Journal of General Education and Humanities
Vol. 5, No. 2, April 2026, pp. 2615 – 2628, https://doi.org/10.58421/gehu.v5i2.1223
ISSN 2963-7147

Multimodal Literacy: A Visual Grammar Analysis of Indonesian EFL Textbooks In Senior High Schools

Raihana Sakdiyah[1], Alemina Perangin-angin[2], T. Thyrhaya Zein[3], T. Silvana Sinar[4], Rahmadsyah Rangkuti[5]
[1,2,3,4,5]University of North Sumatra, Medan, Indonesia

Article Info

Article history:

Received 2026-02-22
Revised 2026-03-03
Accepted 2026-03-23

Keywords:

EFL Textbook
Multimodal Literacy
Social Semiotics
Student Interpretation
Visual Grammar

ABSTRACT

The increasing multimodal design of English as a Foreign Language (EFL) textbooks has not been accompanied by sufficient empirical investigation into how visual elements systematically construct meaning and how students interpret these elements in classroom contexts. This study aims to examine how multimodal literacy is realized through visual grammar in the English Grade X textbook used at MAN 2 Deli Serdang, Indonesia, and to explore how students interpret these multimodal features. This research employed a qualitative descriptive design. Data were collected through multimodal content analysis of selected textbook units using Kress and van Leeuwen's (2006) visual grammar framework and semi-structured interviews with ten EFL students. The analysis focused on representational, interpersonal, and compositional meanings constructed through visual, linguistic, and spatial modes. The findings reveal that the textbook systematically constructs meaning through narrative and conceptual visual representations, strategic use of gaze and social distance to build interpersonal relations, and compositional arrangements that highlight information value and salience. Students reported that these multimodal elements facilitated comprehension of abstract concepts, increased engagement, and enhanced perceived relevance of the learning materials. However, variations in interpretation indicate that visual meaning-making is influenced by students' prior knowledge and literacy experience. These results suggest that multimodal design plays a significant pedagogical role in EFL textbook effectiveness and should be purposefully integrated into instructional material development. The study contributes to multimodal literacy research by integrating visual grammar analysis with students' interpretive perspectives in the Indonesian EFL context.

Corresponding Author:

Raihana Sakdiyah
University of North Sumatra, Medan, Indonesia
Email: sakdiyahraihana@gmail.com

1. INTRODUCTION

In the context of English as a Foreign Language (EFL) education in Indonesia, printed textbooks remain the primary instructional resource, particularly at the senior high school and Madrasah Aliyah levels. Although digital materials are increasingly accessible, classroom practices continue to rely heavily on government-approved textbooks, positioning them as central mediators of linguistic input, cultural representation, and pedagogical direction. Research has shown that Indonesian EFL textbooks significantly shape language exposure, task design, and classroom interaction patterns [1]. Therefore, the way meaning is constructed within textbooks directly influences students' learning experiences.

In recent years, EFL textbooks have undergone notable design transformations. Contemporary textbooks no longer depend solely on written language but incorporate images, typography, layout, and color as integral meaning-making resources. This shift reflects broader changes in communication practices, where meaning is constructed through the interaction of multiple semiotic modes [2], [3]. As communication becomes increasingly multimodal, learners are required to interpret visual, linguistic, and spatial elements simultaneously, making multimodal literacy a crucial component of classroom learning [4], [5].

The theoretical foundation of multimodal analysis is grounded in social semiotics, particularly the visual grammar framework proposed by Kress and van Leeuwen [6]. According to this framework, images construct meaning through three metafunctions: representational (depicting participants and processes), interpersonal (establishing relationships between image and viewer), and compositional (organizing visual elements to create salience and information value). From a pedagogical perspective, multimodal elements can enhance comprehension and retention when properly designed, as suggested by the Cognitive Theory of Multimedia Learning and Dual Coding Theory [7], [8]. Thus, images in EFL textbooks should not be viewed as decorative additions but as structured semiotic resources that actively contribute to meaning-making.

