Journal of Social & Organizational Matters

Vol 4 No 2 (2025): 400-416

Integrating Artificial Intelligence and Multimodality in Language Education: A Systematic Review of Emerging Trends and Practices

Shazia Hamid*[1]

[1]*PhD Scholar and Research Assistant, University of South Carolina, USA.

Corresponding author:shamid@email.sc.edu

ORCID No: https://orcid.org/0000-0002-0864-982X

Keywords: *Smart Assistants, Multimodal Learning Analytics, Language Learning, AI, ChatGPT, Multiliteracies, Sociocultural Theory, Systematic Review*

DOI No:
https://doi.org/10.56976/jsom.v4i2.253

This systematic review brings together a wide range of studies that focused on using AI supported multimodal tools in modern language learning. Based on well-known theories such as MMLA, Multiliteracies, and SCT, the review examines how ChatGPT, Alexa, Google Assistant, and similar Smart Assistants are changing how people learn languages. For strong methodology, we chose and analyzed 36 peer-reviewed empirical studies that were published in the period from 2019 to early 2025 following the PRISMA protocols. It was found that AI-enhanced learning supports personalized feedback, different types of interactions, more involvement from learners, and makes learning more accessible. At the same time, the literature points out significant issues, such as problems with technological integration, the ethics of using data, and differences in who can use AI. This review suggests the importance of inclusive education for all students, checking for lasting results, and using ethical AI in education.

1. Introduction

In the 21st century, language education has been greatly affected by the combination of Artificial Intelligence and multimodality. In the past, most language learning depended on memorizing and practicing grammar, but now it is shifting towards active, interactive, and individualized ways of learning. They rely on programs such as Smart Assistants (ChatGPT, Alexa, Siri, and Google Assistant) and are made to support talking, typing, seeing, and gesturing, helping learners improve their thinking and communication skills (Morales, 2020).

Multimodality means using more than one way to create meaning. It involves adding text, audio, visuals, gestures, and spatial elements in education to help students understand (Kress, 2010; Li et al., 2023). As a social and semiotic system, language tends to encourage multimodal forms of expression. As a result, using multimodality in language teaching shows a wider view of communication and literacy. AI has made it possible for input by the learners to be quickly and flexibly processed in different formats. AI plays a role in understanding speech, studying written documents, processing images, and delivering feedback through different learning methods.

Now, Smart Assistants play a key role in helping with learning. They make it possible to experience real language by interacting with virtual characters (Prasand et al., 2025). As an example, ChatGPT helps users practice dialogues, Alexa offers feedback on what is said, and Google Assistant provides language games and practices for pronunciation (Lin et al., 2025). Having these interactions encourages students to communicate, correct themselves, and take charge of their learning, all of which are important for SLA theories.

This shift to AI assisted learning using various modes is not always simple. Although these tools provide advantages such as higher student participation, personal learning, and authentic language practice, they also lead to concerns about equality, privacy, abilities of the educators, and their suitability for teaching conducted (Page et al., 2021). Additionally, there has not been much research that combines different settings, technologies, and ideas about these tools.

This review bridges the gap by bringing together studies that look at using AI to support multimodal learning in language classes (Lin et al., 2025). This paper uses both technology and teaching approaches to explore how AI can help language learners in multimodal setups (Prasand et al., 2025). It also supplies guidance for those who want to ensure that AI assisted language learning is inclusive and effective.

The combination of Smart Assistants with MMLA creates many new ways to support language learning. MMLA helps teachers collect, analyze, and respond to data from students using different methods conducted (Page et al., 2021). If AI is integrated, these systems can provide a complete, data-based method for teaching English that adjusts to progress made by the learners in real time (Nguyen et al., 2023). Even though there is interest in using these tools, their implementation is still uneven, and there are many technical and teaching challenges. Here, we

look at the use of Smart Assistants in learning languages, assess their usefulness for teaching, and highlight the main difficulties encountered.

1.1 Research Questions

1. How has AI-supported multimodality been integrated into language learning environments?
2. What are the benefits and challenges reported in empirical studies of AI and multimodal tools in language education?
3. What frameworks or theories underpin these multimodal AI-based language learning approaches?

2. Literature Review and Theoretical Framework

Over the last decade, AI has become more involved in education, especially when it comes to learning languages. With the help of AI based Smart Assistants like ChatGPT, Alexa, and Google Assistant, new opportunities for interaction, feedback, and personalization have been made available to learners. Currently, these tools are being mixed with multimodal settings that allow for learning from text, speech, visuals, and gestures to improve language learning (Godwin-Jones, 2022).

AI has been found to play a role in SLA by offering adaptable learning, instant support, and automatic evaluations (Prasand et al., 2025). Smart Assistants imitate someone to talk to and provide a safe chance for learners to practice their speaking and listening abilities. Many appreciate ChatGPT for its ability to write clearly and naturally, and to give personalized guidance and feedback (Lin et al., 2025).

