological perspective, this review shows that many studies remain concentrated at the descriptive-qualitative level, highly controlled experiments on specific components, or perception-based surveys. Studies that test authentic classroom models targeting the four language skills together are still limited. Dahl-Leonard et al. (2024) show that technology-based literacy instruction has a positive overall effect in elementary settings, but the moderators of this effect remain unclear. Li et al. (2022) show that retention needs greater attention. Hadad et al. (2023) show that self-appraisal alone is not a sufficient indicator of performance. Tong & An (2026) show that assessment quality depends on rubrics that can capture intermodal relationships. Meneses et al. (2023), Condie & Pomerantz (2020), Jensen et al. (2025), Huang et al. (2024), and Maijala (2023) further indicate that classroom interaction, implementation context, emotional climate, and teacher reflection all shape learning outcomes, yet these variables are rarely examined together.

This methodological pattern suggests a clear direction for future research in Indonesian language learning. Studies should use mixed-methods or quasi-experimental designs that combine process, outcome, retention, and implementation-context measures. Assessment should also move beyond single test scores to include language-skill measures, multimodal portfolios, classroom-interaction documentation, student reflection, and teacher reflection. In this sense, the contribution of this review is not only substantive but also methodological, because it clarifies how the field can build stronger, more relevant scientific evidence.

Significance of the Research Findings and Their Contribution to the Field of Indonesian Language Learning in Elementary Schools.

The findings of this study are important because they show that Indonesian language learning in elementary schools provides a strategic context for developing an integrative multimodal model. The subject includes reading, writing, listening, speaking, presenting, literary appreciation, reflection, and the development of linguistic identity. These dimensions fit well with multimodal orchestration.

The first contribution is theoretical. This study shifts the discussion of multimodality from a narrow focus on media to a broader focus on intermodal relations, learning ecology, and the design of language experience. The second contribution is pedagogical. This study organises previous findings into a design pattern that can be translated into a learning sequence: multimodal reading or listening, meaning processing, artefact production, presentation, reflection, and revision. The third contribution is methodological. This study emphasises the need for performative evaluation and multimodal portfolios rather than relying only on self-perception or a single score. The fourth contribution is practical. This study provides a scientific basis for Indonesian language teachers in elementary schools to design learning that is more contextual, participatory, and aligned with how children construct meaning today.

More broadly, this review bridges fragmented international literature, both pedagogical and computational, into a more focused orientation for Indonesian language learning in elementary schools. It also shows that developing a multimodal learning model for this context still has significant originality potential, as the current literature does not yet provide a model that explicitly integrates the four language skills, learning emotions, social interaction, student differentiation, and evidence-based assessment within a single framework.

Implications of the Research for Theory, Teaching Practice, Assessment, and Future Research Agendas

Theoretically, this study implies that multimodality should be positioned as a pedagogical design framework in Indonesian language learning rather than as an additional medium. The theory of Indonesian language learning in elementary schools, therefore, needs to more explicitly integrate relationships among modes, language functions, affective experience, and classroom social context.

In teaching practice, teachers need to design activities that link each mode with clear skill targets. Visuals can support reading inference, audio can support pronunciation and prosody, text can support discourse structure, and dialogue can support pragmatic response and confidence in language use. The literature also shows the importance of gradual scaffolding. Without cognitive-load regulation, multimodality may create distraction or reduce retention.

At the school level, these findings emphasise the importance of organisational support, teacher training, and the classroom emotional climate. The success of multimodality depends not only on devices but also on teacher efficacy,

teacher noticing, reflective practice, and the quality of classroom interaction. In assessment, teachers and researchers need rubrics that evaluate language quality, intermodal coherence, the function of each mode, student participation, and evidence of revision.

For future research, this study points to the need to test multimodal models through mixed-methods or quasi-experimental designs that include short- and long-term measurement, classroom observation, and documentation of students' multimodal products. Future studies should also examine how such models work across grade levels, urban and non-urban settings, and students with different learning needs. A future multimodal model for Indonesian language learning in elementary schools should therefore include not only activity syntax but also implementation tools such as teaching modules, assessment rubrics, classroom management guides, peer feedback strategies, and teacher reflection frameworks.