Despite the increasing multimodal nature of EFL textbooks, research in Indonesian contexts has tended to emphasize either teachers' perceptions of multimodal literacy [9], [10] or general descriptions of multimodal features in teaching materials [5], [11]. Some recent studies have applied multimodal or social semiotic analysis to textbooks [12], [13], [14], yet these studies primarily focus on textual or visual representation without integrating students' interpretive perspectives. In addition, critical discourse-oriented analyses have examined values and representations in textbooks [15], [16], but they do not specifically investigate how visual grammar constructs meaning at the metafunctional level.

This indicates a significant research gap. While previous studies acknowledge the presence of multimodal elements in Indonesian EFL textbooks, limited empirical research systematically analyzes how representational, interpersonal, and compositional meanings are constructed through visual grammar and how students interpret these meanings in actual classroom settings. Without such analysis, it remains unclear whether multimodal textbook designs effectively support comprehension and engagement or merely function as aesthetic enhancements.

Responding to this problem, the present study investigates the multimodal realization of visual grammar in the English Grade X textbook used at MAN 2 Deli Serdang, North Sumatra. Specifically, this study aims: (1) To analyze how visual, linguistic, and spatial modes construct representational, interpersonal, and compositional meanings using the visual grammar framework. (2) To explore how students interpret these multimodal elements and how such interpretations influence their comprehension, engagement, and perceived relevance of the learning materials.

By combining systematic visual grammar analysis with students' interpretive responses, this study extends previous research that primarily focuses on descriptive or perception-based approaches. The findings are expected to contribute theoretically to multimodal and social semiotic studies in EFL textbook research and practically to inform textbook designers, curriculum developers, and teachers in creating more pedagogically purposeful and critically informed multimodal materials. Ultimately, this research seeks to support the development of multimodal literacy practices that are responsive to the needs of Indonesian senior high school learners.

2. METHOD

This study employed a qualitative descriptive approach, utilizing multimodal content analysis to examine how meaning is constructed and interpreted in an EFL textbook. The qualitative design aligns with interpretive research traditions that emphasize contextualized meaning-making and in-depth analysis [17].

The analysis was grounded in Kress and van Leeuwen's [6] visual grammar framework. The selection of semi-structured interviews as a complementary method is consistent with qualitative research practices aimed at exploring participants' perspectives and experiences. The number of participants (ten students) was considered adequate to achieve thematic saturation, as suggested by Guest, Bunce, and Johnson [18].

The interview questions explored students' interpretations of visual elements and their role in supporting comprehension, drawing on cognitive multimedia learning theory [7] and dual coding theory [8], which posit that learning is enhanced when verbal and visual information are processed through interconnected cognitive systems. Data analysis integrated visual grammar analysis and thematic analysis, followed by triangulation to enhance credibility and analytical rigor.

3. RESULTS AND DISCUSSION

3.1. Multimodal Elements in the Textbook

The textbook integrates photographs, illustrations, layout features, colors, and icons consistently across its pages, forming a cohesive multimodal learning environment. These multimodal components are not randomly distributed; rather, they are deliberately designed to function as meaning-making resources that complement and reinforce linguistic input. In line with multimodal discourse theory, meaning emerges from the orchestration of semiotic modes that work together in patterned and socially shaped ways. Previous studies on EFL textbooks have similarly found that visual and spatial resources play a significant role in structuring pedagogical meaning rather than merely decorating the page.

Photographs and illustrations provide concrete visual representations that help students visualize concepts, situations, and cultural contexts presented in the written text, thereby reducing cognitive load and supporting comprehension, especially for learners with limited language proficiency. From a cognitive perspective, this aligns with Mayer's [7]m multimedia learning theory and Paivio's [8] dual coding theory, which explain that combining verbal and visual input enhances retention and understanding because learners process information through interconnected verbal and non-verbal channels. Furthermore, visual scaffolding has been identified as an important pedagogical strategy in multimodal literacy development, particularly in language classrooms where learners rely on contextual cues to infer meaning [4].

