



Unembodied Multimodal Resources in Senior High School EFL Classrooms: A Multimodal Interaction Analysis

Mhd Reza Anwar¹, Tien Rafida², Maslathif Dwi Purnomo³

^{1,2,3}Master's Program in English Education, Faculty of Tarbiyah Science and Teacher Training,
State Islamic University of North Sumatra Medan, Indonesia

Article Info	Abstract
Received: 2026-07-11 Revised: 2026-08-18 Accepted: 2026-08-18 Keywords: <i>unembodied multimodal resources, EFL classroom interaction, multimodal interaction analysis, meaning construction, EFL pedagogy</i> DOI: 10.24256/ideas.v14i1.12243 Corresponding Author: Mhd Reza Anwar reza0333244009@uinsu.ac.id Master's Program in English Education, Faculty of Tarbiyah Science and Teacher Training, State Islamic University of North Sumatra Medan, Indonesia	<i>This study examines how unembodied multimodal resources construct meaning in EFL classroom interaction at the senior high school level. Grounded in Sigrid Norris's Multimodal Interaction Analysis (MIA) framework and social semiotic theory (Kress & van Leeuwen, 2001, 2006), the study investigates the meaning-constructing mechanisms through which non-bodily semiotic resources such as digital video media, projected presentation slides, physical material artifacts, and environmental spatial arrangements operate in authentic EFL instruction. Data were collected through video-recorded classroom observations of an experienced English teacher at SMA Kemala Bhayangkari, North Sumatra, Indonesia, across two classroom meetings covering procedure text and argumentative text genres. Analysis of nineteen video-captured still images reveal that meaning is constructed not through any single resource in isolation but through intermodal ensembles in which projected visual content, spoken teacher language, and embodied action converge. Five key mechanisms of meaning construction are identified: intermodal complementarity, indexical anchoring through gesture, compositional organization of projected layouts, temporal sequencing of resources, and kinesthetic affordances of physical objects. The findings demonstrate that unembodied resources serve multiple integrated pedagogical functions in EFL teaching, supporting content comprehension, enabling genre-level metalinguistic analysis, and structuring classroom participation. These results affirm the fundamentally multimodal character of EFL pedagogy and offer evidence-based implications for EFL instructional design and teacher education.</i>

1. Introduction

The rapid development of digital technology and media has significantly transformed educational practices, including English as a Foreign Language (EFL) teaching and learning (Kress & van Leeuwen, 2001; Jewitt, 2014). Language learning is no longer limited to written and spoken texts; instead, it increasingly involves diverse semiotic resources including images, colors, layout, typography, symbols, and digital media (Mayer, 2009; Kress, 2010). This development reflects a broader paradigm shift in linguistics and discourse studies, one that challenges the assumption that language alone constitutes the primary means of meaning-making and instead positions communication as inherently multimodal.

Within the field of multimodal studies, communicative modes are classified as either embodied or unembodied (Norris, 2004, 2011). Embodied modes involve bodily actions such as gestures, gaze, posture, and facial expressions, while unembodied modes encompass non-bodily semiotic resources such as written texts, projected visuals, material artifacts, layout, and spatial arrangements. In EFL classroom contexts, both categories of modes work together to construct meaning, though existing scholarship has predominantly examined embodied modes, leaving unembodied resources underexplored as a distinct analytical category.

The focus on embodied interaction in multimodal classroom research is well-established. Studies such as those by Goodwin (2000), Norris (2004, 2011), and Qin et al. (2021) demonstrate how gestures, gaze, and posture function as integral meaning-making resources in instructional interaction. Research on EFL classrooms specifically has confirmed that teachers routinely deploy embodied modes to scaffold comprehension, manage participation, and regulate discourse (Lim et al., 2022). However, the non-bodily semiotic environment of the EFL classroom, encompassing the visual displays, technological infrastructure, physical objects, and spatial arrangements that frame every instructional episode, has received considerably less sustained analytical attention.

