



VNU Journal of Foreign Studies

Journal homepage: <https://jfs.ulis.vnu.edu.vn/>

## EMBODIED MULTIMODALITY IN ENGLISH LANGUAGE TEACHING: AFFORDANCES FOR ENRICHMENT AND EMPOWERMENT

Edwin Creely<sup>1,\*</sup>, Mary Walker<sup>2</sup><sup>1</sup>*Faculty of Education, Monash University, Melbourne, Victoria, Australia*<sup>2</sup>*Queensland University of Technology, Brisbane, Queensland, Australia*

Received 25 November 2025

Revised 30 January 2026; Accepted 16 June 2026

**Abstract:** In this conceptual article, we explore the possibilities of multimodality, both digital and non-digital, in English language teaching (ELT) and its capacity to enrich learning and empower learners. We draw on embodied cognition, digital multimodal composing (DMC), and semiotics to propose a way of thinking about multimodality as an entanglement of sensory, cognitive, and sociocultural processes. From an embodied cognition perspective, language learning is conceived as grounded in the body, with gesture, movement, and sensory engagement shaping meaning-making. DMC extends these embodied processes into digital spaces, where learners can weave together text, image, sound, and movement to communicate in authentic and creative ways. Semiotics offers a lens to analyse how meaning is produced and interpreted across modes, revealing the cultural and communicative affordances of multimodal expression. By bringing these perspectives together, we aim to show how multimodality can expand communicative repertoires, foster deeper engagement, and create inclusive opportunities for expression, especially for linguistically diverse learners. Three implications for ELT practice are offered: 1) the need to integrate multimodal design into everyday teaching, 2) to develop assessment practices that value the interplay of modes, and 3) to create learning environments that utilise learners' embodied and cultural resources.

**Keywords:** English language teaching (ELT), embodied cognition, multimodality, digital literacies

---

\* Corresponding author.

Email address: [edwin.creely@monash.edu](mailto:edwin.creely@monash.edu)[https://doi.org/ 10.63023/2525-2445/jfs.ulis.5671](https://doi.org/10.63023/2525-2445/jfs.ulis.5671)

# ĐA PHƯƠNG THỨC HIỆN THÂN TRONG GIẢNG DẠY TIẾNG ANH: TÍNH NĂNG TƯƠNG TÁC LÀM PHONG PHÚ TRẢI NGHIỆM HỌC TẬP VÀ TRAO QUYỀN CHO NGƯỜI HỌC

Edwin Creely<sup>1</sup>, Mary Walker<sup>2</sup>

<sup>1</sup>*Khoa Giáo dục, Đại học Monash, Melbourne, Victoria, Ô-xtrây-li-a*

<sup>2</sup>*Đại học Công nghệ Queensland, Brisbane, Queensland, Ô-xtrây-li-a*

Nhận bài ngày 25 tháng 11 năm 2025

Chỉnh sửa ngày 30 tháng 01 năm 2026; Chấp nhận đăng ngày 16 tháng 6 năm 2026

**Tóm tắt:** Trong bài nghiên cứu này, chúng tôi khám phá những khả năng của tính đa phương thức, bao gồm cả dạng số hóa và phi số hóa, trong giảng dạy tiếng Anh, cũng như khả năng của nó trong việc làm phong phú trải nghiệm học tập và trao quyền cho người học. Căn cứ trên nhận thức hiện thân, soạn thảo kỹ thuật số đa phương thức và ký hiệu học, chúng tôi đề xuất cách tiếp cận xem xét tính đa phương thức như sự đan xen giữa các quá trình cảm giác, nhận thức và văn hóa – xã hội. Từ góc nhìn của nhận thức hiện thân, việc học ngôn ngữ được quan niệm là dựa trên cơ thể, với cử chỉ, chuyển động và sự tương tác giác quan định hình quá trình tạo ra ý nghĩa. Sự sáng tạo đa phương thức kỹ thuật số mở rộng các quá trình mang tính hiện thân vào không gian số, nơi người học có thể kết hợp văn bản, hình ảnh, âm thanh và chuyển động để giao tiếp theo những cách thức chân thực và sáng tạo. Ký hiệu học cung cấp một lăng kính phân tích cách thức ý nghĩa được tạo ra và diễn giải thông qua các phương thức khác nhau, từ đó làm rõ những khả năng biểu đạt về mặt văn hóa và giao tiếp của diễn ngôn đa phương thức. Bằng cách kết hợp các góc nhìn này, chúng tôi muốn chỉ ra rằng tính đa phương thức có thể mở rộng nguồn phương tiện giao tiếp của người học, thúc đẩy sự tham gia sâu hơn vào quá trình học tập và tạo ra những cơ hội biểu đạt mang tính hệ thống, đặc biệt đối với những người học có nền tảng ngôn ngữ đa dạng. Bài viết đưa ra ba hàm ý đối với thực tiễn giảng dạy tiếng Anh: 1) Cần tích hợp thiết kế đa phương thức vào hoạt động giảng dạy hàng ngày; 2) Cần phát triển các hình thức đánh giá coi trọng sự tương tác và phối hợp giữa các phương thức biểu đạt khác nhau; 3) Cần xây dựng môi trường học tập tận dụng hiệu quả các nguồn lực hiện thân và văn hóa của người học.

*Từ khóa:* giảng dạy tiếng Anh, nhận thức hiện thân, tính đa phương thức, năng lực số

## 1. Introduction

The pedagogical approaches in English language teaching (ELT) continue to evolve in response to globalisation, technological advancement, and deepening understandings of how humans learn and communicate (Block & Cameron, 2002). The emergence of generative artificial intelligence (GenAI) presents unique opportunities for multimodal language learning, enabling learners to engage with AI interlocutors that can respond through text, image, and increasingly, voice and gesture (Kohnke et al., 2023). Traditional approaches to ELT, whilst valuable in establishing foundational linguistic competence, increasingly appear insufficient in preparing students for the multimodal communicative demands of contemporary life (Jewitt, 2008). As English maintains its position as a global lingua franca, the need for pedagogical approaches that acknowledge and utilise the full spectrum of human communicative resources, including AI-mediated interactions, becomes ever more pressing (Pennycook, 2017). Current ELT practices often privilege verbal and written modes whilst marginalising the rich array of embodied, visual, and digital resources in new interwoven material learning spaces that students naturally employ

in meaning-making in their everyday literacy practices (Stein, 2008; Toohey, 2019).

The integration of embodied and multimodal practices in language teaching, now enhanced by generative AI capabilities, represents more than a methodological innovation; it constitutes a fundamental reimagining of what language learning entails (Lee et al., 2025; Toohey et al., 2015). When learners engage their bodies through gesture and movement, combine visual and auditory modes in digital compositions, collaborate with AI tools for multimodal content creation, and draw upon their full sensory apparatus in communicative acts, they access deeper cognitive and affective dimensions of language acquisition (Jusslin et al., 2022; McCafferty & Stam, 2008; Yan et al., 2024b). These practices acknowledge that communication has always been multimodal and that effective language pedagogy must reflect this reality, including the emerging reality of human AI collaboration (Norris, 2004). The benefits extend beyond enhanced linguistic competence to encompass increased learner engagement, improved retention, greater inclusivity for diverse learning styles, and the development of critical multimodal literacy skills essential for twenty first century communication, including the ability to critically evaluate and creatively utilise AI generated content (Al-Kadi & Ahmed, 2025; Choi & Yi, 2016; Cope & Kalantzis, 2015; Kohnke et al., 2023).

