SELTICS Journal: Scope of English Language Teaching, Literature and Linguistics
Pendidikan Bahasa Inggris, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muslim Maros
Vol. 9, No. 1, June 2026, p-ISSN: 2623-2642, e-ISSN: 2655-5417

Interactive Meaning in Visual Images: A Multimodal Analysis of the English for Nusantara EFL Textbook

Soni Ardian Fasya[1]*, Arif Suryo Priyatmojo[2]
Universitas Negeri Semarang, Indonesia [1) 2)]
soaniardianfasya@students.unnes.ac.id [1)], arifsuryo.unnes@gmail.com [2)]
**Corresponding author*

ABSTRACT

This qualitative study employs a multimodal discourse analysis framework developed by Kress and van Leeuwen to examine how visual images in the *English for Nusantara Kelas IX* textbook construct interactive meaning for ninth-grade EFL learners in Indonesia. The analysis focuses on dimensions of interactive meaning: social distance, perspective, gaze, and modality. A total of 31 images depicting human participants were analyzed. The findings reveal that most images are represented through medium shots (77.41%) and oblique angles (96.78%), tending to position learners primarily as observers rather than symbolically involved participants. Additionally, the use of simplified and stylized visual modality emphasizes visual simplicity while potentially reducing contextual immersion. These findings suggest that while the textbook's visual design may foreground instructional clarity and content delivery, it may provide fewer visual cues for interpersonal involvement and participatory positioning. The study offers analytically informed recommendations for textbook developers and educators to adopt a more balanced visual grammar that integrates both observational and participatory framing to enhance learner engagement.

Keywords : Multimodal analysis, Interactive Meaning, EFL textbook, Visual Images, Language learning

How to Cite : Fasya, S. A., & Priyatmojo, A. S. (2026). Interactive Meaning in Visual Images: A Multimodal Analysis of the English for Nusantara EFL Textbook. *Seltics Journal: Scope of English Language Teaching Literature and Linguistics*, 9(1), 33–43. https://doi.org/10.46918/seltics.v9i1.3247

Article History : Received: 25-04-2026 | Revised: 05-06-2026 | Accepted: 06-06-2026

INTRODUCTION

Textbooks play an integral in formal education by providing one or all of the organized components of teaching and learning. Textbooks support systematic instructional organization, learning progression, and enable students to learn systematically and sequentially, thereby building on what they have already learned. Ellis (1997) states that "textbooks assist teachers in presenting the content of their lessons in accordance with good pedagogical practice and help with delivering the content in a way that is contextually relevant for the student". Textbooks can also be used by students independently to study the material that they are required to complete (such as for an assignment), to reinforce their learning. The Indonesian government mandates the use of nationally developed textbooks such as *English for Nusantara* in its schools, thus ensuring that English as a Foreign Language (EFL) instruction is provided uniformly in Indonesia's many types of schools. Nationally developed EFL textbooks will also serve as a means to standardize cultural and social representations of culture, ethnicity, etc. (Setyono & Widodo, 2019). Increasingly, the use of visual representation in contemporary EFL textbooks represents a more modern way of helping students understand text,

increasing motivation students to learn, and providing contextual support for language learning. The trend toward inclusion of visual elements signifies a substantial pedagogical shift in the understanding of how learners derive meaning from language learning materials. Specifically, language learning materials are considered to develop shared meaning through both verbal and visual resources, as opposed to solely presenting meaning through written text itself.

Modern textbooks also incorporate a wide range of visual elements such as images/drawings. This incorporation serves both aesthetic and educational functions. According to studies, visual aids may enhance student motivation, assist in accommodating different learning styles, assist learners with difficulties in understanding complex concepts and/or abstract concepts and Wang and Hemchua (2022) state that visual representations can aid with understanding cultures, as well as assist with promoting cultural learning in EFL (English as a Foreign Language) classes. The use of visuals has also been shown to improve memory formation and retention when combined with verbal text (Birdsell, 2017; Triacca, 2017). With the continued development of digital technologies, visual materials are becoming interactive, expanding to the pedagogical potential of visual aids within language learning environments (Bikowski & Casal, 2018).

Textbooks are being progressively integrated with both visual and verbal components; this is indicative of larger societal trends in the area of multimodal learning. According to Lim and Tan-Chia (2023) multimodal learning refers to the utilization of various semiotic modes, such as pictures, text and layout, to create meaning. For example, through the integration of multiple modes within their curricular materials, educators could help their students develop better multimodal literacy skills, which then assist them in interpreting and creating meaning across various modes of representation. With regards to language acquisition, the integration and use of multiple modes of representation is particularly important as not only do words carry meaning but so do the visual design of the page and any interactional cues which are embedded within the pictures on a given page.