The visuals in the textbook are strategically aligned with themes of each lesson, such as identity, nationalism, interpersonal communication, and social values. For example, in a unit discussing national heroes, a grayscale portrait of Soekarno is presented alongside a descriptive text to reinforce concepts of patriotism and historical identity. Such integration of image and text reflects representational strategies in visual grammar, where images symbolically encode abstract values and social meanings. Similarly, in lessons on technology and modern communication, images of students using smartphones in everyday settings are used to promote relevance and relatability for learners. Thematic alignment between visuals and lesson content supports coherent meaning construction and strengthens contextual understanding [19].

Beyond photographs and illustrations, the textbook employs various graphic features, including arrows, icons (such as light bulbs to indicate ideas), and color-coded sections to highlight key instructions and newly introduced vocabulary. These elements function as strategies of visual salience and framing, directing students' attention to important information and supporting efficient navigation of the text. Research on visual literacy in Indonesian EFL textbooks similarly emphasizes that layout organization, salience, and framing significantly influence how students prioritize and interpret information [13]. From a pedagogical standpoint, such design features align with multimodal reading comprehension principles, which emphasize that layout and visual signaling systems contribute to structured, accessible meaning-making [20].

From a spatial perspective, the layout maintains a clear hierarchy by distinguishing headings, subheadings, images, and instructional content. The effective use of white space helps prevent cognitive overload and enables students to process information more easily. Tasks and explanations are often enclosed in colored boxes, which enhances visual coherence and supports comprehension. These compositional strategies demonstrate how textbooks apply structured visual design to regulate reading paths and guide interpretive sequencing [6]. As noted in multimodal pedagogy research, such visual organization contributes to clearer instructional communication and supports students' independent navigation of texts [21]

The integration of visual and textual modes fosters stronger connections between what students read and what they see by enabling meaning to be constructed through complementary semiotic resources. Visuals serve as pedagogical scaffolding, particularly for learners who struggle to process linguistic input on their own. By providing contextual

cues, visual representations help students infer meaning, anticipate content, and build mental models of the topics discussed. This multimodal support facilitates comprehension and retention, especially in EFL contexts where learners' language proficiency may still be developing. Similar findings have been reported in studies of multimodal literacy practices in Indonesian classrooms, which show that integrated text-image design enhances interpretive engagement and vocabulary development [10].

Multimodal elements also play a significant role in shaping students' emotional engagement with learning materials. Students reported higher levels of interest and motivation when lessons incorporated expressive visuals or culturally familiar figures that reflected their social and educational environment. Such visuals create a sense of relevance and personal connection, making learning experiences more meaningful and less abstract. By acknowledging learners' identities and cultural contexts, multimodal resources contribute to a more inclusive learning environment, thereby enhancing learners' confidence and willingness to participate. This supports arguments in multimodal literacy research that visual design influences not only cognitive comprehension but also affective and interpersonal dimensions of classroom learning [4].

The table below summarizes how different multimodal components appear in the textbook and their primary educational functions:

Table 1. Multimodal Elements and Their Pedagogical Functions in the Analyzed EFL Textbook

Multimodal Element	**Example from Textbook**	**Educational Purpose**
Photographs	Soekarno's portrait; students using smartphones	Reinforce themes of patriotism, technology, and daily life
Illustrations	Cartoon-style drawings in dialogue activities	Add humor, simplify abstract ideas
Layout Design	Boxes, headings, subheadings	Improve text organization, guide reading flow
Icons and Symbols	Light bulbs, arrows, warning signs	Emphasize instructions, signal important ideas
Color Coding	Blue for instructions, orange for vocabulary	Categorize content, increase visual salience
White Space and Framing	Margins around images and text blocks	Reduce clutter, aid navigation
Cultural Figures and Contexts	Local heroes, student uniforms, food icons	Create relevance, connect learning to students' real lives

