
INTEGRATING GESTURAL AND VISUAL MODES IN CLASSROOM DISCOURSE: A MULTIMODAL ANALYSIS OF TEACHER-STUDENT INTERACTIONS

Zalfa Salsabila Azzahra, Risma Julistiana, Sundari Purwaningsih

Universitas Perjuangan Tasikmalaya

zalfasalsa0911@gmail.com

Abstract

Classroom interaction involves the exchange of meaning through various communication channels beyond spoken language. This study explores how English teachers integrate gestural and visual modes in classroom discourse to enhance student engagement and understanding at SMK Bhakti Kencana Tasikmalaya. Using a qualitative design, data were collected through classroom observation, video documentation, and semi-structured interviews with six students. The analysis applied the Multimodal Discourse Analysis Approach, focusing on the identification and classification of gestures and visual elements utilized by teachers in their instructional interactions. The findings reveal that teacher strategically combine deictic gestures and visual features such as the use of color, typographic emphasis, and layout to enhance the salience and comprehension of key instructional points. Teacher deployment of beat gestures, combined with structured visual arrangements, effectively organizes learning sequences and facilitates clearer task segmentation. The study concludes that the integration of gestural and visual modes significantly enhances meaning-making and student engagement by creating coherence between verbal explanations and visual representations. These multimodal strategies not only clarify learning content but also increase student motivation, attention, and understanding, making classroom interactions more inclusive and effective.

Keywords: Multimodal discourse analysis; gestural mode; visual mode; classroom interaction; EFL teaching

INTRODUCTION

Classroom interaction involves the continuous exchange of meaning between teachers and students through multiple modes of communication. In addition to spoken language, teachers frequently employ non-verbal resources such as gestures, facial expressions, gaze, and spatial positioning, all of which function as semiotic tools that enhance comprehension and engagement. Previous research demonstrates that these multimodal cues play an essential role in instructional processes; for instance, Pramono (2020) found that teachers strategically utilize space, gestures, and gaze to organize learning and support students' understanding.

Multimodal discourse analysis provides a relevant framework for examining how various semiotic resources work together during classroom communication. According to Jewitt (2009), meaning is constructed through the orchestration of multiple modes, including language, gesture, images, sound, and objects. This perspective highlights that teaching is not solely dependent on verbal explanations, but also involves bodily movements, spatial arrangements, and visual

elements that contribute to meaning-making. Semiotic theory further supports this view, with Kress and van Leeuwen (2006) defining modes as culturally shaped sets of resources such as visuals, gestures, and written text used to convey meaning. In classroom contexts, teachers often integrate several modes simultaneously, making it important to analyze how these modes interact in shaping students' understanding.

Both gestural and visual modes have been shown to significantly enhance learning. Jakobson and Axelsson (2021) found that visual representations help students grasp complex concepts by linking them to prior experiences, while gestures reinforce and clarify verbal explanations, particularly in foreign language learning. Lim's (2019) Systemic Functional Multimodal Discourse Analysis (SF-MDA) framework illustrates how gestures operate as meaningful semiotic resources that support ideational, interpersonal, and textual functions in classroom discourse. These perspectives underline the need to investigate how teachers integrate gestural and visual modes in real instructional settings.

This study addresses how an experienced EFL teacher at SMK Bhakti Kencana Tasikmalaya integrates gestural and visual modes during instructional discourse with eleventh-grade students. The research specifically examines: (1) Teacher integrate gestural and visual modes in their interactions with students during classroom discourse (2) Learners perceive the effectiveness of multimodal approaches in enhancing their language acquisition and motivation.

The significance of this investigation extends across theoretical and practical dimensions. Theoretically, it contributes to multimodal discourse analysis by providing detailed empirical evidence of mode integration within an underrepresented educational context. The study responds to calls for context-specific research that examines how universal semiotic principles manifest within particular cultural and institutional settings (Jewitt, 2013). Practically, findings offer insights for EFL teacher education and professional development, highlighting specific multimodal strategies that may enhance instructional clarity and student engagement. By documenting both teaching practices and learner responses, the research provides evidence-based guidance for pedagogical innovation in language classrooms.