In this article we examine the theoretical foundations and practical possibilities of embodied multimodality in ELT contexts. Drawing on theories of embodied cognition (Wilson, 2002), multimodal semiotics (Jewitt et al., 2016), and digital multimodal composing (Shin et al., 2020), an integrated framework is proposed to understand and implement multimodal approaches in English language classrooms. Through a detailed thought experiment, we envision pedagogical spaces where learners' full communicative repertoires are recognised, developed, and celebrated, ultimately arguing that such approaches not only enrich the learning experience but fundamentally empower learners to become agentive, creative communicators in multilingual, multimodal worlds (García & Li, 2014).

## **2. Literature Review**

Language learning represents a complex interplay of social and cognitive processes that extend far beyond the acquisition of grammatical structures and vocabulary. Contemporary research in applied linguistics increasingly recognises language learning as a dynamic, situated practice embedded within sociocultural contexts and mediated by embodied experience (Atkinson, 2011). This socio-cognitive perspective challenges traditional dichotomies between mind and body, individual and social, and suggests that language emerges from the continuous interaction of cognitive, physical, and social resources (Larsen-Freeman, 2018). Vygotsky's (1978) sociocultural theory laid crucial groundwork for understanding learning as fundamentally social, whilst recent developments in embodied cognition have illuminated the critical role of sensorimotor experience in shaping linguistic understanding (Barsalou, 2008). The integration of these perspectives reveals language learning as an ecological process wherein learners actively construct meaning through their embodied interactions with environments, artefacts, and other language users (van Lier, 2004).

The status of English as a global lingua franca profoundly shapes contemporary ELT contexts and pedagogical imperatives. Jenkins (2015) argues that English as a lingua franca (ELF) challenges traditional native speaker norms and necessitates pedagogical approaches that prepare learners for diverse, multilingual communicative contexts. In these settings, successful communication often depends less on adherence to standard grammatical forms than on strategic deployment of multiple communicative resources, including gesture, visual aids, and digital tools

(Seidlhofer, 2011). Canagarajah (2013) advocates for a translingual approach that recognises learners' full linguistic repertoires as resources rather than deficits, suggesting that effective ELF communication emerges through the creative negotiation of meaning across linguistic and modal boundaries. This perspective aligns with multimodal approaches that similarly value diverse communicative resources and challenge monomodal, monolingual ideologies (Blommaert, 2010).

Current ELT practices reflect varying degrees of engagement with multimodal and embodied approaches, though traditional text-based methodologies continue to dominate many contexts. Communicative language teaching (CLT), whilst emphasising meaningful interaction, often remains constrained by logocentrism that privileges verbal communication over other modes (Larsen-Freeman & Anderson, 2013). Task based language teaching (TBLT) has created opportunities for more authentic, multimodal communication, yet implementation frequently falls short of fully integrating embodied and digital dimensions (Ellis, 2018). Some innovative approaches, such as drama-based pedagogy and total physical response, explicitly incorporate bodily engagement, demonstrating improved learning outcomes and increased motivation (Piazzoli, 2018; Asher, 2009). However, these remain marginalised within mainstream ELT, often viewed as supplementary rather than central to language learning processes (Winston, 2012).

The emergence of digital literacy practices has begun transforming language learning landscapes, though integration remains uneven across contexts. Digital technologies enable new forms of multimodal composition in which learners can combine text, images, sound, and movement to create meaning (Mills, 2016). Social media platforms, video creation tools, and collaborative digital spaces offer authentic contexts for multimodal communication that mirror real world language use (Reinhardt & Thorne, 2019). Research on digital storytelling in language learning contexts demonstrates how multimodal composition can enhance linguistic development whilst fostering critical digital literacy skills (Oskoz & Elola, 2016). Yi and Angay-Crowder (2016) document how multilingual youth leverage digital tools to construct complex, multimodal texts that express identities and experiences unavailable through monomodal composition. Lambert (2013) shows how digital storytelling empowers marginalised voices through multimodal narrative construction. These studies suggest that digital multimodal practices not only support language learning but also empower learners to become critical, creative producers of meaning (Barnes & Tour, 2023).

Despite growing recognition of multimodality's importance, significant gaps persist between theoretical understanding and classroom implementation (Yi, 2014). Empirical research has noted that the rapid emergence of generative AI technologies has created both opportunities and challenges, as educators grapple with integrating these technologies whilst maintaining pedagogical integrity (Compagnoni, 2025; Kasneci et al., 2023). Teacher education programmes rarely prepare educators to design and assess multimodal learning experiences or to critically evaluate AI-enhanced learning environments, perpetuating traditional, monomodal approaches (Lotherington & Jenson, 2011). Assessment practices continue privileging written texts, failing to recognise or value multimodal competencies that learners develop through digital and embodied practices, including their emerging abilities to collaborate with AI for meaning making (Kress & Selander, 2012). Furthermore, limited access to digital technologies and AI tools in many educational contexts creates inequities in opportunities for multimodal learning (Warschauer & Matuchniak, 2010). These challenges underscore the need for comprehensive frameworks that can guide the integration of embodied and digital multimodal practices, including ethical AI use, into everyday ELT contexts whilst addressing practical constraints and ensuring equitable access to multimodal learning opportunities.

### 3. Conceptual Ideas

The concept of embodied cognition fundamentally reconceptualises the relationship among body, mind, and language, with profound implications for understanding language-learning processes. Embodied cognition theory posits that cognitive processes, including language comprehension and production, are grounded in sensorimotor experience and bodily states (Shapiro, 2019). This perspective challenges traditional computational models of mind that treat cognition as abstract symbol manipulation, proposing instead that thinking emerges through our physical interactions with the world (Clark, 2008). For language, this means that understanding concepts and constructing meaning involve activating sensorimotor simulations grounded in bodily experience. When learners encounter the word "grasp," for instance, motor areas associated with hand movements become active, suggesting that comprehension involves embodied simulation rather than abstract processing (Glenberg & Gallese, 2012). This embodied foundation of language has significant implications for pedagogy, suggesting that engaging learners' bodies through gesture, movement, and multisensory experience can facilitate deeper linguistic understanding and retention (Mavilidi et al., 2015). This is supported by sensory and embodied learning offering deeper understandings of social interactions, cultural contexts, and this can influence cognitive processes (Sweller, 2022) The sensory aspect of embodied learning affects the thinking and learning processes and can support students to remember the learning content more comprehensively (Alibali & Nathan, 2012; Khatin-Zadeh et al., 2021; Shapiro et al., 2020). This offers not only improved understanding for the learners but a way to improve learning performance and reduce cognitive load (Lyu & Deng, 2024; Zou et al., 2025).