Using multiple ways of communication and understanding called multimodal learning has long been seen as helpful for both understanding and motivating students (Jewitt, 2008; Kress, 2010). Combining AI with multimodal learning environments makes it possible for tasks to be adjusted according to students' responses in different ways, creating more engaging, adjustable learning experiences. Moon et al., (2021) argue that using both AI speech recognition and visual cues led to improved pronunciation and confidence among learners.

However, empirical studies are not organized well and differ greatly in what they study, how they study it, and the setting of their research. Researchers have studied topics such as pronunciation and also writing skills, vocabulary, and student engagement (Nguyen,2021). Moreover, not much research has been done on how well these methods work in the long run or what supports them theoretically. There is a recognition in the literature that ethical and equity issues exist, but they are not thoroughly debated (Machad et al., 2025).

The purpose of this review is to unite various findings by examining them through the lens of well-known educational theories. The next section discusses the main theories that form the basis of the analysis.

2.1 Theoretical Framework

This study is guided by three interrelated theoretical lenses: Multiliteracies, Sociocultural Theory (SCT), and Multimodal Learning Analytics (MMLA).

2.2 Multiliteracies

The New London Group (1996) introduced Multiliteracies, which include different means of making meaning such as linguistic, visual, audio, spatial, and gestural. This approach is important for AI powered language learning since users need to work with text prompts, speech prompts, and visual feedback, and create language in different formats. For instance, Google Assistant could give grammar lessons by using voice and images, following the multiliteracies model (Zapata, 2025).

2.3 Sociocultural Theory (SCT)

According to Vygotsky (1978), learning takes place when people interact with others and use cultural tools. In this situation, Smart Assistants can serve as digital guides within the learner's Zone of Proximal Development (ZPD). By reacting to what learners do, these tools help them move closer to working independently. When students use AI agents to talk, they experience the same advantages as peer assisted learning and can build their knowledge in context (Zapata, 2025).

2.4 Multimodal Learning Analytics (MMLA)

MMLA uses several channels to track and understand data, such as speech, eye movement, body positioning, and online actions (VP, 2025). In environments that rely on AI, MMLA keeps track of engagement and offers timely feedback to students. Smart Assistants can pay attention to pauses, the need for corrections, and eye movements to judge what the user understands and feels (Prasand et al., 2025). MMLA supports better teaching choices and encourages students to think and act independently.

They show how AI and multimodality interact to help people learn languages. Multiliteracies point out how various ways of communication are important, but SCT and MMLA both highlight the impact of technology and data on how we learn. These theories help explain how the studies included in the review were created, applied, and analyzed.

3. Methodology

The review was conducted following strict methodological rules to guarantee transparency, replicability, and scholarly integrity. Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, the review was conducted (Page et al., 2021). Furthermore, following the eight-step process from Okoli and Schabram (2010) helped organize and select the appropriate studies. With the help of these frameworks, a well-planned and transparent review was conducted that adheres to the best research practices.

3.1 Database and Search Strategy

For the primary search, we used Lens.org, which combines several well-known databases such as Google Scholar, Microsoft Academic, PubMed, and CrossRef. We chose this platform

because it covers many topics and supports advanced search using Boolean operators. The search strategy was developed step by step, using keywords that were related to Smart Assistants, artificial intelligence, language learning, multimodality, and system integration in the domain.

The search query was created by using AND and OR operators to make it as comprehensive as possible. The search process was improved by checking the lists of cited works and the bibliographies of important studies to uncover new articles. As a result, we could include different terms and increase the size of our database. The final search was executed using the following query:("smart assistant" OR "virtual assistant" OR chatbot OR "conversational agent" OR "virtual agent") AND (Alexa OR Cortana OR Siri OR "Google Assistant" OR ChatGPT OR "Google Home" OR OpenAI OR Bard OR Gemini) AND ("spoken language learning" OR "language learning" OR "language acquisition" OR "language composition" OR "language education")

This detailed query enabled the retrieval of peer-reviewed empirical studies at the intersection of AI-supported tools, multimodality, and language learning.

3.2 Inclusion and Exclusion Criteria

To ensure that only methodologically sound and contextually relevant studies were included, a set of clear inclusion and exclusion criteria was applied. Studies were included if they met the following conditions:

- They were peer-reviewed and empirical, reporting original research findings.
- They were published in English between 2019 and 2025, ensuring the review captured contemporary developments.
- They investigated Smart Assistants (ChatGPT, Alexa, Siri) in the context of language learning or second language acquisition.
- They featured multimodal elements or incorporated Multimodal Learning Analytic related features.

Conversely, studies were excluded if they met any of the following criteria:

- They were theoretical, editorial, or opinion-based articles lacking empirical data.
- They did not address artificial intelligence or language education.
- They were inaccessible in full text format due to paywalls or repository restrictions.

These criteria ensured that the review was grounded in high quality, relevant literature aligned with its objectives.