Research Limitations

This study has several limitations that should be considered when interpreting its scope and generalizability. First, the reviewed literature is dominated by non-Indonesian contexts, higher education settings, and a substantial number of computational studies. Therefore, transferring these findings to Indonesian language learning in elementary schools still requires conceptual adaptation and further empirical testing.

Second, many reviewed studies focus on a single skill, task type, or outcome area, such as vocabulary, reading, phonology, digital literacy, AI literacy, or learning emotions. As a result, this synthesis is based on a still-fragmented body of literature. Third, several studies important for discussing effectiveness and model design still have limitations in sample size, intervention duration, single-context implementation, or highly controlled designs. The available evidence, therefore, does not yet fully reflect authentic Indonesian-language classroom implementation in elementary schools.

Fourth, the literature still shows limitations in assessment, because many studies do not combine evidence on process, outcomes, retention, and implementation context. Fifth, because this study is a Systematic Literature Review, its contribution lies in synthesis and mapping rather than direct testing of a proposed model in real classrooms. Sixth, the heterogeneity of theories, methods, and outcome indicators means that generalisation must be made with care. This study should therefore be read as a conceptual and methodological foundation for model development rather than as a final judgment about a single intervention that is certainly the most effective.

Seventh, some conceptually useful studies, especially computational studies, offer insights into fusion, attention, coherence, and intermodal conflict but do not provide direct pedagogical guidance for elementary school classrooms. For this reason, the main limitation of this study also becomes the next research agenda: the need to test an Indonesian-language multimodal learning model directly in elementary schools using contextual designs, performance-evidence-based assessment, and detailed documentation of classroom processes. Even with these limitations, the study still provides a clear theoretical foundation and a focused direction for future empirical development.

CONCLUSION

This study demonstrates that multimodal learning models have strong potential to enhance elementary school students' language skills when learning is designed through the purposeful integration of text, visuals, audio, gestures, social interaction, and meaning-making activities. The review shows that multimodal learning is not merely the use of varied media, but a pedagogical design characterised by production tasks, intermodal interaction, gradual scaffolding, authentic contexts, and opportunities for reflection. Its influence tends to be positive across reading, writing, speaking, listening, vocabulary, and communicative competence, although its effectiveness depends on instructional design quality, teacher readiness, student characteristics, and classroom conditions. The study also reveals that the literature remains fragmented conceptually, methodologically, and in assessment practices, with relatively few models specifically integrated for Indonesian language learning in elementary schools.

This study contributes to the field by consolidating scattered findings into a more coherent framework and by showing that elementary students' language development is more meaningful when supported by multimodal and participatory learning experiences. It also highlights the need for assessment systems that capture not only test results, but also learning processes, multimodal artefacts, classroom interaction, and student engagement. For future research, more direct empirical studies are needed to test integrated multimodal models for listening, speaking, reading, and writing across diverse school contexts. Future studies should also develop stronger assessment rubrics and examine both short- and long-term effects, so that theoretically sound, practical, replicable, and contextually relevant multimodal Indonesian language-learning models can be developed.

REFERENCE

Alsubaie, M. A. (2022). Distance education and the social literacy of elementary school students during the Covid-19 pandemic. *Heliyon*, *8*(7), e09811. https://doi.org/10.1016/j.heliyon.2022.e09811

Ayetiran, E. F., & Özgöbek, Ö. (2024a). A Review of Deep Learning Techniques for Multimodal Fake News and Harmful Languages Detection. *IEEE Access*, *12*, 76133–76153. https://doi.org/10.1109/ACCESS.2024.3406258

Ayetiran, E. F., & Özgöbek, Ö. (2024b). An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate speech and offensive language detection. *Information Systems*, *123*, 102378. https://doi.org/10.1016/j.is.2024.102378

Bai, Y., & Lei, S. (2025). Cross-language dissemination of Chinese classical literature using multimodal deep learning and artificial intelligence. *Scientific Reports*, *15*(1), 21648. https://doi.org/10.1038/s41598-025-05921-1