This gap in the literature is significant for several reasons. First, unembodied resources constitute the persistent material and visual substrate of classroom interaction; they exist before the teacher speaks and remain after speech has ceased. Second, projected slides, videos, and physical artifacts are widely deployed in contemporary EFL instruction, particularly in digitally resourced senior high school contexts. Third, understanding how these resources construct meaning has direct implications for the design of EFL instructional materials and the preparation of EFL teachers. Addressing this research gap, the present study asks: How do unembodied multimodal resources construct meaning in EFL classroom interaction at the senior high school level?

This single, focused research question is investigated through Norris's (2004, 2011) Multimodal Interaction Analysis framework, applied to authentic EFL classroom data collected at SMA Kemala Bhayangkari, North Sumatra, Indonesia. The study contributes to the multimodal analysis of EFL classroom interaction by providing a typologically grounded and mechanism-focused account of how unembodied semiotic resources function as meaning-making tools in senior high school English instruction.

Literature Review

Multimodality is an area of inquiry that conceptualizes communication as the integration and orchestration of multiple modes, each carrying distinct affordances and meaning potentials (Jewitt, 2009; Kress, 2010). Rooted in Halliday's (1978) systemic functional theory, multimodal scholars argue that ideational, interpersonal, and textual metafunctions are realized not only through language but also through visual, gestural, spatial, and material semiotic resources (Kress & van Leeuwen, 2006). A mode is defined

as a socially and culturally organized set of semiotic resources used for meaning making; language, images, gesture, gaze, and spatial arrangement all function as modes through recognizable structures and conventions.

Kress and van Leeuwen's (2001, 2006) social semiotic framework provides tools for analyzing the three dimensions of meaning carried by visual and multimodal resources: representational meaning (what is depicted or described), interactional meaning (the relationship constructed between resource producer and receiver), and compositional meaning (how elements are organized within a visual field). These analytical categories are applied in the present study to examine how unembodied multimodal resources operate in EFL classroom discourse.

Sigrid Norris's Multimodal Interaction Analysis (MIA) provides the primary theoretical and methodological framework for this study. Norris (2004) builds on social semiotics and mediated discourse theory to argue that communication is inherently multimodal and that interaction is constituted through the coordinated deployment of multiple semiotic modes, including gesture, gaze, posture, proxemics, movement, material objects, and spatial arrangements. A central analytical concept is modal density, which describes both the number of modes active at any moment (modal complexity) and the intensity or foregrounding of any particular mode (modal strength).

MIA further distinguishes between higher-level actions (broader social goals such as explaining a concept) composed of lower-level actions (moment-by-moment behaviors such as pointing at a board), and between embodied modes (produced through bodily action) and unembodied modes (embedded in material, visual, and environmental resources). Norris (2011) argues that physical objects and spatial configurations within the interactional environment are active participants in meaning-making, not merely background context. This theoretical position is central to the present study's analytical approach.

Unembodied communicative modes refer to non-bodily semiotic resources that are present in an interaction and contribute to meaning-making. These include written texts, images, visual displays, layout, objects, and spatial arrangements used to support communication (Norris, 2004; Jewitt, 2014). In educational settings, these modes organize information, guide attention, and support learners' understanding during classroom interaction.

Despite their ubiquity, unembodied modes have rarely been analyzed as a distinct analytical category in EFL classroom research. Wang (2016) examined communicative modes in instructional interaction and found that meaning was constructed through multiple modes including visual resources, but embodied modes remained the primary focus. Lim et al. (2022) studied how Singapore EFL teachers deploy multiple semiotic resources including visual displays, confirming that multimodal competence is integral to effective language teaching. Qin et al. (2021) analyzed the multimodal resources used by Chinese EFL teachers in lesson lead-ins, demonstrating how video and gesture support student engagement. However, none of these studies foreground unembodied resources as a primary analytical category or systematically examine the mechanisms through which they construct meaning. This gap provides the rationale for the present study.