While the use of visuals in language textbooks can assist with teaching and learning, visual integration also brings several pedagogical concerns that should be investigated through a systematic approach. One major concern involves the possibility of visual-textual misalignment. When procedural text and corresponding imagery lack functional alignment, students are likely to have difficulty comprehending the material presented in the passage and/or receive ambiguous instructions from the teacher (Mayer, 2020). Similarly, an abundance of images or images that are poorly placed in their sequence may increase the amount of extraneous cognitive load that students experience and affect the manner in which they acquire the target language (Sweller et al., 2019). Another area not addressed in prior research concerns the use of semiotic consistencies in visual framing; when illustrations in a textbook have inconsistent use of perspective, distance, or modality, this will no longer depict the learner as a participant in meaning-making, but as a passive receptor of information (Bateman et al., 2017). Previous studies have evaluated the ways in which images are used functionally and representationally in EFL materials, and the ways in which interactive meanings are constructed through visual grammar have not been examined systematically. This lack of systematic research is especially critical in nationally distributed junior high school textbooks, as visual design choices greatly influence student position, level of engagement with pedagogy, and the pedagogical interaction experience.

Although previous studies have examined cultural representation and the general functions of visuals in EFL textbooks, limited research has specifically investigated how interactive meaning is constructed through visual elements such as social distance, gaze, perspective, and modality. Existing studies tend to focus primarily on representational meaning or ideological issues, while the

interpersonal relationship between images and learners receives less attention. Furthermore, few studies have applied Kress and van Leeuwen's visual grammar framework to Indonesian junior high school EFL textbooks, particularly government approved textbooks used at the national level. This oversight is significant because multimodal textbook analysis must interrogate not only what images depict, but how visual grammar constructs learner subject positions and symbolic access to knowledge (Weninger, 2021).

Building on these instructional considerations, the present study addresses a specific gap: how interactive meaning is constructed through visual elements such as distance, angle, and perspective in Indonesian junior high school EFL textbooks. By using Kress and van Leeuwen's framework for a multimodal analysis, the study will analyze how visual features (e.g., distance/camera angle/perspective/visual mode) provide learners with a context for positioning them and creating interaction/engagement with the text book and the activities in the text book within the context of the learners' cognitive engagement and participation with language learning. Therefore, this study contributes to multimodal EFL textbook research by applying an interactive meaning framework to *English for Nusantara Kelas IX* by Damayanti et al. (2022), a nationally distributed Indonesian EFL textbook that has received limited scholarly attention. The research question for this study is: what are the types of interactive meanings that are represented in 'English for Nusantara Kelas IX?

The objective of this research lies in its specific focus on interactive meaning in visuals within learners' EFL textbook, an area that has received limited scholarly attention. This study examines the interaction between visual and verbal modes in language learning materials.

METHODS

This study employed qualitative content analysis informed by multimodal discourse analysis. Descriptive frequency counts and percentages were used to summarize the distribution of visual categories identified in the corpus, while interpretation remained qualitative in nature. Its primary focus is to analyze the content of various visual elements found in the textbook titled *English for Nusantara Kelas IX*. Rather than investigating lived experiences or social behaviors, the analysis focuses on the semiotic resources embedded within instructional visuals, specifically how social distance, gaze, perspective, and modality position learners in relation to represented content. Guided by Kress and van Leeuwen (2020) visual grammar framework, the study treats textbook images as structured communicative artifacts that mediate pedagogical interaction and shape symbolic learner engagement.

Data Collection Procedure

Data were collected through documentation. The primary data source was the textbook *English for Nusantara Kelas IX*, published by the Indonesian Ministry of Education. Documentation was selected because it enables systematic examination of visual and textual materials within their instructional context (Lüpke, 2010). Using this approach, the researcher identified and collected visual images relevant to the study's focus on interactive meaning. Only images containing human participants were selected for further analysis because they allowed examination of interpersonal relations between the represented participants and viewers. The collected images were then categorized according to the analytical framework proposed by Kress and van Leeuwen.