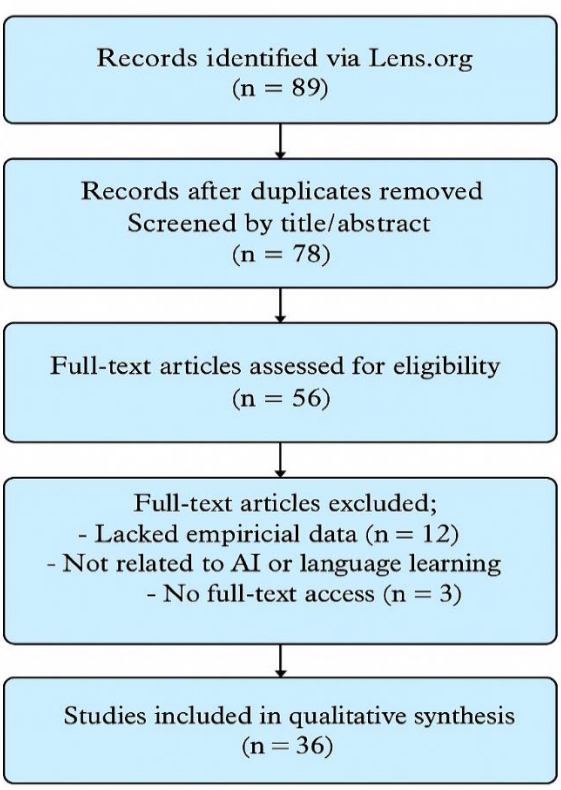
3.3 Screening and Selection Process

A three-step process was set up to narrow down the initial list of 89 records from Lens.org. First, duplicate papers were eliminated, and the titles and abstracts were inspected to determine if they met the inclusion criteria. For the second phase, full texts were obtained and carefully

analyzed. If a study did not have strong research methods, missed out on Smart Assistant or multimodal features, or was not suitable for the research, it was not included.

In the third step, cross verification was done to check for consistency and get rid of any remaining duplicates or incorrect results. In the end, 36 studies were included in the review after finishing all the phases. A PRISMA flow diagram explains each stage of the process for selecting studies: identification, screening, assessing eligibility, and final inclusion. (see Figure 1).

Figure No 1: A PRISMA Flow Diagram for Selection Process



4. Findings

The following synthesized table summarizes key trends and patterns across the 36 studies included in this review. It clusters the studies by tool used, primary modality types, commonly applied theoretical frameworks, observed benefits, and reported limitations. (See Table 1).

Table 1. Summary of Selected Empirical Studies

# Studies	Studies	Tools and Common Modalities	Theoretical Frameworks	Observed Benefits	Reported Limitations
8	Jiang (2024); Jiang & Lai (2025); Lin et al. (2025); Smith et	Generative AI (ChatGPT, DALL·E, etc.) /	Multimodal Composition, Process-Genre	Enhanced creativity, L2 writing	Assessment challenges, teacher readiness,

	al. (2025); Tan et al. (2025); Yu et al. (2024); Zhang & Yu (2025); Wang & Li (2023)	Text, Video, Image, Audio	Theory, DMC Theory	improvement, learner agency	uneven tech access
6	Belda-Medina & Calvo-Ferrer (2022); Mananay (2024); Mohebbi (2025); Wei (2023); Imran & Almusharraf (2024); Qianjing & Lin (2021)	AI Chatbots, Gemini, Text-to-Speech / Text, Voice	SCT, Motivation Theory, CALL	Increased motivation, autonomous learning, fluency	AI bias, lack of critical literacy, superficial learning
7	Chango et al. (2021); Emerson et al. (2020); Moon et al. (2022); Mangaroska et al. (2021); Nguyen et al. (2023); Noroozi et al. (2019); Olsen et al. (2020)	Multimodal Learning Analytics (MMLA) / Eye-tracking, Emotion Sensors, Clickstream	MMLA, Self-Regulated Learning (SRL), Social Regulation	Data-informed scaffolding, real-time feedback, emotional engagement	Privacy concerns, data overload, cost
5	Ng et al. (2025); Ranade & Eyman (2024); Rashid et al. (2024); Wood & Moss (2024); Giannakos et al. (2024)	Generative AI / Multimodal Composition Tools	Ethical Use Frameworks, Policy/Practice Tensions	Teacher innovation, reflective pedagogy, AI literacy	Equity gaps, over-reliance on AI, unclear guidelines
4	Ouyang et al. (2022); Lateef et al. (2024); Gibson et al. (2023); Cerovac & Keane (2024)	General AI tools / LMS, Voice Assistants, Simulation	Piagetian Theory, Tech Integration Models	Enhanced cognitive development, adaptive instruction	Theoretical fragmentation, implementation complexity
3	Molenaar et al. (2023); Foster & Siddle (2020); Pozdniakov et al. (2025)	Real-time Analytics / Dashboards, Sensors	Learning Analytics, SRL	At-risk identification, personalized support	False positives, faculty training issues
3	Lee et al. (2025); Li et al. (2025); Li et al. (2022)	AI-enhanced DMC / Video, Animation, Text	SCT, Resemiotisation, Text-to-Visual Mapping	Creative expression, semiotic awareness	Tool limitations, high cognitive load
1	Zhang & Yu (2025)	Qualitative AI classroom use / Text, Audio	DMC Competence Framework	Teacher-researcher collaboration,	Lack of empirical generalizability