2. Method

This study adopts a qualitative interpretive research design, employing Multimodal Interaction Analysis (Norris, 2004, 2011) as the primary analytical framework. Qualitative design is appropriate because the study seeks to understand the meaning-constructing mechanisms through which unembodied resources function in naturally occurring

classroom interaction, a question that requires in-depth, context-sensitive analysis rather than statistical measurement.

Data were collected from video recordings of EFL classroom sessions conducted in one eleventh-grade class at SMA Kemala Bhayangkari, North Sumatra, Indonesia. The study focuses on a single experienced English teacher who has extensive senior high school EFL teaching experience and consistently employs a range of instructional resources. Two classroom meetings were recorded: Meeting 1, focused on procedure text, and Meeting 2, focused on report and argumentative text. For ethical considerations, the identities of the school, teacher, and students are anonymized. Informed consent was obtained from all parties prior to recording.

Video recordings captured the full duration of each classroom session. The recording device was positioned to capture the teacher's instructional practices, including spoken language, written texts on the whiteboard, projected slides, classroom artifacts, and digital media use. Following recording, nineteen still images were identified at specific timestamps as analytically significant moments in which unembodied multimodal resources were prominently deployed. These images were selected for their relevance to the research question, and their illustration of the range of resource types and meaning-constructing mechanisms present in the data.

Analysis started through three iterative stages. First, repeated viewing of the recordings enabled identification of significant multimodal episodes, and the resources present within them. Second, identified images were systematically analyzed according to MIA categories: types of modes present, modal density and complexity, foreground-background organization, and the relationship between unembodied and embodied modes. Third, meaning-constructing mechanisms were interpreted in relation to Kress and van Leeuwen's (2006) representational, interactional, and compositional meaning dimensions. Trustworthiness was ensured through prolonged engagement with data, triangulation across verbal transcripts and visual documentation, and systematic audit trail maintenance following Lincoln and Guba's (1985) credibility, transferability, dependability, and confirmability criteria.

3. Result

The analysis of nineteen classroom interaction images across two EFL lessons reveals that unembodied multimodal resources construct meaning through five distinct but interconnected mechanisms. These mechanisms are examined below across the three metafunctional dimensions of representational, interactional, and compositional meaning.

Types of Unembodied Multimodal Resources Identified

Before examining meaning-constructing mechanisms, it is necessary to establish the range of unembodied resources present in the data. Four major categories were identified: (1) digital video media, comprising tutorial and documentary videos projected to the class; (2) projected presentation slides, including PowerPoint-style textual, structural, and comparative slides displayed via ceiling-mounted projector; (3) physical material artifacts, specifically a real clothing item used as realia in a procedure text lesson; and (4) environmental spatial and technological resources, encompassing classroom layout, furniture arrangement, and technological infrastructure. This typology extends Norris's (2004, 2011) general references to written texts, images, and spatial arrangements by specifying sub-types that carry distinct semiotic affordances and serve different pedagogical purposes.

A significant distinction within the category of projected visual resources lies between digital video and PowerPoint slides. Video, as observed in Meeting 1 (timestamp 2:46) and Meeting 2 (timestamp 12:04), deploys temporal sequencing, motion, and visual narrative to construct meaning about processes and events. Projected slides deploy spatial layout, typography, and comparative structure to present declarative knowledge in analytically organized form. These two sub-types of projected media serve different pedagogical purposes and construct meaning through different mechanisms, confirming Bezemer and Kress's (2016) argument that the materiality of a resource shapes its meaning potential.

Intermodal Complementarity

The most pervasive mechanism of meaning construction across both meetings is intermodal complementarity: the mutual reinforcement of meaning between unembodied visual resources and spoken teacher language. This mechanism is most clearly exemplified at Meeting 1, timestamp 4:19, where the teacher's instruction "Read this carefully" assumes the presence of written text on the projected screen and can only make instructional sense in combination with that visual resource. Neither the spoken instruction nor the projected text by itself conveys the full instructional meaning; meaning is generated at the interface between the two modes.