Research in gesture studies provides compelling evidence for the embodied nature of language and its pedagogical applications. McNeill (2005) demonstrates that gesture and speech form an integrated system, with gestures not merely accompanying but co-constituting linguistic meaning. Alibali & Nathan (2012) found that teachers and learners produced gestures when explaining mathematical concepts and ideas that were essential to learning and demonstrated that mathematical cognition was embodied. In language learning contexts, gestures serve multiple functions: they facilitate word retrieval, support comprehension of abstract concepts, and provide additional communicative resources when linguistic competence is limited (Macedonia & von Kriegstein, 2012). Studies show that encoding new vocabulary with gestures significantly improves retention compared to verbal-only learning, with neural imaging revealing that gesture-enriched words activate broader neural networks encompassing motor regions (Kelly et al., 2009). Tellier (2008) found that teaching French vocabulary to children through gesture resulted in significantly better memorisation than visual or verbal only instruction. These findings suggest that pedagogical approaches that systematically integrate gesture and movement can enhance language acquisition by engaging embodied cognitive processes that support memory consolidation and conceptual understanding.

Multimodality and semiotics provide theoretical frameworks for understanding how meaning emerges through the orchestration of multiple semiotic resources. Kress (2010) argues that communication has always been multimodal, with different modes offering distinct affordances for meaning making. Each mode possesses specific materialities and cultural histories that shape its communicative potential: images offer spatial simultaneity, sound provides temporal sequencing, and gesture enables embodied demonstration (Bezemer & Kress, 2016). From a social semiotic perspective, meaning makers select and combine modes based on available resources, communicative purposes, and social contexts (Halliday & Matthiessen, 2014). This multimodal orchestration involves complex processes of transduction,

where meaning is moved across modes, and transformation, where meaning shifts through modal re-expression (Kress & van Leeuwen, 2001). Newfield (2014) demonstrates how transduction across modes in pedagogical contexts creates opportunities for expanded meaning-making. Understanding these processes is crucial for language education, as learners must develop not only linguistic competence but also multimodal communicative competence to navigate contemporary communicative landscapes. Digital multimodal composing represents the convergence of embodied, multimodal, and digital practices in creating new possibilities for meaning-making and identity construction. DMC involves the deliberate design and production of digital texts that integrate multiple modes to achieve communicative purposes (Smith et al., 2017).

The integration of generative AI tools has expanded these possibilities exponentially, enabling learners to collaborate with AI for image generation, text refinement, voice synthesis, and even video creation, though this requires critical evaluation of AI contributions and understanding of their limitations (Jeon & Lee, 2023). Unlike traditional writing, DMC requires composers to consider modal affordances, design principles, and audience engagement across multiple semiotic channels (Hafner, 2015). This process demands both technical skills and critical multimodal literacy, including AI literacy for language learners, as composers must make informed decisions about modal selection, combination, transformation, and the appropriate integration of AI-generated elements. In language learning contexts, DMC enhanced by AI collaboration offers opportunities for learners to express complex meanings that may exceed their current linguistic capabilities whilst developing twenty first century literacy skills (Tour, 2015). The creative agency involved in multimodal composition, now augmented by generative AI partnerships, can be particularly empowering for linguistically diverse learners, who can draw upon their full semiotic repertoires and AI assistance rather than being constrained by limited proficiency in English (Hafner, 2014; Pack & Maloney, 2023). Through AI-enhanced DMC, learners become critical designers of meaning, actively constructing knowledge and identity through the orchestration of diverse communicative resources whilst developing essential skills in evaluating and ethically utilising AI-generated content.

While the use of AI has significant potential, it is also important to acknowledge that literature on generative AI in language education presents a more contested picture than the preceding discussion may suggest. While proponents emphasise AI's capacity to provide personalised feedback, scale learner support, and diversify pedagogical resources, critical scholarship identifies significant countervailing concerns (Yan et al., 2024a). Studies document notable limitations in accuracy and intelligibility, particularly for deeper structural and pragmatic aspects of learning, while surface-level corrections may foster overreliance that undermines the development of autonomous language skills (Barrot, 2023). The impact of AI on critical thinking also remains contested, with some evidence indicating that AI-assisted learning may reduce cognitive engagement and independent reasoning (Kasneci et al., 2023). Ethical dilemmas surrounding academic integrity, algorithmic bias, and equitable access persist, and longitudinal evidence regarding learning outcomes remains sparse. In sum, the transformative potential of AI must be subject to ongoing empirical scrutiny and implemented with vigilance with awareness of these documented limitations and debates.

#### **4. Possibilities in ELT**

The integration of embodied multimodality in English language teaching (ELT) opens transformative possibilities that extend far beyond traditional pedagogical boundaries. Current

research demonstrates that multimodal approaches can significantly enhance vocabulary acquisition (Rowe & Miller, 2016), improve grammatical understanding, and foster more authentic communicative competence (Early & Norton, 2012). By recognising language as fundamentally multimodal and embodied, educators can design learning experiences that engage learners' full sensory and cognitive capacities, creating deeper and more durable learning outcomes (Lapaire, 2016; Rosborough, 2014).

Contemporary classroom implementations of multimodal pedagogy reveal promising developments across diverse contexts. Anderson and Macleroy (2016) document how multilingual learners in London schools use digital storytelling to navigate complex identity negotiations whilst developing sophisticated language skills. In Japanese university contexts, Humphries and Burns (2015) show how integrating gesture and movement into pronunciation instruction significantly improves learners' prosodic awareness and production. These empirical studies suggest that multimodal approaches not only support traditional learning objectives but also develop new competencies essential to twenty-first-century communication (Pacheco & Smith, 2015).

The pedagogical affordances of different modes offer unique opportunities for addressing persistent challenges in language education. Visual modes can scaffold comprehension for learners with limited linguistic proficiency, whilst gestural modes provide embodied anchors for abstract grammatical concepts (Morett, 2014). Audio modes enable attention to prosodic features often neglected in text-based instruction, and digital modes facilitate authentic audience engagement that motivates meaningful communication (Vandommele et al., 2017). When orchestrated thoughtfully, these modal resources create rich semiotic environments that support diverse learning styles and linguistic backgrounds (Guichon & McLornan, 2008).

### **5. A Thought Experiment: The Fully Embodied and Multimodal ELT Classroom**

Envisioning a fully embodied and multimodal English language classroom requires reimagining both physical and pedagogical architectures to support diverse forms of meaning-making. Consider a learning space where traditional desk rows have been replaced with flexible zones: a movement area with mirrors and recording equipment for embodied language work, collaborative digital stations equipped with tablets and creation software including generative AI tools, a maker space with art supplies and manipulatives, and quiet reflection pods for individual multimodal composition with AI assistants. The walls display QR codes linking to student-created augmented reality experiences, whilst interactive projections respond to learner movements, transforming gestures into visual poetry through AI-mediated interpretation. This environment acknowledges that language learning happens through bodies in space and through collaboration with intelligent systems, requiring pedagogical environments that facilitate movement, creation, human AI interaction, and multimodal expression.