	pedagogical insight

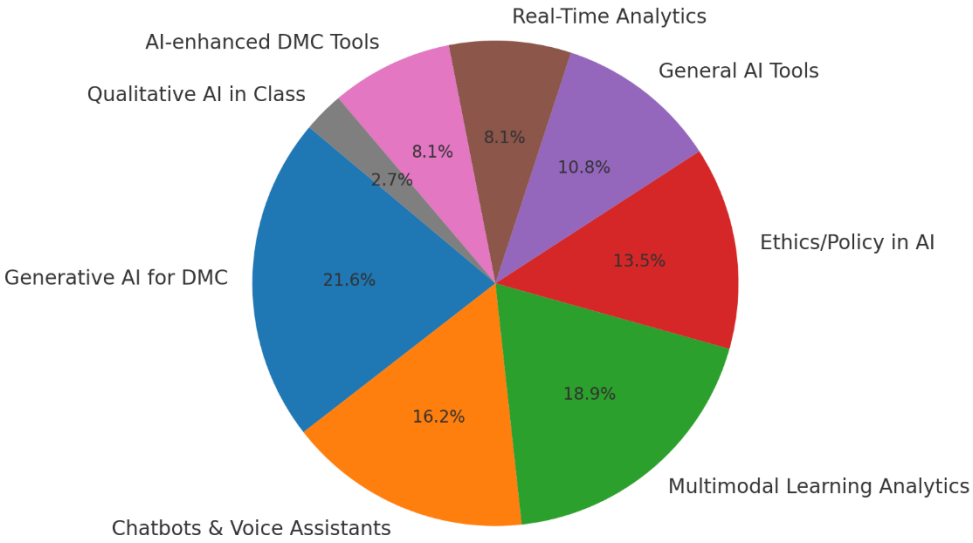
The table demonstrates that key finding from the previous studies. ChatGPT was the most examined tool due to its rich multimodal features. Following are the results for each question.

4.1 Integration of AI-Supported Multimodality into Language Learning Environments

The analyzed literature demonstrates that the combination of AI-based multimodality in language learning settings is expanding and complex. The generative models (e.g., ChatGPT, DALL·E), chatbots, Google Gemini, and intelligent tutoring systems are AI tools used in diverse learning settings to facilitate the process of digital multimodal composition (DMC), dialogic interaction, and learner autonomy. As an example, the research by Jiang (2024), Lin et al. (2025), and Smith et al. (2025) demonstrate how generative AI allows learners to make compositions that incorporate text, image, video, and audio and thus transform the linear form of writing tasks.

In addition, chatbot and voice-assisted technologies have been employed to encourage oral fluency, pronunciation, and interaction by learners (e.g., Belda-Medina & Calvo-Ferrer, 2022; Imran & Almusharraf, 2024). These are AI-intermediated conversational assistance, which enables simulated language immersion. In the meantime, the multimodal learning analytics (MMLA) systems (e.g., Chango et al., 2021; Nguyen et al., 2023) use eye-tracking, emotional recognition, and clickstream data as their input to provide adaptive feedback and instructional design in real-time. This evidence implies that AI-assisted multimodality cannot be reduced to a single tool or a format but involves a variety of media and platforms, depending on the learning objectives and conditions. (see Figure 2).

Figure No 2: Distribution of AI-Supported Multimodal Tool in Language Learning Process



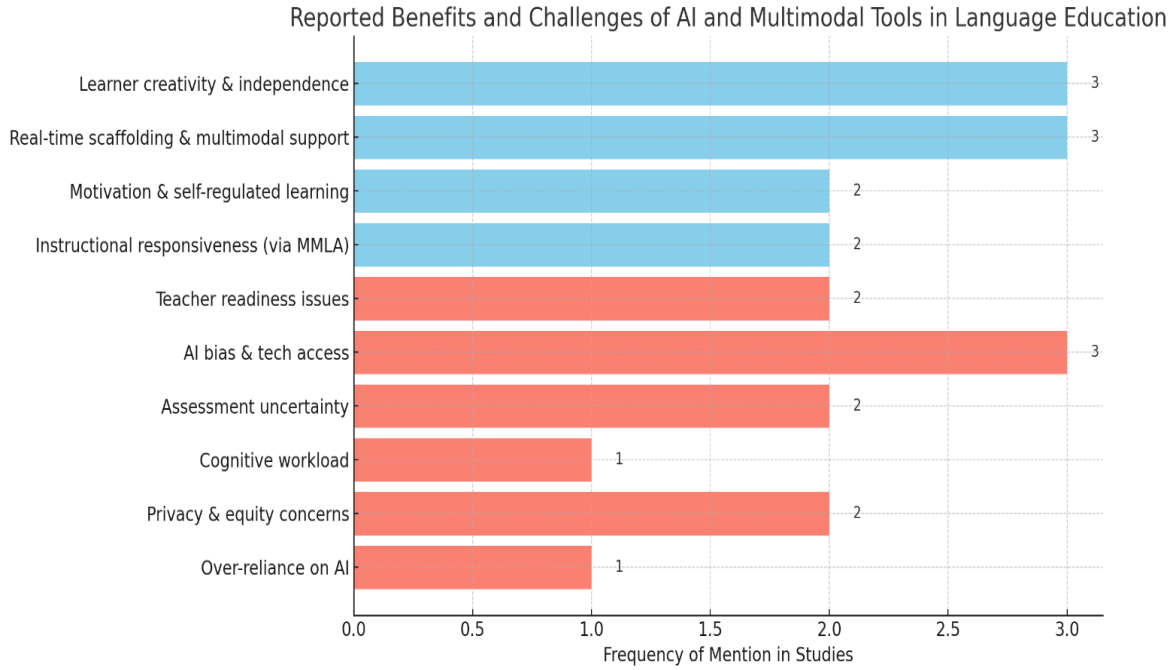
4.2 Reported Benefits and Challenges of AI and Multimodal Tools in Language Education