The unit of analysis in this study was the individual visual image containing human participants. In cases where a textbook illustration consisted of multiple panels that formed a single instructional sequence on the same page, the illustration was treated as one analytical unit because the panels collectively conveyed a unified communicative purpose. Coding was based on the dominant visual

characteristics of the overall illustration rather than on each panel separately. This procedure prevented duplication of data and ensured consistency across the corpus.

Data Analysis

Multimodal discourse analysis provides a rigorous methodological lens for examining how semiotic resources co-deploy to construct pedagogical relationships, making it particularly suited to textbook visual analysis (O'Halloran, 2021). Guided by this approach, *t*his study employs the multimodal discourse analysis framework developed by Kress and van Leeuwen to examine interactive meaning in visual images. Interactive meaning concerns the interpersonal relationship constructed between represented participants and viewers through visual design. The analysis focused on four dimensions of interactive meaning: social distance, gaze, perspective, and modality. These dimensions were examined using Kress and van Leeuwen's (2020) visual grammar framework to investigate how textbook images construct interpersonal relationships between represented participants and viewers.

A coding scheme was developed based on Kress and van Leeuwen's (2020) visual grammar framework. Each image was examined and assigned to categories representing four dimensions of interactive meaning. Social distance was coded as close shot, medium shot, or long shot according to the visible proximity between represented participants and viewers. Gaze was coded as demand when represented participants looked directly toward the viewer and as offer when no direct eye contact was established. Perspective was coded as frontal when participants were visually oriented toward the viewer and oblique when they were shown from the side or at an angle. Modality was analyzed through visual realism indicators, particularly color saturation, background detail, and degree of stylization. Images were examined to identify the presence of these modality markers based on Kress and van Leeuwen's (2020) visual grammar framework. Coding was conducted manually through repeated examination of each image, and the dominant category for each analytical dimension was recorded.

This multimodal framework aligns with contemporary understandings that language acquisition unfolds within inherently multimodal contexts, where communication is expressed and perceived through diverse channels embedded in social interactions according to Karadöller et al. (2025) and critical discourse analysis principles that treat visual design as a socially situated meaning-making practice, not merely a decorative element (Machin & Mayr, 2023).

Data analysis followed a three-phase qualitative procedure (Creswell & Poth, 2018): (1) data reduction, filtering the visual corpus to images depicting human participants to enable interpersonal framing analysis; (2) data display, organizing coded features into an analytical matrix for pattern comparison across distance, perspective, gaze, and modality; and (3) verification, cross-checking interpretations against the visual grammar framework. This screening, eligibility, inclusion sequence is summarized in Figure 1.

To enhance analytical reliability, the coding process was conducted through repeated examination of the visual corpus using a predefined coding scheme derived from Kress and van Leeuwen's framework. Initial coding decisions were reviewed and cross-checked against the theoretical definitions of each category. Ambiguous cases were re-examined to ensure consistency between visual evidence and category assignment. This validation procedure helped reduce subjective interpretation and improve coding consistency throughout the analysis.