Boaventura, D., Neves, A. T., Santos, J., Pereira, P. C., Luís, C., Monteiro, A., Cartaxana, A., Hawkins, S. J., Caldeira, M. F., & Ponces De Carvalho, A. (2021). Promoting Ocean Literacy in Elementary School Students Through Investigation Activities and Citizen Science. *Frontiers in Marine Science*, *8*, 675278. https://doi.org/10.3389/fmars.2021.675278

Carter, H., & Abbott, J. (2024). Literacy Teachers in the Making: A Look at Teacher Candidates' Experiences as they Tutor Elementary Students. *Literacy Research and Instruction*, *63*(1), 79–101. https://doi.org/10.1080/19388071.2023.2167676

Chung, K., Kim, S., Jang, Y., Choi, S., & Kim, H. (2024). Developing an AI literacy diagnostic tool for elementary school students. *Education and Information Technologies*, *30*, 1013–1044. https://doi.org/10.1007/s10639-024-13097-w

Condie, C., & Pomerantz, F. (2020). Elementary students' literacy opportunities in an age of accountability and standards: Implications for teacher educators. *Teaching and Teacher Education*, *92*, 103058. https://doi.org/10.1016/j.tate.2020.103058

Dahl-Leonard, K., Hall, C., & Peacott, D. (2024). A meta-analysis of technology-delivered literacy instruction for elementary students. *Educational Technology Research and Development*, *72*(3), 1507–1538. https://doi.org/10.1007/s11423-024-10354-0

Ding, A.-C. E., Glazewski, K., & Pawan, F. (2022). Language teachers and multimodal instructional reflections during video-based online learning tasks. *Technology, Pedagogy and Education*, *31*(3), 293–312. https://doi.org/10.1080/1475939X.2022.2030790

Engman, M. M. (2021). A worksheet, a whiteboard, a teacher-learner: Leveraging materials and colonial language frames for multimodal indigenous language learning. *Classroom Discourse*, *12*(1–2), 75–100. https://doi.org/10.1080/19463014.2020.1856696

Farías, M., & Véliz, L. (2016). Efl Students' Metaphorical Conceptualizations Of Language Learning. *Trabalhos Em Linguística Aplicada*, *55*(3), 833–850. https://doi.org/10.1590/010318135146185751

Garcia, M. (2026). Multilingual language learning in a multimodal metaverse: A multidimensional study of communicative, affective, and cognitive development. *Innovation in Language Learning and Teaching*, 1–27. https://doi.org/10.1080/17501229.2026.2621262

Gladys, A., & Vetriselvi, V. (2024). Sentiment analysis on a low-resource language dataset using multimodal representation learning and cross-lingual transfer learning. *Applied Soft Computing*, *157*, 111553. https://doi.org/10.1016/j.asoc.2024.111553

Goo, M., Myers, D., Maurer, A. L., & Serwetz, R. (2020). Effects of Using an iPad to Teach Early Literacy Skills to Elementary Students With Intellectual Disability. *Intellectual and Developmental Disabilities*, *58*(1), 34–48. https://doi.org/10.1352/1934-9556-58.1.34

Hadad, S., Watted, A., & Blau, I. (2023). Cultural background in digital literacy of elementary and middle school students: Self-appraisal versus actual performance. *Journal of Computer Assisted Learning*, *39*(5), 1591–1606. https://doi.org/10.1111/jcal.12820

Hagerman, M. S., Cotnam-Kappel, M., Turner, J.-A., & Hughes, J. M. (2022). Literacies in the Making: Exploring elementary students' digital-physical meaning-making practices while crafting musical instruments from recycled materials. *Technology, Pedagogy and Education*, *31*(1), 63–84. https://doi.org/10.1080/1475939X.2021.1997794

Heo, Y., Kang, S., & Seo, J. (2023). Natural-Language-Driven Multimodal Representation Learning for Audio-Visual Scene-Aware Dialog System. *Sensors*, *23*(18), 7875. https://doi.org/10.3390/s23187875

Hermes, M. R., Engman, M. M., Meixi, & McKenzie, J. (2023). Relationality and Ojibwemowin in Forest Walks: Learning from Multimodal Interaction about Land and Language. *Cognition and Instruction*, *41*(1), 1–31. https://doi.org/10.1080/07370008.2022.2059482