In this thought experiment, a typical learning sequence might begin with embodied vocabulary introduction through a technique I call "semantic embodiment mapping," now enhanced with AI visual generation. Rather than presenting new vocabulary through translation or definition, the teacher guides learners in exploring words through physical experience, whilst AI systems generate corresponding visual representations in real time. For the concept "resilience," learners first experience physical resistance through pushing against walls or partners, feeling muscle tension and release. Simultaneously, they prompt generative AI to create images that capture their embodied understanding, comparing AI interpretations with their

physical experiences. They then create gesture sequences that capture the essence of bouncing back, using their bodies to understand the metaphorical extension from physical to psychological resilience, whilst AI tools help them generate multilingual explanations and visual metaphors. These embodied explorations are captured through video, with AI transcription and analysis tools helping learners identify patterns in their gestural language, building personal gesture dictionaries that serve as embodied memory aids (Alibali & Nathan, 2012).

Grammar instruction transforms through what we term "syntactic choreography," where grammatical structures become embodied through movement patterns, and AI-generated visual narratives. Present perfect tense, often challenging for learners whose languages lack similar structures, becomes a physical journey through space augmented by AI-generated timeline visualisations. Learners position themselves along a timeline stretched across the floor, physically moving from past positions to present locations whilst carrying objects representing completed actions that maintain present relevance. AI voice assistants provide real time feedback on their verbal productions, whilst generative AI creates personalised visual stories that reinforce the grammatical concept. They create human sentences, with individuals embodying different grammatical elements, physically rearranging themselves to explore word order variations and their semantic implications, whilst AI systems generate multiple linguistic variations for comparison. This embodied approach, enhanced by AI scaffolding, makes abstract grammatical concepts tangible, allowing learners to literally feel the connections between form and meaning whilst receiving personalised AI-generated examples relevant to their interests and proficiency levels (Huang & Wang, 2023; Jusslin, 2022; Witte, 2026).

Digital multimodal composing projects in this imagined classroom provide authentic contexts for language use whilst developing critical twenty-first-century literacies, including AI literacy. In one extended project, learners create "identity portraits" that weave together multiple modes to express their linguistic journeys and cultural identities, collaborating with generative AI as a creative partner. Beginning with audio recordings of their pronunciation evolution analysed by AI for patterns and progress, learners layer in visual timelines showing their encounters with English across different contexts, with AI helping generate culturally relevant imagery. They use AI text generation tools critically, editing and adapting suggestions to maintain their authentic voice whilst improving linguistic accuracy. They incorporate family photographs enhanced with AI generated artistic filters, create word clouds from significant texts in their lives with AI assistance in identifying themes, and design soundscapes mixing languages, music, and ambient sounds from meaningful places, with AI helping to compose transitional elements. Using tools like Adobe Spark integrated with generative AI capabilities, learners assemble these elements into interactive digital stories that resist linear narrative, instead offering viewers multiple pathways through their multilingual, multimodal experiences. Throughout, learners develop critical AI literacy, learning to evaluate AI suggestions, maintain authorial control, and understand the ethical implications of AI generated content (Pokrivcakova, 2019).

Collaborative multimodal projects foster genuine communicative purpose whilst building intercultural competence and AI collaboration skills. Groups engage in "cultural remix" projects where they select proverbs or idioms from their various linguistic backgrounds and translate them across modes and cultures, using AI as a research assistant and creative collaborator. A Chinese proverb about patience might become a stop motion animation using found objects, with AI generating intermediate frames and suggesting visual metaphors, whilst learners maintain creative control over narrative and meaning. A Spanish saying about community could transform into a collaborative mural combining human created visual

metaphors with AI generated elements, creating QR codes linking to multilingual audio explanations recorded by students and enhanced with AI generated background music. These projects require extensive negotiation of meaning as learners must understand concepts deeply enough to transduce them across modes and cultures, whilst critically evaluating AI contributions for cultural appropriateness and accuracy. The process develops both translingual competence and AI literacy as learners recognise connections across languages whilst using English and AI tools as mediating instruments for intercultural exchange (Hong, 2023).

Assessment in this multimodal classroom shifts from standardised testing towards portfolio-based approaches that document learning across modes and over time. Learners compile multimodal portfolios showcasing their developing communicative repertoires through diverse artefacts: video reflections on learning processes, gesture glossaries for new vocabulary, digital stories exploring grammatical concepts created with AI assistance, collaborative multimedia presentations, and augmented reality tours of significant spaces. AI powered assessment tools provide detailed analytics on linguistic development whilst learners maintain reflective journals documenting their AI collaboration processes and critical evaluation skills (Mizumoto & Eguchi, 2023). Assessment criteria value modal aptness, creative meaning making, ethical AI use, and communicative effectiveness rather than narrow linguistic accuracy. Peer assessment becomes crucial, with learners developing multimodal annotation skills as they provide feedback on each other's compositions using voice notes, visual markups, and embedded comments, whilst AI tools help identify patterns and suggest areas for improvement. This approach recognises that communicative competence in contemporary contexts requires ability to orchestrate multiple modes effectively and collaborate critically with AI systems rather than mastery of decontextualised language forms.

## **6. Pedagogical Frameworks and Implementation Strategies**

The implementation of multimodal pedagogy requires systematic frameworks that guide teachers in designing, facilitating, and assessing multimodal learning experiences whilst integrating generative AI tools responsibly. The Learning by Design framework (Cope & Kalantzis, 2015) offers a robust model that integrates multimodal meaning making across experiencing, conceptualising, analysing, and applying knowledge processes, now expanded to include critical AI engagement at each stage. This framework supports teachers in scaffolding learners' development from situated experience through abstract understanding to critical analysis and creative application across modes, whilst developing what AI-enhanced pedagogical reasoning (Ntelioglou et al., 2014).

Teacher roles transform significantly in multimodal classrooms, shifting from knowledge transmitters to designers of learning experiences, facilitators of meaning-making processes, and guides for ethical AI use (Creely, 2024). Teachers must develop multimodal orchestration skills, learning to coordinate multiple modes simultaneously whilst maintaining pedagogical coherence and helping learners navigate AI affordances and limitations (Lim et al., 2022). This involves real-time decision-making about modal selection, timing of modal shifts, appropriate AI tool integration, and responsive adjustments based on learner engagement and understanding. Professional development becomes crucial, requiring sustained support for teachers to develop theoretical understanding, practical skills in multimodal pedagogy, and critical perspectives on AI in education (Tan et al., 2020; Celik et al., 2022).