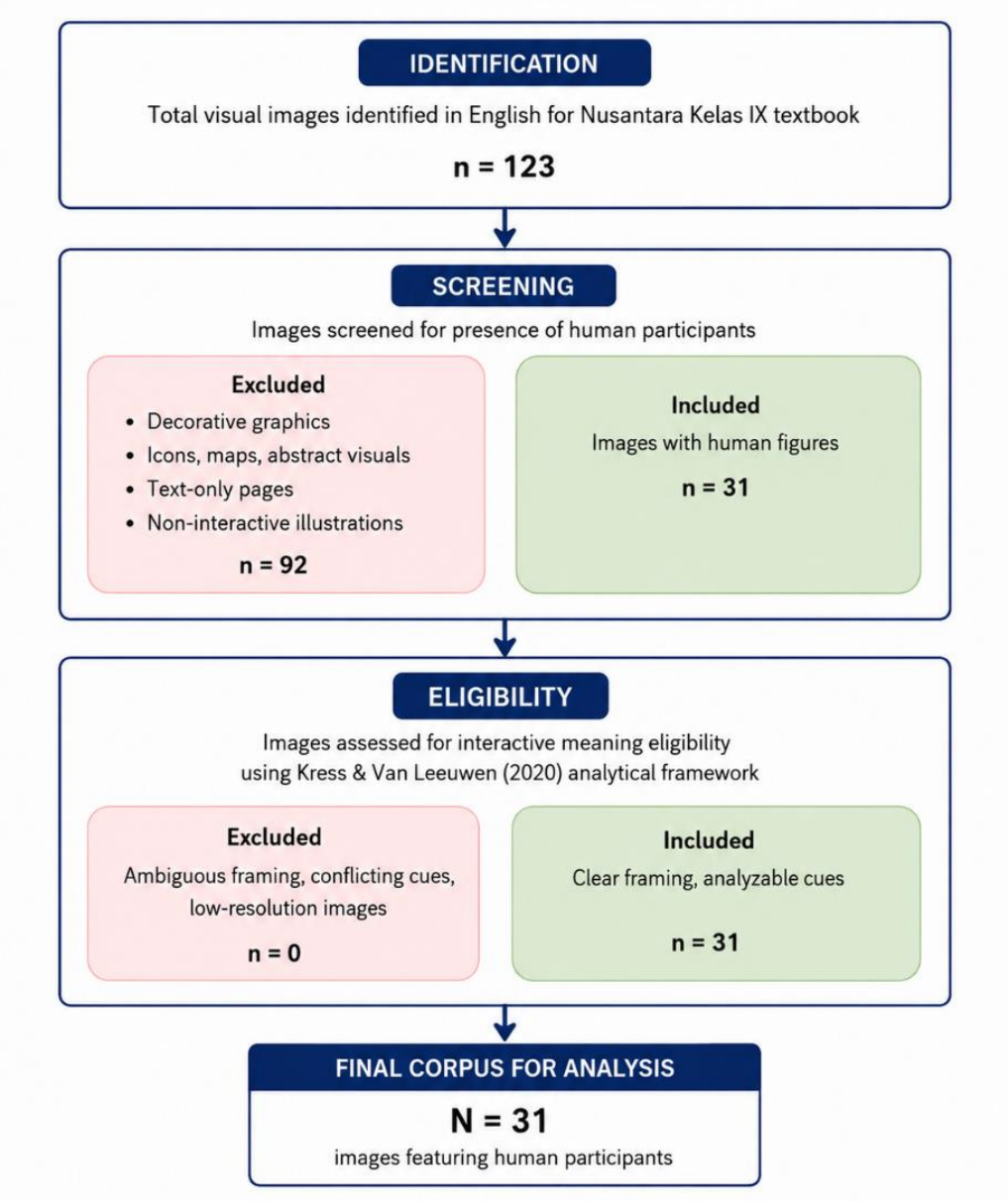

Figure 1. Image Selection Process

Figure 1 presents the process used to select visual data for analysis. From 123 visual images identified in the textbook, only 31 images containing human participants were included because they were considered relevant to interactive meaning analysis

RESULTS AND DISCUSSIONS

Results

Table 1 summarizes the quantitative distribution of interactive meaning categories identified in the visual corpus. Guided by the study's analytical framework, interactive meaning was analyzed through social distance, perspective, gaze, and modality.

Table 1. Results of Interactive Meaning Analysis

Interactive meaning	Category	N	%
Distance	Long-shot	5	16.12
	Medium- Shot	24	77.41

Interactive meaning	Category	N	%
	Close-shot	2	6.45
Perspective	Frontal	1	3.22
	Oblique	30	96.78
Gaze	Offer	30	96.78
	Demand	1	3.22
Modality Indicators	Color saturation present	31	100
	Simplified Background Detail	31	100

Discussion

Social Distance is the distance between how a viewer perceives their physical proximity to participants being represented and how they feel about those participants. The kind of shot utilized in a scene usually indicates the amount of social distance that exists within that scene. Close-up shots (heads/shoulders) create an emotional connection and help develop more emotion and provide better direct interaction for the audience. This emotional proximity increases viewer engagement with the represented action taking place within the scene compared to medium shots (waist/fingers). Medium shots construct a moderate level of intimacy, which exists due to the perception of social closeness created by distance to the participant. This framing creates a sense of relational familiarity without excessive intimacy. Therefore, medium shots establish familiarity while providing an appropriate balance between understanding the learning materials and the ability of the learner to follow while maintaining instructional clarity. Long shots (full body/environment) accentuate the environment in which the participant is situated and create psychological distance by allowing the viewer to observe the participant from a fairly long distance (Machin, 2014).

Attitude/Perspective analysis examines how the spatial relationship between image and viewer are constructed through horizontal and vertical angles. The frontal angle indicates symbolic involvement between viewers and represented participants because the participant is visually oriented toward the viewer. However, interpersonal interaction is more specifically realized through gaze. A direct gaze establishes a demand relationship, whereas the absence of direct gaze creates an offer relationship (Kress & van Leeuwen, 2020), this 'offer' structure positions learners primarily as observers of information rather than co-participants in interaction. Within communicative language teaching frameworks, visual positioning may influence how learners are symbolically oriented toward classroom interaction, particularly whether learning is represented as participatory or transmission-based.