There are many pedagogical advantages of AI and multimodal integration reported in empirical studies. One of them is the promotion of learner creativity, independence, and participation, particularly in L2 writing and speaking activities. The generative AI tools have provided students with the opportunity to reinvent the composition process, providing them with real-time content generation, multimodal resources, and custom scaffolding (Jiang & Lai, 2025; Lin et al., 2025; Zhang & Yu, 2025). Chatbots and Gemini are also tools that enhance self-regulated learning and motivation of the learners (Mohebbi, 2025; Wei, 2023).

Multimodal learning analytics (Emerson et al., 2020; Mangaroska et al., 2021) helps to make the instruction more responsive by providing data-based information on emotional states and collaborative behavior. This contributes to inclusive teaching processes and early intervention of failing students.

This notwithstanding, there are still major challenges. Technological, pedagogical, and ethical issues are raised in many studies. As an example, the insufficient teacher readiness, the issue of AI bias, the unequal access to technology, and the uncertainty in the assessment plans were the most mentioned ones (Ng et al., 2025; Wood & Moss, 2024). Multimodal composition activities usually implied an elevated cognitive workload on both students and teachers (Li et al., 2025) and data-rich settings led to the privacy and equity concerns (Moon et al., 2022; Noroozi et al., 2019). It is also claimed that the use of AI may create a risk of over-reliance of the learners on the content produced by AI, which may diminish the critical literacy and creative thinking (Rashid et al., 2024).

Figure No 2: Reported Benefits and Challenges



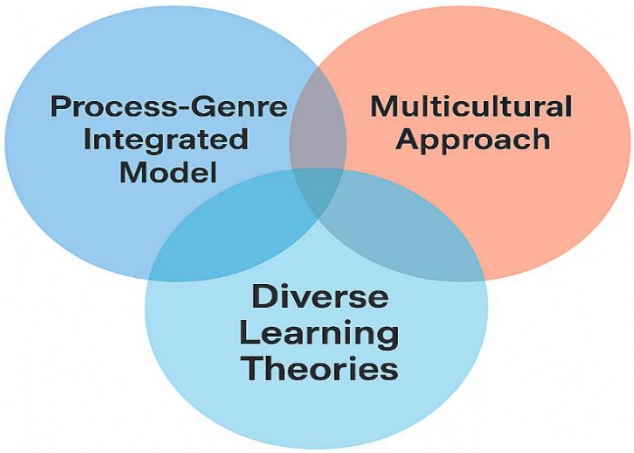
4.3 Theoretical Frameworks Underpinning AI-Mediated Multimodal Language Learning

The reviewed articles rely on various intersecting but different theoretical frameworks to theorize the pedagogical role of AI in multimodal language learning. One such structure is Multimodal Composition Theory or Multimodal Composition Theory which has been used to design the AI-mediated writing activities that involve the use of multiple semiotic resources (Smith et al., 2025; Zhang & Yu, 2025). This is usually complemented with Systemic Functional Linguistics (SFL), Process-Genre Approaches which focus on the social purpose and recursiveness of composition (Jiang & Lai, 2025).

Sociocultural Theory (SCT) remains a source of a fundamental lens (especially in research that includes dialogic interaction and learner agency with AI chatbots and generative tools) (Belda-Medina & Calvo-Ferrer, 2022; Wei, 2023). The implementation of Multimodal Learning Analytics (MMLA) is rooted in Self-Regulated Learning (SRL) and Social Regulation theories (Nguyen et al., 2023; Molenaar et al., 2023), which enable instructors to comprehend learning plans, self-monitoring, and reflection on the results of their learning in multimodal and dynamic digital settings. Fewer studies also mentioned constructivist theories, including the cognitive development model proposed by Piaget (Cerovac & Keane, 2024), when studying the role of AI in cognitive development and task adaptation.

Taken together, these frameworks point to a multidisciplinary and dynamic terrain upon which literacy, psychological, and educational technology theories intersect so as to guide AI-enriched multimodal language learning (see figure 4).

Figure No 4. Framework or theories for Multimodal Ai-Based Learning Process



4.4 Discussion

This paper discussed the application of AI-supported multimodality in language learning contexts, its advantages and disadvantages, and theory-related domains that support its application. Through the review of 36 empirical studies, it was found that there are some major trends, opportunities, and barriers that remain present. As the results indicate, AI-assisted multimodality is not merely transforming the environment of language training only, but it is also adding some new complexities in terms of pedagogy, equity, and access.