Technology integration in multimodal ELT extends beyond simple tool use to encompass critical engagement with digital affordances, constraints, and the ethical

implications of AI use. Virtual reality environments offer immersive contexts for embodied language practice, allowing learners to inhabit different cultural settings and practice contextually appropriate communication with AI-powered conversational agents (Kaplan-Rakowski & Gruber, 2019). Augmented reality applications enable situated language learning that connects classroom learning to real-world contexts, with AI providing personalised, context-aware language support (Christou et al., 2025). Generative AI tools offer unprecedented opportunities for multimodal content creation, from image generation to voice synthesis, though this requires careful pedagogical framing to ensure learners develop critical evaluation skills alongside creative capabilities. However, effective integration requires pedagogical rather than technological drivers, ensuring that digital and AI tools enhance rather than constrain embodied and multimodal learning (Godwin-Jones, 2018).

## **7. Implications for Practice**

The integration of multimodal design into everyday teaching represents a fundamental shift in pedagogical practice that extends beyond occasional multimodal activities to encompass comprehensive redesign of learning experiences. This is imperative if we wish to support the reduction of cognitive load associated with learning and to promote a deeper understanding among learners (Lyu & Deng, 2024; Shapiro et al., 2020). Teachers must develop multimodal pedagogical content knowledge, understanding not only how different modes function semiotically but also how to scaffold learners' development of multimodal competencies within language learning objectives (Lim & Tan, 2018). This involves reconceptualising lesson planning to consider modal orchestration: which modes best support particular learning goals? How can transitions between modes deepen understanding? What modal combinations enable differentiated learning for diverse students? For instance, vocabulary instruction might systematically integrate gesture protocols, visual mapping techniques, and digital annotation tools, with teachers explicitly discussing modal affordances and helping learners develop personal multimodal learning strategies (Royce, 2002).

Professional development programmes must address the multimodal competency gap amongst educators, providing sustained support for teachers to develop confidence in multimodal pedagogy. This includes technical skills for digital multimodal composing, but more critically, theoretical understanding of multimodality and practical strategies for implementation within existing curricular constraints (Edwards-Groves, 2011). Teacher education should model multimodal approaches, with educators experiencing embodied and digital multimodal learning firsthand before attempting implementation (Ajayi, 2011). Collaborative teacher research projects can document effective practices whilst addressing context-specific challenges, building local knowledge about multimodal pedagogy that reflects diverse educational settings (Jewitt, 2008). Schools must also reconsider resource allocation, investing not only in technology but in flexible learning spaces, professional development time, and ongoing support for multimodal innovation.

Assessment practices require fundamental reconceptualisation to value the interplay of modes rather than privileging traditional written texts. This involves developing assessment frameworks that recognise multimodal competencies as legitimate learning outcomes, with criteria that capture the complexity of modal orchestration (Hung et al., 2013). Rather than evaluating linguistic accuracy in isolation, assessment might consider communicative effectiveness across modes, creativity in modal selection and combination, critical awareness of modal affordances and constraints, and ability to adapt modal use for different purposes and

audiences (Pandya, 2018). Portfolio assessment becomes particularly valuable, allowing learners to demonstrate development across time and modes whilst reflecting on their evolving multimodal repertoires (Yancey, 2004). Importantly, assessment must avoid technocentric bias that values digital sophistication over communicative purpose, maintaining focus on meaning-making rather than technical proficiency.

Creating learning environments that utilise learners' embodied and cultural resources demands recognition of the diverse communicative practices they bring to the classroom. This involves conducting multimodal ethnographies of learners' out of school communicative practices, documenting how they naturally orchestrate modes in digital communication, gaming, and social interaction (Domingo et al., 2014). These insights inform pedagogical design that builds on existing competencies rather than dismissing them as irrelevant to academic learning. For multilingual learners, this means valuing translanguaging practices that blend languages and modes, recognising code meshing and multimodal code switching as sophisticated communicative strategies rather than deficiencies (Canagarajah, 2011). Teachers must create inclusive spaces where diverse embodied communicative practices, from cultural specific gestures to digital vernaculars, become resources for collective learning rather than barriers to overcome (Blackledge & Creese, 2017).

The physical and digital infrastructures of learning spaces must evolve to support embodied and multimodal practices. This extends beyond adding technology to fundamentally reconsidering how space enables or constrains different forms of meaning making (Gee, 2004). Flexible furniture arrangements allow rapid reconfiguration for movement activities, collaborative work, and individual composition. Display spaces showcase multimodal work, with digital screens, projection surfaces, and maker space areas enabling diverse forms of expression. Acoustic considerations become important as learners engage in simultaneous verbal, musical, and movement-oriented activities. Digital infrastructure must support multimedia creation and sharing, with robust wifi, device availability, and platforms for multimodal collaboration (Burnett, 2016). However, schools must also address equity concerns, ensuring all learners have access to digital tools and connectivity for multimodal learning both in and outside classroom spaces.

Implementation requires navigating institutional constraints whilst advocating for systemic change that values multimodal literacy. Standardised testing regimes that prioritise discrete linguistic skills create tensions with multimodal approaches that develop integrated communicative competencies (Menken, 2008). Teachers must become skilled at articulating how multimodal practices support traditional learning objectives whilst also developing new competencies essential for contemporary communication (Rowse et al., 2008). This involves strategic documentation of learning outcomes, gathering evidence that multimodal approaches enhance rather than distract from language acquisition. Building stakeholder support requires demonstrating multimodal learning to parents, administrators, and policymakers, helping them understand how these approaches prepare learners for academic, professional, and civic participation in multimodal societies (Walsh, 2010).

Ethical considerations emerge around digital footprints, cultural representation, consent in multimodal composition, and the responsible use of generative AI (Creely, 2024). When learners create digital multimodal texts incorporating personal images, voices, and stories, potentially enhanced with AI generated content, educators must address privacy concerns, digital citizenship, and issues of authorship and authenticity (Burnett & Merchant, 2018). This includes teaching learners about intellectual property in multimodal composition, the ethics of remixing

and appropriation, the importance of disclosing AI use, and the implications of sharing multimodal work online (Kostopolus, 2025). The integration of generative AI raises new ethical questions about originality, academic integrity, and the development of genuine communicative competence, requiring explicit discussion of when and how AI assistance is appropriate (Rudolph et al., 2023). Cultural sensitivity becomes crucial when embodied practices involve touching, proximity, or gestures that may carry different meanings across cultures, and when AI systems may perpetuate cultural biases (Hall, 2012). Teachers must navigate these complexities thoughtfully, creating agreements with learners about sharing, representation, and AI use whilst fostering critical discussion about power, privilege, and algorithmic bias in multimodal texts.

## 8. Conclusion

This exploration of embodied multimodality in English language teaching reveals transformative possibilities for enriching learning experiences and empowering diverse learners. By integrating insights from embodied cognition, multimodal semiotics, digital multimodal composing, and emerging generative AI capabilities, we can envision pedagogical approaches that honour the full spectrum of human and technologically augmented communicative resources (Norton & Toohey, 2011; Lee et al., 2025). The thought experiment presented demonstrates how classrooms can become spaces where gesture, movement, image, sound, digital creation, and AI collaboration converge to support deep, engaged language learning. These approaches acknowledge that effective communication in globalised, digitised, AI-enhanced contexts requires orchestration of multiple modes and critical engagement with intelligent systems, making multimodal and AI competence essential rather than supplementary to language education (Unsworth, 2014).