In this study, modality was examined through visual indicators including color saturation, background detail, and stylization. All analyzed images employed saturated colors and simplified background details, suggesting a stylized visual design that prioritizes instructional clarity. Such visual simplification may foreground informational presentation by directing attention toward instructional content and reducing extraneous visual detail (Zhu & Yang, 2023). For young learners, visual modality choices carry heightened significance, as stylized representations can either scaffold identity exploration or inadvertently limit affective connection to target-language contexts (Stec, 2019). Although the use of stylized modality increases access to the learner, excessive stylized modality may also reduce the amount of contextual richness that is needed for pragmatic use of the language.

The final analytical corpus comprised 31 images featuring human participants, as summarized in Table 1.

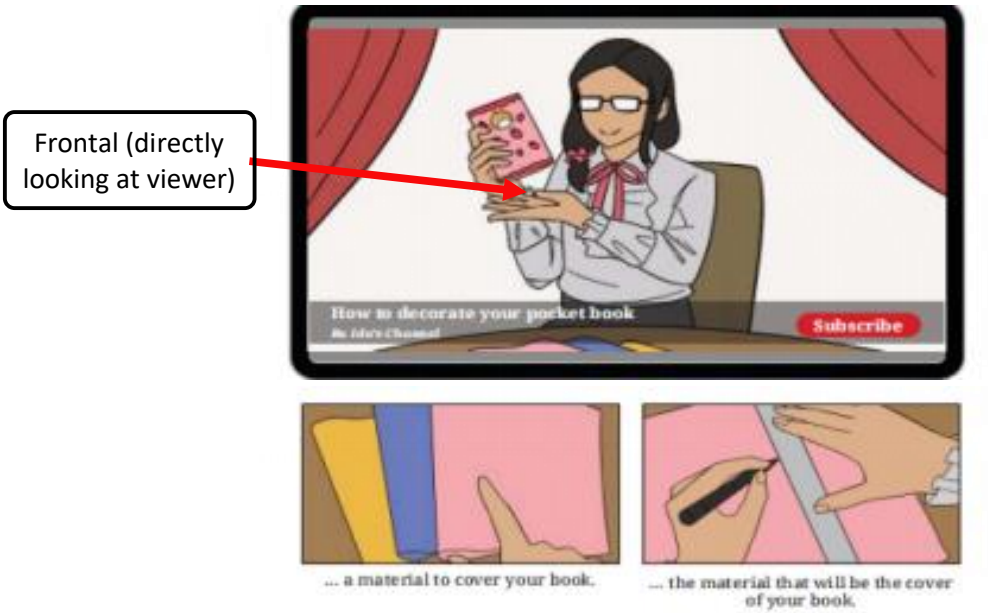

Figure 2. An Image Illustrating Participator Position in EFL Textbook Visuals

The image employs a frontal perspective combined with a demand gaze, creating a stronger interpersonal connection between represented participant and viewer. The central person is positioned to face the camera directly, creating a frontal horizontal angle. According to Kress and van Leeuwen, frontal positioning signifies involvement and interpersonal alignment with the viewer. The participant's direct gaze functions as a visual imperative that implicitly requests viewer attention and engagement thereby encouraging a stronger sense of interpersonal engagement with the represented activity. The medium shot reduces social distance while maintaining sufficient visual clarity for instructional purposes, maintaining an interpersonal connection while providing clear instructions on instructional procedures. Within the illustration, supplementary close-up panels help guide viewer attention through the instructional sequence. Additionally, the use of stylized modalities (use of bright, saturated colors and a simplified background) minimize contextual noise and allow viewers to focus on the demonstrated activity. This participatory framing contrasts with the oblique perspective illustrated in the following example.