Figure. 1



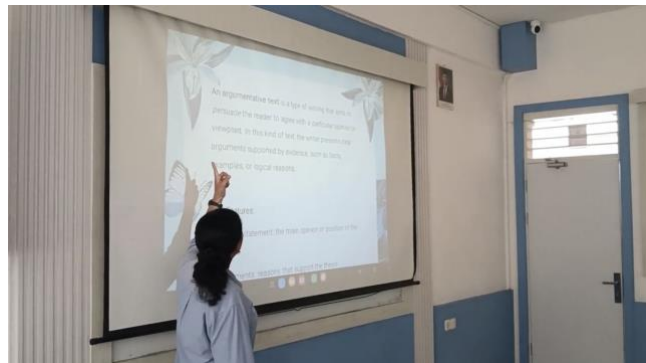
Similarly, at Meeting potential. Intermodalthe directive "You pay attention to the first paragraph" selects a specific portion of the projected text passage and assigns it informational prominence. The projected written passage (unembodied) provides the stable visual reference, while spoken language (verbal mode) navigates the student's attention within it. This pattern of intermodal complementarity, in which unembodied visual resources and speech co-construct meaning by each providing what the other cannot, represents the foundational mechanism through which instruction is achieved in the observed EFL classroom. The finding aligns with Norris's (2004) theoretical claim that meaning in interaction cannot be attributed to any single mode in isolation.

Indexical Anchoring Through Gesture

A second mechanism involves the use of embodied pointing gesture to anchor spoken discourse within specific elements of an unembodied visual resource. This mechanism is most dramatically illustrated at Meeting 2, timestamp 28:31 (Figure 4.10), where the teacher's extended arm and fully raised index finger point directly at projected definitional text while she explains argumentative text features verbally. The pointing gesture creates

a real-time semiotic bridge connecting the spatial location of the written definition on screen with the ongoing verbal discourse.

Figure. 2

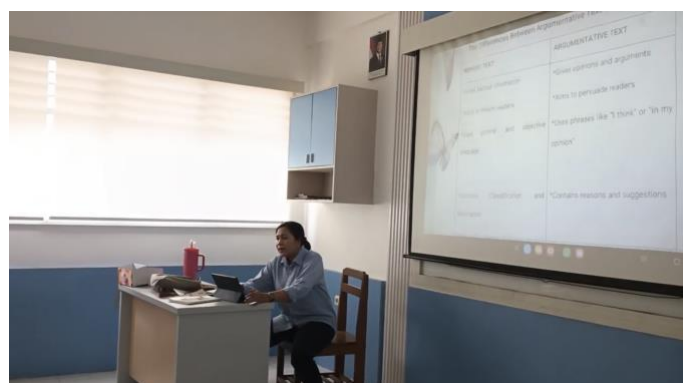


At this moment, the projected written list (unembodied mode), the teacher's pointing gesture (embodied mode), and the spoken utterance "There are five points in argumentative essay" converge to form a highly dense multimodal configuration. This convergence illustrates what McNeill (1992) describes as the integrated gesture-speech system: the gesture and the verbal utterance form a single communicative act, with the gesture establishing the spatial referent and speech providing the linguistic elaboration. Norris (2004) terms this kind of high-density co-deployment of modes a complex modal ensemble and associates it with heightened instructional salience and communicative effectiveness.

Compositional Organization of Projected Layouts

A third mechanism operates at the level of the visual design of projected slides: meaning is constructed through the spatial and typographic organization of information within the slide layout itself. This mechanism is most clearly exemplified in Meeting 2, timestamp 28:07 (Figure 4.9), where the teacher projects a comparison table titled "The Differences Between Report Text and Argumentative Text," organized in two vertical columns. The spatial placement of report text characteristics on the left and argumentative text characteristics on the right enacts a visual grammar of comparison that enables students to perceive genre distinctions simultaneously rather than sequentially.