In summary, the textbook uses multimodal elements systematically and purposefully. Each mode, whether visual, spatial, or symbolic, interacts with the linguistic mode to facilitate interpretation, memory retention, and student engagement. These findings confirm that multimodal design in Indonesian senior high school EFL textbooks operates as a structured semiotic system rather than as decorative supplementation [1]. Consequently, multimodal literacy emerges as a central component of effective EFL pedagogy and textbook development in contemporary classrooms [11].

3.2. Visual grammar analysis

The visual grammar analysis in this study is grounded in Kress and van Leeuwen's [6] framework, which conceptualizes images as structured semiotic systems organized through three metafunctions: representational, interpersonal, and compositional. These metafunctions operate simultaneously, shaping how visual elements construct social reality, position viewers, and regulate interpretive pathways. As emphasized in multimodal theory, visual meaning is never neutral but socially and culturally encoded. Therefore, analyzing textbook images through visual grammar allows a deeper understanding of how pedagogical and ideological meanings are embedded within multimodal design.

1. Representational Meaning

Representational meaning concerns how participants, actions, and concepts are depicted in visual texts. According to Kress and van Leeuwen [6], representational structures are divided into narrative processes (which depict action and unfolding events) and conceptual processes (which present participants in terms of classification, symbolism, or timeless attributes).

Images such as a portrait of Soekarno and students celebrating together convey both narrative and conceptual representations. Narrative visuals depict action and relationships through vectors (such as gaze direction, gestures, or body orientation), while conceptual representations signify abstract ideas like heroism, authority, or identity without showing dynamic action.

Example:

Figure 1. A conceptual representation of heroism: Soekarno's grayscale portrait in the textbook suggests historical reverence.

In Figure 1, the grayscale portrait of President Soekarno functions as a conceptual representation. The image does not depict him performing a specific action; instead, it presents him as a symbolic and authoritative figure. His formal attire, upright posture, and serious facial expression signify leadership, patriotism, and national legacy. The absence of a clear setting or temporal context removes the image from a specific historical moment, underscoring the timeless, institutionalized nature of his role as a national hero.

From a visual grammar perspective, this image exemplifies a symbolic conceptual process in which the participant represents abstract values rather than concrete actions. The grayscale color scheme enhances solemnity and historical distance, reinforcing ideological associations with nationalism and reverence. As noted by Callow [22], visual

design choices, such as color treatment and framing, significantly shape how viewers interpret authority and emotional tone.

Pedagogically, this representation does more than illustrate a historical figure; it encodes moral and civic values aligned with curriculum objectives. Similar findings in Indonesian EFL textbook studies indicate that visual representations often function to transmit cultural and ideological meanings alongside linguistic content [23]. Thus, representational meaning in this textbook serves both as linguistic support and as a value-transmission mechanism.

In contrast, images of students celebrating together reflect narrative processes. The presence of action vectors, such as raised hands, smiling faces, and mutual gaze, creates a dynamic representation of social interaction. These narrative structures depict collaboration and solidarity, visually reinforcing communicative competence and interpersonal skills emphasized in EFL pedagogy.

2. Interpersonal Meaning: Gaze, Distance, and Angle in Visuals Shape How Students Relate to Images.

Interpersonal meaning concerns how images position viewers in relation to represented participants. Through gaze, camera distance, and angle, images establish social relations and power dynamics.

For example, a direct gaze may create a “demand” relationship, inviting interaction, while an indirect gaze creates an “offer,” positioning viewers as observers. Mid-shot, eye-level images typically suggest familiarity and equality between viewer and subject.