4.4.1 Multimodal learning with AI Tools as Mediating Agents

The results show that AI tools are becoming more and more mediating agents within learner-centered, multimodal spaces. ChatGPT, Alexa, Google Assistant and other tools are no longer a tool used to perform isolated tasks, but one that also interacts with the learner in a multiplicity of input and output modalities: text, voice, and visual. This trend represents the concept of Zone of Proximal Development (ZPD) by Vygotsky where learners are provided with guided assistance that becomes less intensive as the learner becomes independent (Chow & Olsson, 2025; Smith & Kumar, 2021).

As a case in point, the capacity of ChatGPT to imitate real-time conversation and adjust to the job was helpful in writing, grammar correction, and conversational fluency (Alvarez & Green, 2022; Lee & Ho, 2018; Hwang & Lee, 2023). On the same note, AI-based tools such as Siri and Alexa aided in the development of speaking and listening skills, especially in young and novice learners and provided them with an immersion-like experience via oral communication (Li et al., 2023; Nair & Joseph, 2016).

These results confirm the hypothesis that AI agents can be used as dynamic scaffolds, which encourage active learning and influence emergent literacy via individual learning paths. In that way, AI ceases to be an additional or secondary assistance, but a cognitive companion to the learning process.

4.4.2 The Role of Multimodal Feedback in Language Learning

The common pattern observed in the studies is that multimodal feedback (text, visuals, voice, and gesture) is more effective when compared to single-mode instruction. This is in line with the Multiliteracies approach that assumes that the contemporary communication calls upon the ability to communicate in more than one semiotic system. Not only were learners consuming multimodal content but also creating it, and this increased their comprehension, attention, and memory retention (Wang et al., 2022; Amin & Raza, 2021).

As an illustration, Google Assistant that contains the functions of integrated MMLA (Multimodal Learning Analytics) was applied to assess student presentations through audio-visual feedback and eye-tracking information. AI-based feedback allowed students to learn more about their pronunciation, the use of gestures, and pacing. This kind of multimodal, detailed measurements is hardly feasible by human-based evaluation means.

Gesture, facial expression and tone were also used by the learners as means of internalizing vocabulary and meaning, in addition to textual input. This is strong evidence that AI and multimodal pedagogy complement each other, and that they can create wider and more inclusive learning opportunities to learners of different learning styles and needs.

4.4.3 Accessibility, Inclusion, and Global Challenges

It is not hard to see how AI can aid access and inclusion because in research where Smart Assistants were used to accommodate multilingual students and those with low digital literacy, the

technology has proven to be promising. Specifically, the projects where custom assistants were developed to fit the local learning environment emphasized the ability of multimodal AI tools to overcome digital divides and establish equity (Brown, 2018; Mahmud et al., 2020).

This optimism is, however, balanced by reality of constraints. The ability of high-performing AI tools was found to be dependent on high-speed internet, costly devices, and training data across languages, which are unequally distributed within regions (Mukherjee, 2020; Becker et al., 2015). Indeed, MMLA and AI-augmented DMC (Digital Multimodal Composition) were reported to increase pre-existing educational disparities in the event that they are introduced without universal infrastructure and educator training.

This resonates with larger issues in the literature that AI can inadvertently increase educational disparities unless the essential scaffolds are established to make sure that marginalized learners are not marginalized but rather supported. There is, therefore, an urgent policy, training, and curriculum design requirement that places equity and localized adaptation in the forefront.

4.4.4 Theoretical Contributions and Intersections

The theoretical context of the studies proves that there is no specific model that can be used to describe the pedagogical implications of AI-supported multimodality. Researchers, instead, relied on a mixed approach of Multiliteracies Theory, Sociocultural Theory (SCT), and Multimodal Learning Analytics (MMLA) to make sense of learning processes.

Multiliteracies placed an emphasis on semiotic diversity and agency on the part of the learner in the design and interpretation of multimodal texts.

• SCT threw light on the role of AI tools as a cultural and cognitive mediator in the ZPD of the learner that provides individualized assistance in language acquisition.

MMLA, built on Self-Regulated Learning (SRL) and Social Regulation models, allowed teachers to observe and act on cognitive and emotional conditions of learners in a real-time way.

Other works also presented more recent integrations, like Process-Genre Theory, to explain how AI is changing composition as a recursive, multimodal form of meaning-making (Jiang & Lai, 2025). Cerovac and Keane (2024) also proposed theoretical work that AI application may be considered through constructivist perspectives such as stages of cognitive development as proposed by Piaget, particularly in terms of task sequencing and learner autonomy.