METHOD

Research Design

This study employed a qualitative research design to explore how multimodal communication naturally occurs in classroom interaction. As noted by Denzin and Lincoln (2011), qualitative

focuses on understanding social phenomena through in-depth contextual analysis. Using a multimodal analysis approach, the study examined how gestural and visual modes are integrated in teacher student interactions and how these modes support learners' understanding and engagement. This approach aligns with Erickson's (2004) emphasis on capturing the nuances of authentic classroom communication.

Research Context and Participants

The study was conducted at SMK Bhakti Kencana Tasikmalaya, located in Indihiang district, Tasikmalaya city, West Java, Indonesia. The selection of this setting was based on the consideration that the school has supportive learning facilities, as well as a variety of student learning styles that allow for the use of multimodality in teaching. The focus of this research is on an English class taught by an experienced teacher.

Student participants consisted of eleventh-grade students enrolled in the Pharmacy class. From this class, six students were selected through stratified random sampling, ensuring gender balance with three male and three female participants, following Cochran's (1953) principles, this sampling is a sampling method in which the population is divided into several subgroups or strata that are homogeneous based on specific characteristics relevant to the survey's objectives.

Data Collection Procedures

Data were collected through classroom observations and semi-structured interviews. Non-participant observations were conducted to document teacher–student interactions without researcher involvement in the learning process. The classroom activities were video-recorded using a digital camera positioned to clearly capture gestures, facial expressions, and visual materials while ensuring minimal disruption to the lesson. Observations were carried out over two to three sessions to obtain sufficient multimodal data.

Semi-structured interviews were also conducted with the teacher to explore their understanding and use of multimodality during instruction. This interview format allowed flexible

probing into teaching practices and perceptions, consistent with Kvale and Brinkmann's (2009) view that semi-structured interviews help uncover deeper meanings behind participants' experiences.

Data Analysis Procedures

The data in this study were analyzed using Miles and Huberman's (1994) framework, which outlines three interactive stages: data reduction, data display, and conclusion drawing and verification.

Data reduction involved selecting and condensing segments from video observations and interview transcripts that were relevant to the research focus. At this stage, the researcher identified classroom moments that clearly demonstrated the integration of gestural and visual modes such as the teacher's hand movements, gaze direction when giving instructions, and the use of color or images in instructional slides that supported lesson delivery.

Data display consisted of organizing the reduced data into structured formats, including multimodal tables, annotated screenshots, and detailed multimodal transcriptions. These displays represented combinations of communication modes body movements, facial expressions, and visual elements making complex interactional data more coherent and interpretable. Selected interview excerpts were also incorporated to strengthen and contextualize the analysis.

The final stage, conclusion drawing and verification, involved interpreting the patterns identified in the displays to understand how various modes contributed to classroom meaning-making. This process was carried out iteratively and supported through triangulation across observations, interviews, and multimodal transcripts. Additional validation was achieved through member checking and consultation with supervisors or peer researchers. Consistent with Miles and Huberman's (1994) emphasis on plausibility and confirmability, emerging interpretations were repeatedly reviewed against the raw data to ensure their credibility and analytical robustness.