Example:

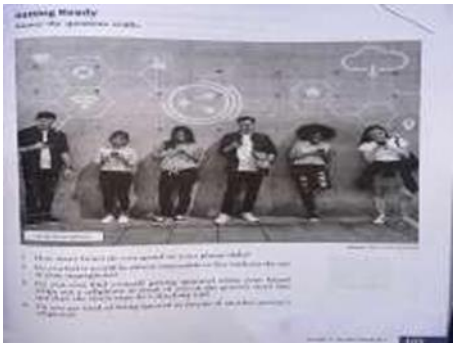

Figure 2. Students using smartphones: an indirect gaze and a side profile invite observation and self-reflection.

In Figure 2, the students are depicted in side profiles and do not establish direct eye contact with the viewer. This indirect gaze constructs an “offer” image, positioning viewers as observers rather than direct interlocutors. The medium shot creates moderate social distance, suggesting familiarity without intimacy. Meanwhile, the eye-level angle indicates equal social power between the viewer and the participants.

This configuration subtly shapes interpretive engagement. Rather than instructing viewers how to feel about smartphone use, the image invites reflection. As Serafini [24] argues, interpersonal positioning in multimodal texts significantly influences reader

interpretation and critical engagement. By avoiding a confrontational direct gaze, the textbook allows space for evaluative reasoning rather than imposing a singular moral stance [15].

Additionally, the close physical distance among the students conveys group cohesion and shared activity, emphasizing the social dimension of technology use. Such representational choices align with communicative language teaching principles, which prioritize interaction and collaborative meaning-making. Research on multimodal literacy in classroom practice similarly highlights that gaze and perspective shape emotional and cognitive involvement in learning [4], [9].

Thus, interpersonal meaning in this textbook is carefully constructed to balance engagement, reflection, and identification, supporting both linguistic comprehension and social awareness.

3. Compositional Meaning

Compositional meaning concerns how visual elements are arranged to create coherence and guide interpretation. According to Kress and van Leeuwen [6], three main principles regulate composition: information value (left/right, top/bottom placement), salience (visual prominence), and framing (connection or separation between elements).

Example:

Figure 3. Humanitarian doctor image: right-side positioning denotes importance and signals new content.

In Figure 3, the image of a humanitarian doctor is positioned on the right side of the layout. Within visual grammar, the right side represents “new information,” while the left side typically represents “given information” [6]. This placement suggests that the doctor’s role introduces a new thematic focus or key learning content.

The doctor is visually salient due to foregrounding, color contrast, and relative size compared to surrounding elements. Salience directs viewer attention and establishes hierarchical importance. The white coat functions as a cultural symbol of medical professionalism and care, while the depicted interaction with patients signifies empathy and service. These symbolic cues enhance the thematic coherence of lessons on social responsibility and humanitarian values.

Framing devices, such as spatial separation between text and image structure, reading flow, and clarification of informational relationships, help clarify informational relationships. Research on multimodal reading comprehension emphasizes that layout

organization significantly affects interpretive sequencing and cognitive processing [20]. Similarly, studies on visual design in textbooks demonstrate that compositional strategies regulate how learners prioritize information [13], [25].

Thus, compositional meaning in this textbook operates as a silent pedagogical organizer. Through placement, salience, and framing, the layout directs students' attention, structures their reading path, and reinforces thematic priorities. The integration of compositional design with representational and interpersonal meanings illustrates how multimodal literacy is embedded within textbook architecture rather than confined to isolated images.

Overall, the visual grammar analysis reveals that the textbook's multimodal design is structured, intentional, and pedagogically meaningful. Representational structures encode ideological and thematic values, interpersonal configurations regulate viewer engagement, and compositional arrangements organize interpretive flow. These metafunctions operate simultaneously, demonstrating that multimodal literacy in Indonesian senior high school EFL textbooks is not incidental but systematically constructed through principles of visual grammar [2], [6].

By analyzing these layers, the study demonstrates that images in EFL textbooks function as semiotic resources that shape comprehension, emotional engagement, and sociocultural interpretation. This confirms that visual grammar analysis provides a powerful framework for understanding how multimodal literacy is realized in educational materials.