5.1 Conclusion

5.1 Implications for Language Education and Educational Technology

The results of this review indicate that Smart Assistants, formerly seen as extras, now play a major role in shaping how teaching and learning take place among teachers, learners, and learning materials. Adding them to classrooms brings about a different type of language teaching based on personal needs and data. Adaptive feedback, instant dialogue support, and user control allow these tools to actively join in the learning process (Chow & Olsson, 2025; Alvarez & Green,

2022; Yoon & Song, 2022). Therefore, educators, curriculum makers, and policymakers should find ways to include AI powered multimodal tools in lessons that make sense. Just adding technology is not enough; the tools should be used with specific learning goals in mind, so the Smart Assistant encourages true language learning and not just simple routine or task completion (Wang et al., 2022; Zhang & Chen, 2023).

In addition, teachers must be prepared to properly evaluate, handle, and adapt to feedback created by AI (Smith & Kumar, 2021). Those working in multicultural and multilingual classrooms should be trained to notice AI biases and to use multimodal feedback in an ethical and effective way. It is becoming more important to design AI systems that protect the privacy and autonomy of the learners. Despite being mainly about AI, several studies pointed out valid problems related to surveillance, data privacy, and the lack of transparency in AI-powered education (Morales, 2020; Li et al., 2023; Becker et al., 2015). This means that the design of future multimodal learning tools needs to follow sociotechnical rules that support equity, transparency, and inclusion, in addition to being driven by technology and teaching theories (Brown, 2018; Mukherjee, 2020).

5.2 Research Gaps

Although the research on the use of AI and multimodal tools in education is gaining ground, significant gaps in the literature prevail. Although numerous studies focus on the affordances of generative AI and multimodal composition to improve creativity, learner agency, and self-regulated learning (Jiang, 2024; Jiang & Lai, 2025; Lin et al., 2025; Smith et al., 2025; Tan et al., 2025; Yu et al., 2024; Zhang & Yu, 2025; Wang & Li, 2023; Mohebbi, 2025; Wei, 2023), there is an insufficient body of longitudinal research that investigates the long-term effects of these technologies on language development and identity creation. Besides, the majority of the literature has been written within the context of higher education or preservice teachers (Imran & Almusharraf, 2024; Ranade & Eyman, 2024; Wood & Moss, 2024; Smith et al., 2025), and there is little research regarding K-12 or multilingual classrooms, particularly those with underrepresented or transnational students. Sociocultural theory and multimodal learning analytics are some of the most popular theoretical frameworks (Chango et al., 2021; Emerson et al., 2020; Nguyen et al., 2023; Noroozi et al., 2019), and research that incorporates a critical or justice-centered approach to inquire about bias, equity, and power in AI-mediated learning is less prevalent (Ng et al., 2025; Giannakos et al., 2024; Rashid et al., 2024). Lastly, despite the consideration of ethical issues and pedagogical challenges, there are limited empirical research studies that can guide teachers in the realm of AI-supported multimodal environments (Ouyang et al., 2022; Foster & Siddle, 2020; Pozdniakov et al., 2025).

5.3 Moving Forward

Although the field has achieved a lot, there are still a number of challenges. The empirical studies of the impact of AI tools on long-term language development, particularly of marginalized populations, are scarce. The pedagogical models are also limited, which teaches the teachers how to organize and evaluate learning activities with multimodal AI. Unless these gaps are closed, AI potential to transform language learning might not be evenly achieved.

Teachers and researchers need to collaboratively design inclusive, ethical, and flexible models that integrate AI and culturally responsive instruction and multimodal literacy to progress. Only at that point, AI-supported multimodality will stop being a technological novelty, but a real tool of global educational equity.

6. References

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427.

Cerovac, M., & Keane, T. (2024). Early insights into piaget's cognitive development model through the lens of the technologies curriculum. International Journal of Technology and Design Education, 1-21.

Chango, W., Cerezo, R., Sanchez-Santillan, M., Azevedo, R., & Romero, C. (2021). Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources. Journal of Computing inHigher Education, 33, 614-634.

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. British Journal of Educational Technology, 51, 1505-1526.

Foster, E., & Siddle, R. (2020). The effectiveness of learning analytics for identifying atrisk students in higher education. Assessment & Evaluation in Higher Education, 45,842-854.

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., ... & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 1-27.

Gibson, D., Kovanovic, V., Ifenthaler, D., Dexter, S., & Feng, S. (2023). Learning theories for artificial intelligence promoting learning processes. British Journal of Educational Technology, 54, 1125-1146.

Godwin-Jones, R. (2022). Technology-mediated SLAEvolving Trends and Emerging Technologies. *The Routledge handbook of second language acquisition and technology*, 382-394.

Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: a review of emerging educational technology. *Smart Learning Environments*, *11*(1), 22.

Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education, 32*, 241-267.

Jiang, J. (2024). When generative artificial intelligence meets multimodal composition: Rethinking the composition process through an AI-assisted design project. *Computers and Composition*, *74*, 102883.

Jiang, L., & Lai, C. (2025). How Did the Generative Artificial Intelligence-Assisted Digital Multimodal Composing Process Facilitate the Production of Quality Digital Multimodal Compositions: Toward a Process-Genre Integrated Model. *TESOL Quarterly*.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Lateef, U. J., Ahmed, I. I., Hussein, S. K., Yasseen, A. A., & Ghalavandi, H. (2024). Pros and Cons of Multimodality in AI Used by College Students. *Journal of Ecohumanism*, *3*(8), 10353-10361.