FINDINGS AND DISCUSSION

a. Findings

Integration of Gestural and Visual Modes

Figure 1. Teacher Used Deictic Gesture and Color and Font Visual Types

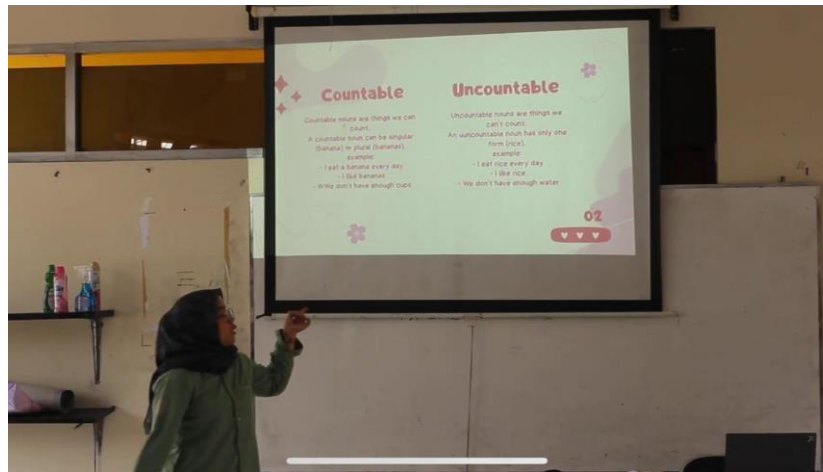


Figure 1 shows that the teacher employed a gestural mode by pointing at the screen with the index finger type a **Deictic Gesture** which was simultaneously integrated with the visual mode, represented by the lesson title “Countable” written in bold pink font type a **Color and Font Visual**. The integration of these two modes effectively strengthened students’ attention toward the main information being conveyed. The pointing gesture served to direct students’ focus to a specific visual element on the screen, namely the lesson title. As stated by McNeill (1992), deictic gestures act as reference markers that help speakers guide listeners’ attention to relevant objects or information. In this context, the teacher’s pointing gesture functioned as a visual cue that bridged the verbal explanation and the displayed visual text, allowing students to more easily grasp the key point of the discussion.

The coordination between the pointing gesture and the visual text “Countable” organized the flow of students’ attention and focused their cognition on the lesson's main topic. The use of bold pink typography provided strong visual emphasis, signaling that the word carried significance in the learning context. Color Pink is also socially perceived as a feminine color because it is often used by women, such as in clothing and toys. This is also related to the gender of the teacher, who is female, so she uses pink in the text of the PowerPoint. This aligns with Kress and van Leeuwen’s

(2006) view that visual elements such as color, font size, and layout possess semiotic functions that contribute to meaning construction.

The meaning of this integration lies in creating a multimodal message in which physical action and visual representation complement one another to convey a unified pedagogical intention. The deictic gesture kinesthetically guides students' attention, while the visual title "Countable" conceptually reinforces the central focus of the lesson. Together, they form an Inter-semiotic meaning, indicating that the teacher is emphasizing the main topic countable nouns and guiding students to focus on it both visually and cognitively. Thus, the integrated use of gesture and visual mode not only signals the lesson topic but also highlights its importance and ensures students' comprehension through simultaneous kinesthetic and visual reinforcement.

Figure 2. Teacher Used Beat Gesture and Font Size Visual Types



Figure 2 shows that the teacher employed a gestural mode in the form of a head nod a type of **Beat gesture** to indicate approval of students' responses or explanations related to the use of how much and how many in expressing quantity. According to Anstey and Bull (2006), multimodal communication such as head nods functions as a gestural resource that conveys agreement, affirmation, and interpersonal support, reinforcing students' understanding while fostering a positive and collaborative classroom atmosphere.

Simultaneously, the visual mode was represented by the text displayed on the screen, particularly the sentence "We use 'how much' and 'how many' to ask about quantities" written in a large, prominent font size, type a **Font Size Visual**. This visual design draws students' attention

and emphasizes the grammatical concept being taught. As explained by Anstey and Bull (2006), visual prominence through size and layout helps to structure meaning and guide students' focus toward essential elements in multimodal communication.

The meaning of this integration lies in the harmonious interaction between bodily movement and visual emphasis, where both modes work together to construct a unified pedagogical message. The teacher's head nod functions as a nonverbal affirmation, signaling to students that their interpretation or response is correct, while the visual text reinforces this approval through its clear and bold presentation of the grammatical rule. Together, these modes form meaning that communicates both cognitive validation and affective support students not only see the rule visually but also feel the teacher's affirmation through gesture.

Figure 3 Teacher Used Deictic Gesture and Color Visual Types.