Figure. 3



Kress (2010) argues that visual modes possess a fundamental affordance unavailable to spoken language: the ability to present multiple elements in simultaneous spatial relationships, enabling comparison in a single perceptual act. The compositional layout of the comparison table exploits this affordance, making the structural contrast between two text types visually available in a way that verbal explanation alone cannot achieve. Kress and van Leeuwen's (2006) concept of information value is relevant here: the given-new organization of the two-column layout, with one text type on each side, activates a compositional logic of contrast that structures the students' reading of the slide. The teacher's subsequent instruction "Take a note okay" further positions the slide as a knowledge resource to be internalized.

Temporal Sequencing of Resource Types

A fourth mechanism of meaning construction operates across time rather than at individual moments: the sequential deployment of different types of unembodied resources across the instructional arc of a lesson creates a scaffolded progression of meaning that moves from iconic representation through symbolic analysis to material enactment. This mechanism is most clearly visible in Meeting 1, which proceeds from digital video (timestamp 2:46) to projected slides (timestamps 4:12–8:03) to physical realia (timestamps 6:36–7:18).

The tutorial video introduces the procedure text concept through visual demonstration, allowing students to observe the complete process before engaging with its linguistic formalization. The projected slides then transform this visual experience into textual and grammatical knowledge, presenting the structural and linguistic features of procedure texts in organized visual-written form. Finally, the physical t-shirt provides the material instantiation of the procedure, enabling students to enact the steps they have watched and analyzed. This progression from video (iconic) to slide (symbolic) to realia (enactive) mirrors the representational mode continuum theorized by Kress (2010) and implements a multimodal scaffolding design that ensures multiple entry points for comprehension. The temporal sequencing of resource types of thus functions as a macro-level meaning construction mechanism that organizes the lesson as a coherent instructional progression.

Kinesthetic Affordances of Physical Objects

The fifth mechanism is specific to physical material artifacts and involves the construction of meaning through tactile, three-dimensional, and kinesthetic interaction with a real-world object. At Meeting 1, timestamp 7:18, the teacher and a student volunteer jointly hold and shake a t-shirt in preparation for the folding procedure, with the teacher instructing "Yeah, shake it, like that." The physical object in this moment becomes a kinetic resource whose movement, flexibility, and material responsiveness construct procedural meaning through embodied action.

Figure. 4



This mechanism is distinct from the others because meaning here is not constructed representationally but enactively. The t-shirt does not stand for or illustrate the procedure; it is the material reality around which the procedure is organized. Norris (2011) notes that material objects in the interactional environment co-form the meaning-making process, and Bezemer and Kress (2016) describe this as the affordance specificity of modes: each mode offers unique possibilities for meaning-making that cannot be replicated by other modes. The kinesthetic affordances of the physical realia the feel of the fabric, the weight of the garment, the spatial experience of folding provide channels for procedural meaning-making unavailable through the video tutorial or projected slides that preceded this activity.

4. Discussion

The five mechanisms identified in the findings collectively demonstrate that unembodied multimodal resources are not supplementary aids to verbal instruction but constitute essential meaning-making tools through which the instructional goals of EFL lessons are organized, scaffolded, and achieved. These finding advances previous scholarship in several important ways.

First, the typological differentiation of unembodied resources into four distinct categories (digital video, projected slides, physical realia, and spatial-technological arrangements) addresses a limitation in Norris's (2004, 2011) original framework, which references unembodied modes in broad categorical terms. The present analysis demonstrates that each category carries unique semiotic affordances that produce different mechanisms of meaning construction; digital video constructs meaning through temporal sequencing, projected slides through compositional organization, physical realia through kinesthetic engagement, and spatial arrangements through participation structure. This granularity is significant for both theory and practice.

Second, the identification of temporal sequencing as a macro-level meaning construction mechanism suggests that multimodal resource orchestration in EFL teaching operates not only at the level of individual moments but across the temporal arc of lessons. This finding extends MIA's predominantly synchronic, moment-by-moment analytical focus and points to the importance of examining how resource deployment is organized across instructional phases as a dimension of multimodal pedagogical design.