The implications for practice demand significant shifts in pedagogical design, assessment frameworks, and learning environments. Teachers must develop new competencies in multimodal pedagogy and AI integration whilst advocating for institutional changes that value diverse forms of meaning-making (Stein & Newfield, 2006). Implementation requires careful attention to equity, ensuring all learners have access to opportunities for multimodal expression and AI-enhanced learning whilst avoiding the creation of new forms of disadvantage (Vasudevan et al., 2010). The challenges are substantial, involving resistance from traditional assessment regimes, resource limitations, ethical concerns about AI use, and the need for extensive professional development. However, the potential benefits, including increased engagement, improved retention, greater inclusivity, development of twenty first century literacies, and preparation for an AI integrated future, justify sustained effort towards multimodal transformation (Al-Kadi & Ahmed, 2025; Jewitt et al., 2016).

At the same time, it is important to acknowledge that the synthesis presented in this article draws upon research conducted across diverse educational contexts, representing many regions, each shaped by distinct institutional cultures, policy environments, resource configurations, and learner populations. The possibilities articulated through the thought experiment, for example, must be understood as heuristic rather than prescriptive, offering conceptual orientations that require careful adaptation to specific teaching and learning ecologies, which is especially relevant between western and Asian contexts. What proves transformative in a well-resourced urban language centre may encounter different affordances and constraints in a regional community college or an under-resourced adult migration programme. Similarly, the integration of generative AI technologies raises varying ethical, practical, and infrastructural considerations across different national and institutional contexts,

where access, policy frameworks, and cultural attitudes toward technology diverge significantly. Future research must attend more systematically to this contextual variability, documenting not only successful implementations but also the specific conditions that enable or constrain multimodal and AI-enhanced pedagogies across settings. Such context-sensitive inquiry will be essential for developing nuanced understandings of how embodied multimodality can be realised in ways that are responsive to local needs, resources, and constraints whilst remaining oriented toward the broader transformative vision outlined in this article.

Moving forward, the ELT community must embrace multimodality and AI literacy not as peripheral to existing practice but as fundamental to contemporary language education. This involves continued research documenting effective multimodal practices across diverse contexts, development of assessment tools that capture multimodal and AI collaborative competencies, and creation of teacher education programmes that prepare educators for multimodal, AI-enhanced pedagogy. As English continues evolving as a global lingua franca in an increasingly AI-mediated world, our pedagogical approaches must similarly evolve to prepare learners for communicative futures we cannot fully anticipate. By centring embodied, multimodal practices and critical AI engagement in language education, we empower learners to become creative, critical, and competent communicators capable of orchestrating the full range of semiotic and technological resources available in our richly multimodal, AI-augmented world (Cope & Kalantzis, 2009).

### References

- Ajayi, L. (2011). Preservice teachers' knowledge, attitudes, and perception of their preparation to teach multiliteracies/multimodality. *The Teacher Educator*, 46(1), 6–31. <https://doi.org/10.1080/08878730.2010.488279>
- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, 21(2), 247–286. <https://doi.org/10.1080/10508406.2011.611446>
- Al-Kadi, A., & Ahmed, R. (2025). Inclusivity and learner engagement through multimodal approaches in English language learning. *Learning and Teaching in Higher Education: Gulf Perspectives*, 1–16. <https://doi.org/10.1108/LTHE-02-2025-0007>
- Anderson, J., & Macleroy, V. (2016). *Multilingual digital storytelling: Engaging creatively and critically with literacy*. Routledge.
- Asher, J. J. (2009). *Learning another language through actions (7th ed.)*. Sky Oaks Productions.
- Atkinson, D. (2011). *Alternative approaches to second language acquisition*. Routledge. <https://doi.org/10.4324/9780203830932>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Barnes, M., & Tour, E. (2023). Empowering English as an additional language students through digital multimodal composing. *Literacy*, 57(2), 106–119. <https://doi.org/10.1111/lit.12319>
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge. <https://doi.org/10.4324/9781315687537>
- Blackledge, A., & Creese, A. (2017). Translanguaging and the body. *International Journal of Multilingualism*, 14(3), 250–268. <https://doi.org/10.1080/14790718.2017.1315809>
- Block, D., & Cameron, D. (Eds.). (2002). *Globalization and language teaching*. Routledge. <https://doi.org/10.4324/9780203193679>
- Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511845307>

- Burnett, C. (2016). Being together in classrooms at the interface of the physical and virtual: Implications for collaboration in on/off screen sites. *Learning, Media and Technology*, 41(4), 566–589. <https://doi.org/10.1080/17439884.2015.1050036>
- Burnett, C., & Merchant, G. (2018). Literacy-as-event: Accounting for relationality in literacy research. *Discourse: Studies in the Cultural Politics of Education*, 41(1), 45–56. <https://doi.org/10.1080/01596306.2018.1460318>
- Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal*, 95(3), 401–417. <https://doi.org/10.1111/j.1540-4781.2011.01207.x>
- Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations*. Routledge. <https://doi.org/10.4324/9780203073889>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Choi, J., & Yi, Y. (2016). Teachers' integration of multimodality into classroom practices for English language learners. *TESOL Journal*, 7(2), 304–327. <https://doi.org/10.1002/tesj.204>
- Christou, E., Vassiliou, P., & Parmaxi, A. (2025). Augmented reality in language learning: a systematic literature review of the state-of-the-art and task design considerations. *Innovation in Language Learning and Teaching*, 1–28. <https://doi.org/10.1080/17501229.2025.2504706>
- Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195333213.001.0001>
- Compagnoni, I. (2025). Pedagogical implications of AI-enhanced digital storytelling in EFL education. *International Journal of Linguistics*, 17(5). <https://doi.org/10.5296/ijl.v17i5.22773>
- Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal*, 4(3), 164–195. <https://doi.org/10.1080/15544800903076044>
- Cope, B., & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 1–36). Palgrave Macmillan. [https://doi.org/10.1057/9781137539724\\_1](https://doi.org/10.1057/9781137539724_1)
- Creely, E. (2024). Exploring the Role of Generative AI in Enhancing Language Learning: Opportunities and Challenges. *International Journal of Changes in Education*, 1(3), 158–167. <https://doi.org/10.47852/bonviewIJCE42022495>
- Domingo, M., Jewitt, C., & Kress, G. (2014). Multimodal social semiotics: Writing in online contexts. In K. Pahl & J. Rowsell (Eds.), *The Routledge handbook of contemporary literary studies* (pp. 251–266). Routledge.
- Early, M., & Norton, B. (2012). Language learner stories and imagined identities. *Narrative Inquiry*, 22(1), 194–201. <https://doi.org/10.1075/ni.22.1.15ear>
- Edwards-Groves, C. (2011). The multimodal writing process: Changing practices in contemporary classrooms. *Language and Education*, 25(1), 49–64. <https://doi.org/10.1080/09500782.2010.523468>
- Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters. <https://doi.org/10.21832/9781788920148>
- García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan. <https://doi.org/10.1057/9781137385765>
- Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge. <https://doi.org/10.4324/9780203594216>
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7), 905–922. <https://doi.org/10.1016/j.cortex.2011.04.010>
- Godwin-Jones, R. (2018). *Using mobile devices in the language classroom: Part of the Cambridge Papers in ELT series*. Cambridge University Press.
- Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85–93. <https://doi.org/10.1016/j.system.2007.11.005>
- Hafner, C. A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly*, 48(4), 655–685. <https://doi.org/10.1002/tesq.138>
- Hafner, C. A. (2015). Remix culture and English language teaching: The expression of learner voice in digital multimodal compositions. *TESOL Quarterly*, 49(3), 486–509. <https://doi.org/10.1002/tesq.238>