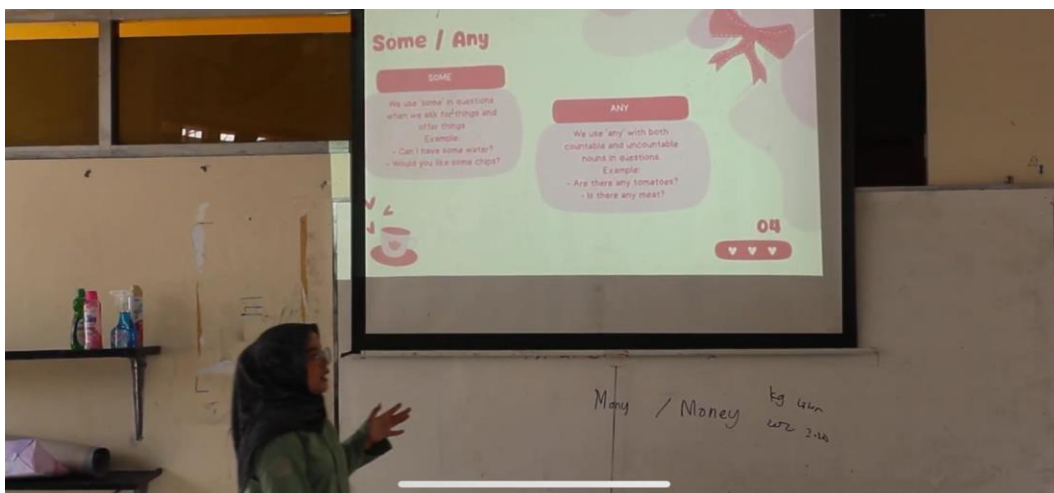


Figure 3 shows that the teacher employed a gestural mode in the form of a pointing hand movement type a **Deictic Gesture** directed toward the projector screen to indicate the word “Any.” This gesture served as a focus marker, guiding students’ attention to a specific part of the lesson. At the same time, the visual mode reinforced the meaning through the presentation of the word “Any” in white font color placed against a pink background (**Color Visual**), white color symbolizes purity and cleanliness because it is essentially a color that appears unstained. It can also create a neutral impression that helps enhance contrast. This contrast highlighted the target word and distinguished it from other textual elements, allowing students to visually perceive its importance. The use of color and placement aligns with Kress and van Leeuwen’s (2006) theory

of visual grammar, which states that visual choices such as contrast, size, and color carry semiotic significance that contributes to meaning construction.

The meaning of this integration lies in the complementary relationship between gesture and visual design, where both modes collaboratively construct a unified pedagogical message. The pointing gesture physically directs and narrows students' visual focus toward the target word, while the visual contrast symbolically emphasizes its grammatical function and relevance. Together, these two modes create connection that communicates both the location (through gesture) and the conceptual importance (through visual emphasis) of the word “Any.”

Figure 4 Teacher used Beat Gesture and Layout, Shape Visual Types

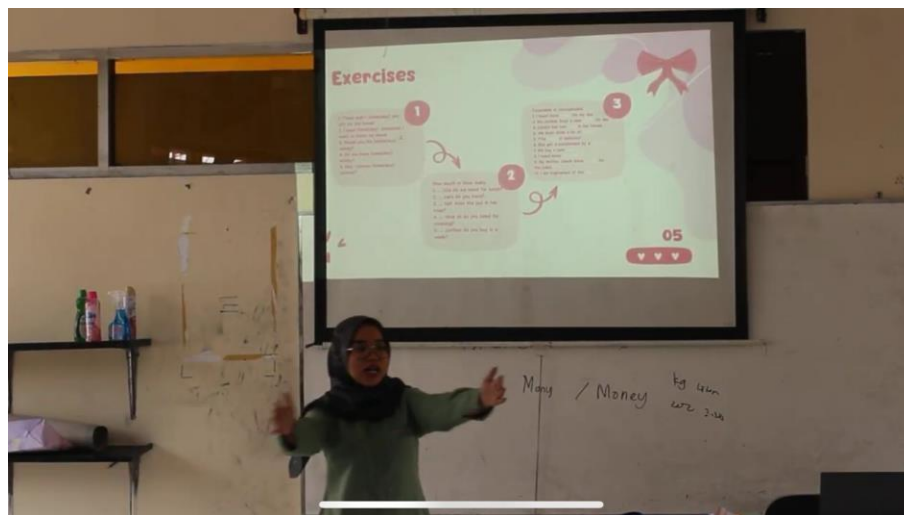


Figure 4 shows that the teacher used a gestural mode through body movement involving changes in position, which were adjusted according to the students' seating arrangement. This movement indicated that the teacher was giving directions and guiding students' attention to each row. Furthermore, the use of open-hand gestures moving to the right and left, categorized as **beat gestures**, functioned as nonverbal signals to indicate the division of tasks or the focus of activities for each group. Meanwhile, the visual mode was represented by the slide presentation displayed on the projector screen, which contained three text sections (**Layout Visual**) consisting of exercises or questions for students to complete. Each section was numbered and connected with directional arrows (**Shape Visual**), showing the sequence of completion from one to three. These visual elements provided a clear structure for the learning activities and helped students understand the stages of the tasks they needed to follow. According to Anstey and Bull (2006), the visual

mode plays an important role in meaning-making by organizing information through elements such as layout, color, direction, and hierarchy, allowing learners to easily comprehend the relationships among different pieces of information.