Hong, J., & Kim, K. (2025). Impact of AIoT education program on digital and AI literacy of elementary school students. *Education and Information Technologies*, *30*(1), 107–130. https://doi.org/10.1007/s10639-024-12758-0

Huang, Y., Xu, W., Sukjairungwattana, P., & Yu, Z. (2024). Learners' continuance intention in multimodal language learning education: An innovative multiple linear regression model. *Heliyon*, *10*(6), e28104. https://doi.org/10.1016/j.heliyon.2024.e28104

Jensen, M. T., Solheim, O. J., & Olsen, E. (2025). Leader support in relation to teacher self-efficacy, classroom emotional climate and students' literacy skills in elementary school. *Scandinavian Journal of Educational Research*, *69*(4), 729–742. https://doi.org/10.1080/00313831.2024.2348451

Li, L., Bai, X., Xu, J., Wang, D., & Jiang, T. (2025). Multimodal learning audio-visual detection for obtaining object-level sound sources in Japanese-language teaching room. *Scientific Reports*, *15*(1), 16632. https://doi.org/10.1038/s41598-025-00588-0

Li, W., Yu, J., Zhang, Z., & Liu, X. (2022). Dual Coding or Cognitive Load? Exploring the Effect of Multimodal Input on English as a Foreign Language Learners' Vocabulary Learning. *Frontiers in Psychology*, *13*, 834706. https://doi.org/10.3389/fpsyg.2022.834706

Lin, J., Zhang, H., & Lin, X. (2022). Prosodic Transfer in English Literacy Skills among Chinese Elementary-Age Students: Controlling for Non-Verbal Intelligence. *Journal of Intelligence*, *10*(4), 114. https://doi.org/10.3390/jintelligence10040114

Maijala, M. (2023). Multimodal postcards to future selves: Exploring pre-service language teachers' process of transformative learning during one-year teacher education programme. *Innovation in Language Learning and Teaching*, *17*(1), 72–87. https://doi.org/10.1080/17501229.2021.1919683

Malone, J., Hui, B., Pandža, N., & Tytko, T. (2025). Eye Movements, Item Modality, and Multimodal Second Language Vocabulary Learning: Processing and Outcomes. *Language Learning*, lang.70007. https://doi.org/10.1111/lang.70007

Melo-Pfeifer, S., & Chik, A. (2022). Multimodal linguistic biographies of prospective foreign language teachers in Germany reconstructing beliefs about languages and multilingual language learning in initial teacher education. *International Journal of Multilingualism*, *19*(4), 499–522. https://doi.org/10.1080/14790718.2020.1753748

Meneses, A., Uccelli, P., & Valeri, L. (2023). Teacher Talk and Literacy Gains in Chilean Elementary Students: Teacher Participation, Lexical Diversity, and Instructional Non-present Talk. *Linguistics and Education*, *73*, 101145. https://doi.org/10.1016/j.linged.2022.101145

Nguyen-Thi, M.-H., Tran, K.-X., & Giang, T.-V. (2025). Exploring the emotional experience in learning Chinese as a second language of students from the multimodal affective perspective: A case study in Vietnam. *Acta Psychologica*, *260*, 105575. https://doi.org/10.1016/j.actpsy.2025.105575

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. *ITL - International Journal of Applied Linguistics*, *173*(1), 2–17. https://doi.org/10.1075/itl.21039.pel

Rahmanu, I. W. E. D., & Molnár, G. (2024). Multimodal immersion in English language learning in higher education: A systematic review. *Heliyon*, *10*(19), e38357. https://doi.org/10.1016/j.heliyon.2024.e38357

Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of Multimodal Glossing on Second Language Vocabulary Learning: A Meta-analysis. *TESOL Quarterly*, *55*(1), 105–133. https://doi.org/10.1002/tesq.579

Scott, J.-A. (2020). (Re)directing a university storytelling troupe for at-risk elementary students for course credit: A story of embodied empathy, literacy, and personal transformation. *Text and Performance Quarterly*, *40*(2), 170–186. https://doi.org/10.1080/10462937.2019.1691742