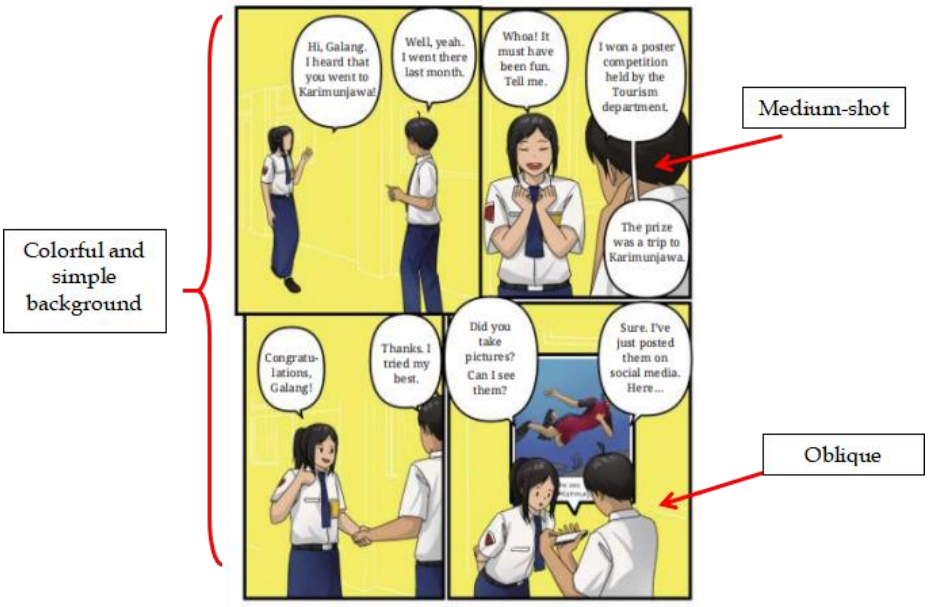

Figure 3. An Image Illustrating Observer Position in EFL Textbook Visuals

This image constructs limited interpersonal engagement through the absence of direct gaze. The represented participants direct their gaze toward one another or objects in the scene, like cellphones. As noted by Kress and van Leeuwen (2020), when a subject's eyes do not face towards the viewer, the visual representation provides an "offer" instead of a "demand" therefore, the visual arrangement frames the learner as an external observer instead of as an active participant. This visual orientation produces a neutral and informational tone commonly associated with educational materials. This interpersonal positioning is realized through the use of specific shot types. The images employ medium-shots framing. This creates the impression that the characters share a relatively familiar social relationship but they are not highly intimate. The participants remain visually accessible through facial expressions and gestures. The medium-shot framing constructs moderate interpersonal distance consistent with everyday classroom interaction. Social distance in this image is significant because it helps the readers understand the relationships between the characters such as interpersonal distance among students. The illustrations employ high color saturation and geometric linearity. The illustrations are moderately stylized rather than fully realistic. They still represent the ongoing interaction and interpersonal dynamics. This illustrative style may foreground content-focused presentation because it is visually accessible and contains limited extraneous visual detail. The image uses this style to direct viewer attention toward the communicative content rather than the visual form itself. The level of realism in the book is appropriately balanced for educational purposes. The image balances simplicity and detail, which may help maintain visual clarity while preserving relevant contextual information.

The results indicate dominant visual patterns in the pictures of the English textbook *English for Nusantara Kelas IX*. Several significant patterns emerged from the analysis. The visuals in *English for Nusantara Kelas IX* can be categorized according to social distance, perspective, gaze, and modality. The analysis revealed that 77.41% of the images employ medium shots, while 16% use long shots. Only two images employ close-up representations (either a close up of their face or one of their upper bodies). The table provides a basis for comparing the different types of images based on distance as illustrated by the images contained in the textbook *English for Nusantara Kelas IX*. The analysis revealed a predominance of the oblique representations, with 30 of the 31 images (96.7%) represented through oblique angles. The gaze analysis further supports this observational pattern. Of the 31 images analyzed, 30 (96.78%) employ offer relations, while only 1 image (3.22%) employ demand relations. The predominance of oblique perspectives, together with the high frequency of offer gaze, may reduce interpersonal involvement between viewers and represented participants. This visual orientation contrasts with findings by Sander et al., (2025), who argue that language acquisition occurs more effectively in multimodal interactions that position learners as active meaning-makers rather than passive recipients. This visual pattern aligns with (Royce & Bowcher, 2013) concept of multimodal communicative competence, which emphasizes that effective second language learning requires learners to simultaneously interpret and negotiate meaning across integrated visual and verbal modes, rather than relying on textual exposition alone.