Student Perceptions of Multimodal Instruction

The interview data indicate that students view multimodal approaches such as gestures, visuals, and interactive activities as highly effective in supporting their English learning and increasing motivation. Students consistently stated that combining multiple modes makes lessons more engaging and easier to understand compared to relying on verbal explanation alone.

Students reported that multimodal input helped them understand abstract concepts, including grammar points like the difference between some and any. As expressed by Student 2:

Excerpt 1

"They help me understand quite a lot. For example, earlier I didn't know the difference between 'some' and 'any,' but with the help of body movements and text, it became clearer."

Another student emphasized the difficulty of understanding without gestures or visuals:

Excerpt 2

"It is very helpful because there are examples on the PowerPoint. If it's only explained verbally, I don't understand as well." (Student 6)

These responses demonstrate that multimodal resources provide concrete visual representations that strengthen comprehension and memory, aligning with Kress & van Leeuwen's (2006) view that meaning is constructed through multiple semiotic modes.

Students also reported increased engagement and motivation when visual and gestural elements were incorporated into the lesson. For example:

Excerpt 3

"I feel more excited because there's something to look at, not just listening, so it's less boring." (Student 3)

Excerpt 4

"I enjoy moving more because there are body movements, and it makes the class more fun." (Student 6)

This supports Anstey & Bull's (2018) framework emphasizing the role of linguistic, visual, audio, gestural, and spatial modes in creating engaging learning experiences.

Most students expressed a clear preference for lessons that combine explanations, visuals, and gestures, stating that multimodal presentation improves clarity and supports learning beyond the classroom:

Excerpt 5

“I prefer using visuals and body movements because if I only read, I don’t really understand. So, with gestures and visuals, I can understand better.” (Student 2)

Excerpt 6

“I prefer learning with body movements and visuals because it’s clearer, and I can also see what the visuals look like, such as pictures that can be photographed to review later.” (Student 3)

Therefore, these findings show that students benefit cognitively and practically from multimodal instruction, which supports both immediate comprehension and independent review.

b. Discussion

The findings show that the teacher’s use of multimodal resources particularly gestural and visual modes significantly enhanced student engagement and supported the construction of meaning during English lessons. The integration of gestures and visual elements illustrates how multiple semiotic systems work together to facilitate comprehension, supporting Anstey and Bull’s (2006, 2018) argument that multimodal communication relies on the interplay of verbal, visual, gestural, spatial, and auditory modes.

A clear example of this integration is seen in Figure 3, where the teacher’s deictic gesture of pointing at the word “Any” directs students’ attention to a specific visual element. As McNeill (1992) explains, deictic gestures serve as reference markers that create shared attentional focus, helping learners connect spoken explanations with visual representations. The visual design white text on a pink background further enhances salience, consistent with Kress and van Leeuwen’s (2006) view that color and contrast shape meaning-making. Together, the gesture and visual emphasis create inter-semiotic coherence (Jewitt, 2009), aligning students’ cognitive focus and strengthening understanding of abstract grammar concepts.

Gestures such as head nods and open-hand movements also contributed to positive interpersonal dynamics. These gestures conveyed affirmation, encouragement, and direction, reinforcing verbal feedback. As McNeill (1992) notes, gestures are integral to cognition and communication, not merely supplementary. This aligns with Anstey and Bull’s (2006) perspective

that gestural modes communicate interpersonal meaning, supporting motivation and active participation.

Visual elements on the PowerPoint slides were equally important in structuring lesson content. Bold pink typography, large font sizes, numbered layouts, and directional arrows highlighted key ideas and clarified task sequences. According to Kress and van Leeuwen (2006), layout, color, and typography function as semiotic resources that guide attention and communicate hierarchy. These design features worked in tandem with the teacher's speech and gestures to form a cohesive multimodal message.