Sheng, H., Shen, X., Du, H., & Yu, X. (2026). Mobile Auslan: A multimodal dialogue-centered sign language learning system. *Computer Vision and Image Understanding*, *265*, 104646. https://doi.org/10.1016/j.cviu.2026.104646

Tong, P., & An, I. S. (2026). Synaesthesia in digital multimodal composing: The case of a mobile-assisted task for learning Chinese as a foreign language. *Computer Assisted Language Learning*, 1–40. https://doi.org/10.1080/09588221.2025.2605538

Umino, T. (2023). Using multimodal language learning histories to understand learning experiences and beliefs of second language learners in Japan. *The Modern Language Journal*, *107*(1), 308–327. https://doi.org/10.1111/modl.12828

Wahyudi, L. (2024). *Watase Uake: Research Collaboration Tools*. Retrieved from https://www.watase.web.id. https://www.watase.web.id

Yang, Y., Yang, Y.-Q., Ren, G., & Yu, B.-G. (2025). Hierarchically trusted evidential fusion method with consistency learning for multimodal language understanding. *Knowledge-Based Systems*, *312*, 113164. https://doi.org/10.1016/j.knosys.2025.113164

Yoon, J., Choi, G., & Choi, C. (2023). Multimedia analysis of robustly optimized multimodal transformer based on vision and language co-learning. *Information Fusion*, *100*, 101922. https://doi.org/10.1016/j.inffus.2023.101922

Yue, M., Jong, M. S.-Y., Dai, Y., & Lau, W. W.-F. (2025). Students as AI literate designers: A pedagogical framework for learning and teaching AI literacy in elementary education. *Journal of Research on Technology in Education*, 1–22. https://doi.org/10.1080/15391523.2025.2449942

Zhou, M., Steinberg, S., Stiso, C., Danish, J. A., & Craig, K. (2024). Using network visualizations to engage elementary students in locally relevant data literacy. *Information and Learning Sciences*, *125*(3/4), 209–231. https://doi.org/10.1108/ILS-06-2023-0069

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to all individuals and institutions that contributed to the completion of this study. We are particularly grateful to the affiliated institution for its academic and administrative support throughout the research process. We also acknowledge the valuable insights, encouragement, and constructive assistance provided by colleagues and other contributors whose support helped improve the quality of this manuscript. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTION STATEMENT

BF conceived the study, designed the review protocol, conducted the literature search, extracted and analysed the data, and drafted the manuscript. DS and BM supervised the study, contributed to the interpretation of the findings, and reviewed the manuscript for important intellectual content. AR and Y help translate and improve the language used in articles and review the discussion sections. BF, DS, BM, Y and AR approved the final version of the manuscript and agree to be accountable for all aspects of the work.

AI DISCLOSURE STATEMENT

The author used Scite AI during the preparation of this work for preliminary literature searches, to obtain a general overview of the research topic, and to identify initial reference-related information relevant to the study. The author also used Wataseuake to identify and analyse the literature in accordance with the PRISMA framework, using international scientific databases, thereby supporting the mapping of research developments in the selected topic area. In addition, Grammarly was used for grammar checking, sentence structure editing, and improving the quality of academic English writing to ensure clarity and linguistic accuracy in the manuscript. After using these tools and services, the author thoroughly reviewed, revised, and edited the content as needed and takes full responsibility for the accuracy, originality, and integrity of the entire article.

***Budi Febriyanto (Corresponding Author)**
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Kota Bandung, Jawa Barat., Indonesia
Email: budifebriyanto@upi.edu

Dadang Sunendar
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Kota Bandung, Jawa Barat., Indonesia
Email: dadangsunendar@upi.edu

Bachrudin Musthafa
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Kota Bandung, Jawa Barat., Indonesia
Email: bachrudin.mustafa@upi.edu

Yuliawati[4]
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi No. 229, Kota Bandung, Jawa Barat., Indonesia
Email: yuliagunawan20@upi.edu

Agus Rofi'i
Universitas Majalengka,
Jl. K.H. Abdul Halim No. 103, Majalengka Kulon, Kabupaten Majalengka, Jawa Barat 45418, Indonesia
Email: agusrafii@unma.ac.id