The findings suggest that the textbook prioritizes informational clarity over interpersonal involvement. The predominance of medium shots, oblique angles, and simplified modality constructs a largely observational viewing position, where learners engage primarily as recipients of visual information rather than as symbolically involved participants. This pattern aligns with broader multimodal scholarship indicating that EFL and multilingual textbooks frequently employ depersonalized imagery to streamline content transmission and reduce cognitive load, often at the expense of immersive interaction (Cremona & Arnaouti, 2019; Jewitt et al., 2025). Similarly, studies of

visual grammar within EFL contexts in Asia have reported that oblique framing and medium-distance shots are commonly employed as techniques for balancing learner accessibility with cultural appropriateness (Qi, 2024). The limited use of frontal perspectives and close shots may reflect a broader tendency in Asian EFL materials to position English as a global commodity to be observed rather than a localized practice to be inhabited (Joo et al., 2020). Collectively, these findings suggest that *English for Nusantara Kelas IX* systematically constructs an observational and information-oriented pedagogical stance that corresponds more closely to traditional, teacher-centred instructional models than to more learner-centred, participatory teaching paradigms. Furthermore, recent visual grammar analysis of Indonesian senior high school EFL textbooks similarly reports a predominance of observational framing, suggesting a systemic pattern across educational levels in nationally mandated materials (Sakdiyah et al., 2026)

Although this design supports structured language exposure and reduces cognitive overload, it is less consistent with contemporary student-centred approaches that position learners as active meaning-makers. Future textbook designers may therefore benefit from adopting a more balanced visual grammar that strategically alternates between observational and participatory framing to support both comprehension and learner engagement. Such visual positioning warrants critical attention, as textbook imagery can semiotically encode power relations that privilege certain learner identities while marginalizing others (Xiang & Yenika-Agbaw, 2021). Integrating critical visual literacy perspectives into multimodal textbook analysis enables educators to examine not only how images construct meaning, but also whose meanings are privileged and how power relations are semiotically encoded (Tran & Lam, 2026). By aligning visual design with modern second language acquisition and multimodal literacy theories, educators and publishers can transform static illustrations into tools that actively scaffold learner interaction.

CONCLUSIONS

The study indicates that the visual design of *English for Nusantara Kelas IX* affords limited opportunities for interpersonal viewer engagement, with interactive meaning largely constrained by text-dominant presentation. Most images employ medium shots (77.41%) and oblique angles (96.78%), and offer gaze relations (96.78%) positioning students as external observers rather than active participants. The consistent use of stylized modality prioritizes curricular alignment and conceptual clarity over affective or experiential connection. The simplified visual design appears consistent with an observational instructional orientation that emphasizes informational presentation and visual clarity over direct interpersonal involvement. A more balanced integration of observational and participatory framing may support stronger learner involvement in multimodal meaning-making.

EFL teachers should use multimodal, learner-centered activities (such as role-play, digital storytelling, or critical visual literacy tasks) to accompany visual representations that are necessarily observational in nature with more participatory visual practices. This enables students to participate in creating meaning from static images to promote interaction between textbook design and classroom instruction. This study proposes a conceptual connection between visual grammar and learner engagement in EFL contexts by drawing on Kress & van Leeuwen (2020) visual grammar framework and contemporary theories of second language acquisition and multimodal literacy. The findings highlight the relationship between visual interactive meaning and learner positioning in EFL materials. The findings suggest that visual elements such as shot type, perspective, and modality may influence how learners are symbolically positioned within EFL instructional materials.

Future research should examine how differences in visual framing among students in the classroom affect learners' motivation to learn, retention of knowledge, and ability to communicate. Additionally, comparative studies using multimodal between different textbook series used at different educational levels and within different national curricula; and further qualitative research may determine how students perceive the impact of stylized versus high modally images on their language learning experience. Additional research should investigate the pedagogical applications of interactive digital and augmented textbooks to inform future multimodal (multiple modes) materials for language learning.

REFERENCES

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis—A problem-oriented introduction*. De Gruyter Mouton. https://doi.org/10.1515/9783110479898

Bikowski, D., & Casal, J. E. (2018). Interactive digital textbooks and engagement: A learning strategies framework. *Language Learning & Technology, 22*(1), 119–136. https://doi.org/10.64152/10125/44584

Birdsell, J. B. (2017). The role of images in ELT textbooks: A case for visual metaphors. *Journal of Liberal Arts Development and Practices, 1*, 9–18.

Cremona, G., & Arnaouti, E. (2019). *Multimodal interpretations of foreign language learning textbooks: Two case studies from Maltese and Greek learning contexts*. Technological Educational Institute of Larissa.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Damayanti, I. L., Febrianti, Y., Suharto, P. P., Fellani, A. J., & Nurlaelawati, I. (2022). *English for Nusantara untuk SMP/MTs kelas IX*. Pusat Perbukuan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal, 51*(1), 36–42. https://doi.org/10.1093/elt/51.1.36

Jewitt, C., Bezemer, J., & O'Halloran, K. (2025). *Introducing multimodality* (2nd ed.). Routledge.