- Hall, J. K. (2012). *Teaching and researching: Language and culture (2nd ed.)*. Pearson Education.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An Introduction to Functional Grammar (4th ed.)*. Routledge. <https://doi.org/10.4324/9780203783771>
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1). <https://doi.org/10.61414/jeti.v5i1.103>
- Huang, T. H., & Wang, L. Z. (2023). Artificial intelligence learning approach through total physical response embodiment teaching on French vocabulary learning retention. *Computer Assisted Language Learning*, 36(8), 1608–1632. <https://doi.org/10.1080/09588221.2021.2008980>
- Humphries, S., & Burns, A. (2015). 'In reality it's almost impossible': CLT-oriented curriculum change. *ELT Journal*, 69(3), 239–248. <https://doi.org/10.1093/elt/ccu081>
- Hung, Y. W., Chiang, C. Y., & Chiu, C. Y. (2013). Using multimodal assessment to explore students' learning of science. *International Journal of Science Education*, 35(14), 2426–2443. <https://doi.org/10.1080/09500693.2012.665194>
- Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice*, 2(3), 49–85. <https://doi.org/10.1515/eip-2015-0003>
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241–267. <https://doi.org/10.3102/0091732X07310586>
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge. <https://doi.org/10.4324/9781315638027>
- Jusslin, S., Korpinen, K., Lilja, N., Martin, R., Lehtinen-Schnabel, J., & Anttila, E. (2022). Embodied learning and teaching approaches in language education: A mixed studies review. *Educational Research Review*, 37, Article 100480. <https://doi.org/10.1016/j.edurev.2022.100480>
- Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the Innovation in Language Learning International Conference* (pp. 552–555). Pixel.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24(2), 313–334. <https://doi.org/10.1080/01690960802365567>
- Khatin-Zadeh, O., Eskandari, Z., Cervera-Torres, S., Ruiz Fernández, S., Farzi, R., & Marmolejo-Ramos, F. (2021). The strong versions of embodied cognition: Three challenges faced. *Psychology & Neuroscience*, 14(1), 16–33. <https://doi.org/10.1037/pne0000252>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Kostopolus, E. (2025). Student use of generative AI as a composing process supplement: Concerns for intellectual property and academic honesty. *Computers and Composition*, 75, Article 102894. <https://doi.org/10.1016/j.compcom.2024.102894>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge. <https://doi.org/10.4324/9780203970034>
- Kress, G., & Selander, S. (2012). Multimodal design, learning and cultures of recognition. *The Internet and Higher Education*, 15(4), 265–268. <https://doi.org/10.1016/j.iheduc.2011.12.003>
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Lambert, J. (2013). *Digital storytelling: Capturing lives, creating community (4th ed.)*. Routledge. <https://doi.org/10.4324/9780203102329>

- Lapaire, J.-R. (2016). The choreography of time: Metaphor, gesture and construal. In R. Gabriel & A. C. Pelosi (Eds.), *Linguagem e cognição: emergência e produção de sentidos* (pp. 217–234). Insular. [https://www.researchgate.net/publication/311439570\\_The\\_choreography\\_of\\_time\\_metaphor\\_gesture\\_and\\_construal](https://www.researchgate.net/publication/311439570_The_choreography_of_time_metaphor_gesture_and_construal)
- Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals*, 51(1), 55–72. <https://doi.org/10.1111/flan.12314>
- Larsen-Freeman, D., & Anderson, M. (2013). *Techniques and principles in language teaching (3rd ed.)*. Oxford University Press.
- Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34(1), 335–359. <https://doi.org/10.1080/10494820.2025.2498537>
- Lim, F. V., & Tan, K. Y. S. (2018). Developing multimodal literacy through teaching the critical viewing of films in Singapore. *Journal of Adolescent & Adult Literacy*, 62(3), 291–300. <https://doi.org/10.1002/jaal.882>
- Lim, F. V., Weninger, C., & Nguyen, T. T. H. (2021). "I expect boredom": Students' experiences and expectations of multiliteracies learning. *Literacy*, 55(2), 102–112. <https://doi.org/10.1111/lit.12243>
- Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics*, 31, 226–246. <https://doi.org/10.1017/S0267190511000110>
- Lyu, C., & Deng, S. (2024). Effectiveness of embodied learning on learning performance: A meta-analysis based on the cognitive load theory perspective. *Learning and Individual Differences*, 116, 102564. <https://doi.org/10.1016/j.lindif.2024.102564>
- Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, 6(3-4), 393–416. <https://doi.org/10.5964/bioling.8931>
- Mavilidi, M. F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review*, 27, 413–426. <https://doi.org/10.1007/s10648-015-9337-z>
- McCafferty, S. G., & Stam, G. (Eds.). (2008). *Gesture: Second language acquisition and classroom research*. Routledge. <https://doi.org/10.4324/9780203866993>
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226514642.001.0001>
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Multilingual Matters.
- Mills, K. A. (2016). *Literacy theories for the digital age: Social, critical, multimodal, spatial, material and sensory lenses*. Multilingual Matters. <https://doi.org/10.21832/9781783094639>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, 98(3), 834–853. <https://doi.org/10.1111/modl.12125>
- Newfield, D. (2014). Transformation, transduction and the transmodal moment. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 100–114). Routledge.
- Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge. <https://doi.org/10.4324/9780203379493>
- Norton, B., & Toohey, K. (2011). Identity, language learning, and social change. *Language Teaching*, 44(4), 412–446. <https://doi.org/10.1017/S0261444811000309>
- Ntelioglou, B. Y., Fannin, J., Montanera, M., & Cummins, J. (2014). A multilingual and multimodal approach to literacy teaching and learning in urban education: A collaborative inquiry project in an inner city elementary school. *Frontiers in Psychology*, 5, 533. <https://doi.org/10.3389/fpsyg.2014.00533>
- Oskoz, A., & Elola, I. (2016). Digital stories: Bringing multimodal texts to the Spanish writing classroom. *ReCALL*, 28(3), 326–342. <https://doi.org/10.1017/S0958344016000094>