Student interview data further reinforce these findings. Learners reported that gestures and visuals made lessons clearer, more engaging, and easier to follow, particularly when dealing with abstract grammar points such as the distinction between some and any. Students also described multimodal lessons as more motivating and less monotonous, supporting Anstey and Bull's (2018) claim that multimodal environments enhance engagement by activating multiple sensory channels. The synchronization of gestural cues with visual aids allowed students to process information through both visual and kinesthetic modes, improving comprehension and retention.

Overall, the study demonstrates that the strategic integration of gestural and visual modes creates a cohesive multimodal environment that significantly supports students' meaning-making, motivation, and understanding of grammatical concepts.

CONCLUSION

This study concludes that the integration of gestural and visual modes plays a crucial role in enhancing communication and meaning-making in classroom discourse. The findings reveal that teachers strategically employ gestures such as pointing, nodding, and open-hand movements alongside visual elements such as colors, font size, and layout to clarify learning content and attract students' attention. These multimodal combinations create coherence between verbal explanations and visual representations, resulting in more effective knowledge transfer.

From a semiotic perspective, gestures function as dynamic resources that not only support but also complement spoken language by conveying interpersonal and ideational meanings. Meanwhile, visual modes structure the learning content, emphasize key points, and aid in the cognitive organization of information. When these two modes are synchronized, they create an inter-semiotic relationship that strengthens comprehension and engagement.

Students' responses indicate that multimodal teaching increases their motivation, attention, and understanding. Learners perceive that visual and gestural cues make English lessons clearer, more interactive, and less monotonous. The integration of these modes also supports learners with different learning styles, making classroom interactions more inclusive and meaningful.

Hence, this study emphasizes that multimodality is an essential component of effective pedagogy. The use of gestural and visual resources should not be viewed as supplementary, but as integral elements of the teaching process that contribute to a richer and more engaging learning environment.

REFERENCES

- Anstey, M., & Bull, G. (2006). Teaching and learning multiliteracies: Changing times, changing literacies. International Reading Association.
- Anstey, M., & Bull, G. (2018). Foundations of multiliteracies. Routledge.
- Cochran, W. G. (1953). Sampling techniques. John Wiley & Sons.
- Denzin, N. K., & Lincoln, Y. S. (2011). The Sage handbook of qualitative research (4th ed.). Sage.
- Erickson, F. (2004). Talk and social theory: Ecologies of speaking and listening in everyday life. Polity Press.
- Ilmi, N. U., & Dewi, R. (2022). Students' perceptions of multimodal teaching in EFL classrooms. *Journal of Language Teaching and Research*, 13(4), 632–640.
- Jakobson, B., & Axelsson, M. (2021). The role of images and concrete objects in supporting students' conceptual understanding. *International Journal of Science Education*, 43(5), 785–802.
- Jewitt, C. (2009). The Routledge handbook of multimodal analysis. Routledge.
- Jewitt, C. (2013). Multimodal methods for researching digital technologies. In S. Price, C. Jewitt, & B. Brown (Eds.), *The SAGE handbook of digital technology research* (pp. 250–265). Sage.
- Kress, G., & van Leeuwen, T. (2006). Reading images: The grammar of visual design (2nd ed.). Routledge.
- Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing* (2nd ed.). Sage.
- Lim, F. V. (2019). Analysing classroom discourse: A systemic functional multimodal discourse analysis approach. *Social Semiotics*, 29(2), 219–244.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
- Norris, S. (2004). Analyzing multimodal interaction: A methodological framework. Routledge.
- Pramono, A. (2020). Teachers' use of space, gestures, and gaze in EFL classrooms. *Journal of English Education*, 8(2), 112–124.
(Catatan: Jika jurnal berbeda, bisa disesuaikan nanti.)
- Qin, J., & Wang, L. (2021). Multimodal competence in EFL teaching: A classroom-based study. *Linguistics and Education*, 63, 101–115.
- Yana, Y., & Abbas, M. (2022). The impact of visual media on students' conceptual understanding. *Indonesian Journal of Educational Research*, 11(1), 45–55.