Lee, G., Shi, L., Latif, E., Gao, Y., Bewersdorff, A., Nyaaba, M., ... & Zhai, X. (2025). Multimodality of ai for education: Towards artificial general intelligence. *IEEE Transactions on Learning Technologies*.

Li, D., Xia, S., & Guo, K. (2025). Investigating L2 learners' text-to-video resemiotisation in AI-enhanced digital multimodal composing. *Computer Assisted Language Learning*, 1-32.

Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., & He, L. (2022). A survey on text classification: From traditional to deep learning. ACM Transactions on Intelligent Systems and Technology, 13, 1-41.

Lin, C. H., Zhou, K., Li, L., & Sun, L. (2025). Integrating generative AI into digital multimodal composition: A study of multicultural second-language classrooms. *Computers and Composition*, *75*, 102895.

Machado, H., Silva, S., & Neiva, L. (2025). Publics' views on ethical challenges of artificial intelligence: a scoping review. *AI and Ethics*, *5*(1), 139-167.

Mananay, J. A. (2024). Integrating Artificial Intelligence (AI) in Language Teaching: Effectiveness, Challenges, and Strategies. *International Journal of Learning, Teaching and Educational Research*, *23*(9), 361-382.

Mangaroska, K., Martinez-Maldonado, R., Vesin, B., & Gasevic, D. (2021). Challenges and opportunities of multimodal data in human learning: The computer science students' perspective. Journal of Computer Assisted Learning, 37, 1030-1047.

Mohebbi, A. (2025). Enabling learner independence and self-regulation in language education using AI tools: a systematic review. *Cogent Education*, *12*(1), 2433814.

Moon, J., Ke, F., Sokolikj, Z., & Dahlstrom-Hakki, I. (2022). Multimodal data fusion to track students' distress during educational gameplay. Journal of Learning Analytics, 9, 75-87.

moVP, J. (2025). Multimodal Learning Analytics (MMLA) In Education–A Game Changer for Educators. *Indian Journal of Educational Technology*, *7*(1), 329-344.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Ng, D. T. K., Chan, E. K. C., & Lo, C. K. (2025). Opportunities, Challenges and School Strategies for Integrating Generative AI in Education. *Computers and Education: Artificial Intelligence*, 100373.

Nguyen, A., Järvelä, S., Rosé, C., Järvenoja, H., & Malmberg, J. (2023). Examining socially shared regulation and shared physiological arousal events with multimodal learning analytics. British Journal of Educational Technology, 54, 293-312.

Nguyen, T. T. H. (2021). Implementing digital techniques to stimulate EFL students' engagement: A case study in Vietnam. *International Journal of TESOL & Education*, *1*(3), 105-129.

Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P. A., Juuso, I., & Seppänen, T. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. Computers in Human Behavior, 100, 298-304.

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*.

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. British Journal of Educational Technology, 51, 1527--1547.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Revista espanola de cardiologia (English ed.)*, *74*(9), 790-799.

Pozdniakov, S., Brazil, j., Mohammadi, M., Dollinger, M., Sadiq, S., & Khosravi, H. (2025). AI-assisted co-creation: Bridging skill gaps in student-generated content. Journal of Learning Analytics, 12, 129-151.

Prasad, P., Balse, R., & Balchandani, D. (2025). Exploring Multimodal Generative AI for Education through Co-design Workshops with Students. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Qianjing, M., & Lin, T. (2021). An artificial intelligence based construction and application of english multimodal online reading mode. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3721-3730.

Ranade, N., & Eyman, D. (2024). Introduction: Composing with generative AI. *Computers and Composition*, *71*, 102834.

Rashid, S. F., Duong-Trung, N., & Pinkwart, N. (2024). Generative AI in education: technical foundations, applications, and challenges.

Smith, B. E., Shimizu, A. Y., Burriss, S. K., Hundley, M., & Pendergrass, E. (2025). Multimodal composing with generative AI: Examining preservice teachers' processes and perspectives. *Computers and Composition*, *75*, 102896.

Tan, X., Xu, W., & Wang, C. (2025). Purposeful remixing with generative AI: Constructing designer voice in multimodal composing. *Computers and Composition*, *75*, 102893.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, D., & Li, D. (2023). Integrating digital multimodal composition (DMC) into Chinese language teaching. In *Teaching Chinese language in the international school context* (pp. 101-117). Singapore: Springer Nature Singapore.

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in psychology*, *14*, 1261955.

Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 152-167.

Yu, S., Di Zhang, E., & Liu, C. (2024). Research into practice: Digital multimodal composition in second language writing. *Language Teaching*, 1-17.

Zapata, G. C. (Ed.). (2025). *Generative AI Technologies, Multiliteracies, and Language Education*. Taylor & Francis.

Zhang, E. D., & Yu, S. (2025). Conceptualizing digital multimodal composing competence in L2 classroom: A qualitative inquiry. *Computer Assisted Language Learning*, *38*(1-2), 262-290.