- Pacheco, M. B., & Smith, B. E. (2015). Across languages, modes, and identities: Bilingual adolescents' multimodal codemeshing in the literacy classroom. *Bilingual Research Journal*, 38(3), 292–312. <https://doi.org/10.1080/15235882.2015.1091051>
- Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, 57(4), 1571–1582. <https://doi.org/10.1002/tesq.3253>
- Pandya, J. Z. (2018). *Exploring critical digital literacy practices: Everyday video in a dual language context*. Routledge. <https://doi.org/10.4324/9781315102672>
- Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge. <https://doi.org/10.4324/9781315225593>
- Piazzoli, E. (2018). *Embodying language in action: The artistry of process drama in second language education*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-77962-1>
- Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, 7(3), 135–153. <https://doi.org/10.2478/jolace-2019-0025>
- Reinhardt, J., & Thorne, S. (2019). Digital literacies as emergent multifarious literacies. In N. Arnold & L. Ducate (Eds.), *Engaging language learners through CALL* (pp. 208–239). Equinox.
- Rosborough, A. (2014). Gesture, meaning-making, and embodiment: Second language learning in an elementary classroom. *Journal of Pedagogy*, 5(2), 227–250. <https://doi.org/10.2478/jped-2014-0011>
- Rowe, L. W., & Miller, P. M. (2016). Designing for diverse learners: Using universal design and multimodal pedagogy in the classroom. In M. Kalantzis, B. Cope, & D. Zapata (Eds.), *International perspectives on education* (pp. 199–216). Common Ground.
- Rowsell, J., Kosnik, C., & Beck, C. (2008). Fostering multiliteracies pedagogy through preservice teacher education. *Teaching Education*, 19(2), 109–122. <https://doi.org/10.1080/10476210802040799>
- Royce, T. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly*, 36(2), 191–205. <https://doi.org/10.2307/3588330>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.
- Shapiro, L. (2019). *Embodied cognition (2nd ed.)*. Routledge. <https://doi.org/10.4324/9781315180380>
- Shapiro, L., Bell, K., Dhas, K., Branson, T., Louw, G., & Keenan, I. D. (2020). Focused multisensory anatomy observation and drawing for enhancing social learning and three-dimensional spatial understanding. *Anatomical Sciences Education*, 13(4), 488–503. <https://doi.org/10.1002/ase.1929>
- Shin, D. S., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing*, 47, 100714. <https://doi.org/10.1016/j.jslw.2020.100714>
- Smith, B. E., Pacheco, M. B., & de Almeida, C. R. (2017). Multimodal codemeshing: Bilingual adolescents' processes composing across modes and languages. *Journal of Second Language Writing*, 36, 6–22. <https://doi.org/10.1016/j.jslw.2017.04.001>
- Stein, P. (2008). *Multimodal pedagogies in diverse classrooms: Representation, rights and resources*. Routledge. <https://doi.org/10.4324/9780203935804>
- Stein, P., & Newfield, D. (2006). Multiliteracies and multimodality in English in education in Africa: Mapping the terrain. *English Studies in Africa*, 49(1), 1–21. <https://doi.org/10.1080/00138390608691341>
- Sweller, J. (2022). The role of evolutionary psychology in our understanding of human cognition: Consequences for cognitive load theory and instructional procedures. *Educational Psychology Review*, 34(4), 2229–2241. <https://doi.org/10.1007/s10648-021-09647-0>
- Tan, L., Zammit, K., D'warte, J., & Gearside, A. (2020). Assessing multimodal literacies in practice: A critical review of its implementations in educational settings. *Language and Education*, 34(2), 97–114. <https://doi.org/10.1080/09500782.2019.1708926>
- Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8(2), 219–235. <https://doi.org/10.1075/gest.8.2.06tel>

- Toohey, K. (2019). The onto-epistemologies of new materialism: Implications for applied linguistics pedagogies and research. *Applied Linguistics*, 40(6), 937–956. <https://doi.org/10.1093/applin/amy046>
- Toohey, K., Dagenais, D., Fodor, A., Hof, L., Nuñez, O., Singh, A., & Schulze, L. (2015). "That sounds so coooool": Entanglements of children, digital tools, and literacy practices. *TESOL Quarterly*, 49(3), 461–485. <https://doi.org/10.1002/tesq.236>
- Tour, E. (2015). Digital mindsets: Teachers' technology use in personal life and teaching. *Language Learning & Technology*, 19(3), 124–139. <https://doi.org/10.64152/10125/44437>
- Unsworth, L. (2014). Multiliteracies and metalanguage: Describing image/text relations as a resource for negotiating multimodal texts. In D. Leu, J. Coiro, M. Knobel, & C. Lankshear (Eds.), *Handbook of research on new literacies* (pp. 377–404). Routledge.
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Kluwer Academic. <https://doi.org/10.1007/1-4020-7912-5>
- Vandommele, G., Van den Branden, K., Van Gorp, K., & De Maeyer, S. (2017). In-school and out-of-school multimodal writing as an L2 writing resource for beginner learners of Dutch. *Journal of Second Language Writing*, 36, 23–36. <https://doi.org/10.1016/j.jslw.2017.05.010>
- Vasudevan, L., DeJaynes, T., & Schmier, S. (2010). Multimodal pedagogies: Playing, teaching and learning with adolescents' digital literacies. In D. Alvermann (Ed.), *Adolescents' online literacies: Connecting classrooms, media, and paradigms* (pp. 5–25). Peter Lang.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy*, 33(3), 211–239. <https://doi.org/10.1007/BF03651836>
- Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Educational Research*, 34(1), 179–225. <https://doi.org/10.3102/0091732X09349791>
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>
- Winston, J. (2012). *Second language learning through drama: Practical techniques and applications*. Routledge. <https://doi.org/10.4324/9780203805114>
- Witte, A. (2026). Embodiment, enaction, and the lived body in foreign language learning: A novel conception of action-oriented language education. *Frontiers in Language Sciences*, 4, Article 1761548. <https://doi.org/10.3389/flang.2025.1761548>
- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024a). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839–1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2024b). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- Yancey, K. B. (2004). Postmodernism, palimpsest, and portfolios: Theoretical issues in the representation of student work. *College Composition and Communication*, 55(4), 738–761. <https://doi.org/10.2307/4140669>
- Yi, Y., & Angay-Crowder, T. (2016). Multimodal pedagogies for teacher education in TESOL. *TESOL Quarterly*, 50(4), 988–998. <https://doi.org/10.1002/tesq.326>
- Yi, Y. (2014). Possibilities and challenges of multimodal literacy practices in teaching and learning English as an additional language. *Language and Linguistics Compass*, 8(4), 158–169. <https://doi.org/10.1111/lnc3.12076>
- Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouweland, K., & Paas, F. (2025). The synergy of embodied cognition and cognitive load theory for optimized learning. *Nature Human Behaviour*, 9, 877–885. <https://doi.org/10.1038/s41562-025-02152-2>