Third, the identification of the comparative visual display mechanism, operating through the compositional affordances of projected layout, reveals that unembodied resources can perform specifically metalinguistic functions in EFL instruction that verbal explanation alone cannot achieve. Making the structural differences between text types simultaneously visible through a two-column comparative layout enables a form of genre

awareness that depends on the visual affordances of the projected medium. This finding has direct implications for EFL materials design and pedagogical planning at the senior high school level, where genre knowledge is a central curricular objective.

These findings are broadly consistent with Qin et al.'s (2021) study of Chinese EFL teachers and Lim et al.'s (2022) study of Singaporean EFL teachers, both of which confirm the importance of multimodal resource deployment in EFL instruction. However, the present study's exclusive focus on unembodied resources and its mechanism-centered analysis provide a more detailed account of how non-bodily semiotic resources specifically operate as meaning-making tools, contributing a dimension of analysis absent from previous work.

5. Conclusion

This study investigated the meaning-constructing mechanisms of unembodied multimodal resources in senior high school EFL classroom interaction through Multimodal Interaction Analysis (Norris, 2004, 2011) and social semiotic theory (Kress & van Leeuwen, 2001, 2006). Analyzing nineteen video-captured still images from two EFL classroom meetings, the study identified five mechanisms through which unembodied resources construct meaning: intermodal complementarity, indexical anchoring through gesture, compositional organization of projected layouts, temporal sequencing of resource types, and kinesthetic affordances of physical objects. These mechanisms operate across representational, interactional, and compositional meaning dimensions and are realized through four categories of unembodied resource: digital video media, projected presentation slides, physical material artifacts, and environmental spatial and technological resources.

The findings prove that meaning in EFL classroom interaction arises not from any single mode but from intermodal ensembles in which projected visual content, spoken teacher language, embodied action, and material objects converge. The temporal sequencing mechanism in particular reveals that multimodal resource orchestration in EFL teaching is not only a moment-by-moment phenomenon but a macro-level instructional design decision. The metalinguistic function of comparative slide layouts offers evidence that unembodied resources can make the formal properties of language visible in ways uniquely afforded by visual media.

For EFL practice, these findings support deliberate and principled attention to unembodied resource deployment in lesson planning, materials design, and teacher education. Specifically, EFL teachers are encouraged to sequence resources progressively from iconic to symbolic to enactive forms, design projected slides that exploit visual-compositional affordances for making genre and language contrasts visible and employ physical realia where lesson content involves procedural or material knowledge. Future research should employ continuous video analysis with multimodal transcription to examine real-time co-deployment of modes, integrate analysis of student multimodal responses, and examine the relationship between specific unembodied resource types and measurable language learning outcomes.

6. References

- Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166–195. <https://doi.org/10.1177/0741088307313177>
- Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage Publications.
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489–1522. [https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X)
- Goodwin, C. (2007). Interactive footing. In E. Holt & R. Drew (Eds.), *Reporting talk: Reported speech in interaction* (pp. 16–46). Cambridge University Press.
- Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning*, 56(1), 155–196.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241–267.
- Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.
- Jewitt, C. (2014). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (2nd ed., pp. 15–30). Routledge.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, 69, 101048. <https://doi.org/10.1016/j.linged.2022.101048>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge.
- Norris, S. (2011). *Identity in (inter)action: Introducing multimodal (inter)action analysis*. De Gruyter Mouton.
- Qin, Y., Wang, P., & Zhang, L. J. (2021). Multimodal analysis of EFL teachers' classroom lead-ins in Chinese secondary schools. *Frontiers in Psychology*, 12, 793495. <https://doi.org/10.3389/fpsyg.2021.793495>
- Wang, C. (2016). Communicative modes in instructional interaction: A multimodal analysis. *Journal of Language Teaching and Research*, 7(4), 680–688.
- Walsh, S. (2011). *Exploring classroom discourse: Language in action*. Routledge.