Joo, S. J., Chik, A., & Djonov, E. (2020). The construal of English as a global language in Korean EFL textbooks for primary school children. *Asia-Pacific Journal of Education, 40*(1), 68–84. https://doi.org/10.1080/02188791.2019.1627636

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language, 45*(6), 673–710. https://doi.org/10.1177/01427237241290678

Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Lim, F. V., & Tan-Chia, L. (2023). *Designing learning for multimodal literacy: Teaching viewing and representing*. Routledge.

Lüpke, F. (2010). Research methods in language documentation. *Language Documentation & Description, 7*, 55–104. https://doi.org/10.25894/ldd227

Machin, D. (2014). *Visual communication*. De Gruyter Mouton.

Machin, D., & Mayr, A. (2023). *How to do critical discourse analysis: A multimodal introduction* (3rd ed.). SAGE Publications. https://doi.org/10.4135/9781036212933

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

O'Halloran, K. (2021). Multimodal discourse analysis. In K. Hyland, B. Paltridge, & L. L. C. Wong (Eds.), *The Bloomsbury companion to discourse analysis* (2nd ed.). Bloomsbury Academic.

Qi, J. (2024). Analyzing image meanings in Chinese EFL textbooks: A multimodal perspective. *Theory and Practice in Language Studies, 14*(8), 2498–2509. https://doi.org/10.17507/tpls.1408.23

Royce, T., & Bowcher, W. L. (Eds.). (2013). *Multimodal communicative competence in second language contexts*. Routledge.

Sakdiyah, R., Perangin-angin, A., Zein, T. T., Sinar, T. S., & Rangkuti, R. (2026). Multimodal literacy: A visual grammar analysis of Indonesian EFL textbooks in senior high schools. *Journal of General Education and Humanities, 5*(2), 2615–2628. https://doi.org/10.58421/gehu.v5i2.1223

Sander, J., Zhang, Y., & Rowland, C. F. (2025). Language acquisition occurs in multimodal social interaction: A commentary on Karadöller, Sümer and Özyürek. *First Language, 45*(6), 780–784. https://doi.org/10.1177/01427237251326984

Setyono, B., & Widodo, H. P. (2019). The representation of multicultural values in the Indonesian Ministry of Education and Culture-endorsed EFL textbook: A critical discourse analysis. *Intercultural Education, 30*(4), 383–397. https://doi.org/10.1080/14675986.2019.1548102

Stec, M. (2019). Identity and multimodality of cultural content in ELT coursebooks for YLS. In *European Proceedings of Social and Behavioural Sciences* (Vol. 72, pp. 274–288). Future Academy. https://doi.org/10.15405/epsbs.2019.11.25

Sweller, J., Paas, F., & van Merriënboer, J. J. G. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tran, H. L., & Lam, Q. D. (2026). Multimodal discourse analysis in textbooks: A scoping review and semi-bibliometric synthesis (1995–2025). *VNU Journal of Foreign Studies, 42*(1), 64–80. https://doi.org/10.63023/2525-2445/jfs.ulis.5602

Triacca, S. (2017). Teaching and learning with pictures: The use of photography in primary schools. *Proceedings, 1*(9), Article 952. https://doi.org/10.3390/proceedings1090952

Wang, Y., & Hemchua, S. (2022). Can we learn about culture by EFL textbook images? A semiotic approach perspective. *Language Related Research, 13*(3), 479–499. https://doi.org/10.52547/LRR.13.3.19

Weninger, C. (2021). Multimodality in critical language textbook analysis. *Language, Culture and Curriculum, 34*(2), 133–146. https://doi.org/10.1080/07908318.2020.1797083

Xiang, R., & Yenika-Agbaw, V. (2021). EFL textbooks, culture and power: A critical content analysis of EFL textbooks for ethnic Mongols in China. *Journal of Multilingual and Multicultural Development, 42*(4), 327–341. https://doi.org/10.1080/01434632.2019.1692024

Zhu, Y., & Yang, N. (2023). Visual images and image-text relations in ELT textbooks for young learners. *Language Teaching for Young Learners, 5*(2), 196–216. https://doi.org/10.1075/ltyl.00035.zhu