3.3 Student Responses to Multimodal Texts

Based on interviews with ten students, several key themes emerged that demonstrate how multimodal elements significantly influence students' meaning-making processes. In line with Gunther Kress and Theo van Leeuwen [6], meaning is constructed not only through written language but also through visual and spatial resources that function as semiotic modes. The findings of this study reveal that images, symbols, layout design, and color choices work together to shape students' comprehension and engagement with textbook content.

These modes play an important role in supporting interpretation. Images illustrate abstract ideas and trigger emotional responses; symbols function as visual shortcuts to key information; and layout design influences how easily students navigate and prioritize content. As Kress and van Leeuwen argue, visual structures carry representational, interactive, and compositional meanings that guide viewers in interpreting texts. This perspective is clearly reflected in students' responses throughout the interviews.

The purpose of this analysis is to understand how these visual components support students in constructing meaning from textbook materials. When written language alone is difficult to understand, visuals provide contextual clues that clarify and make content more accessible. Student responses indicate that multimodal features guide interpretation, assist vocabulary development, and enhance motivation through emotional engagement. This aligns with the broader objective of multimodal literacy, which emphasizes the ability to interpret and implicitly engage with various semiotic resources in educational texts [16].

Table 2. Students' Responses to the Role of Multimodal Elements in EFL Textbooks

Description	Student Quotes
Visuals serve as previews to help students anticipate lesson content.	- The pictures really help me understand the topic... it makes me feel like I know what to expect. (Student 1) - When I saw it, I immediately thought about history and leadership. (Student 2) - I guessed it was about friendship or social events—and I was right (Student 3)

Students reported that images serve as anticipatory frameworks, helping them predict lesson topics before reading the text. This reflects what Kress and van Leeuwen describe as the representational function of images, where visuals depict participants, actions, and conceptual meanings. By observing visual cues such as clothing, facial expressions, or setting, students activate prior knowledge and prepare cognitively for the reading task. This previewing process reduces anxiety and increases confidence in engaging with English texts.

Table 3. Enhanced Learning Through Text–Image Integration

Description	Student Quotes
Images combined with text improve comprehension and support guessing unfamiliar words.	- When they are on the same page, I can connect them right away. (Student 6) - If I don‘t understand the text, sometimes the image explains it better. (Student 5)

Students emphasized the importance of proximity between text and image. When visuals are positioned close to related written content, students can immediately integrate both modes. This aligns with the compositional meaning in visual grammar, particularly the concept of information value and salience. The integration of image and text helps learners infer meaning, especially when encountering unfamiliar vocabulary. Rather than relying solely on translation, students use visual context as a strategy for comprehension, demonstrating developing multimodal literacy skills.

Table 4. Influence of Emotional Expressions in Visuals

Description	Student Quotes
Emotional cues from visuals impact student mood and motivation.	- A smiling person makes me enjoy the reading. (Student 7) - A smiling face makes the lesson feel more relaxed. (Student 9)

Emotional expressions in images were found to influence students' affective responses. Smiling faces, friendly gestures, and warm color tones created a relaxed learning atmosphere. In visual grammar terms, this relates to interactive meaning, particularly gaze and facial expression, which establish a relationship between the image and the viewer. Students reported feeling more motivated and less pressured when visuals conveyed positive emotions. This suggests that multimodal elements contribute not only to cognitive comprehension but also to emotional engagement in EFL learning.

Table 5. Role of Symbols and Icons

Description	Student Quotes
Symbols like light bulbs and arrows help identify key ideas and sequences.	- When I see icons like light bulbs, I think it means an idea or something important. (Student 2) - Sometimes there are arrows that show steps or ideas. (Student 8)

Symbols and icons function as semiotic markers that organize information hierarchically. Students recognized recurring icons, such as light bulbs to signal "important ideas" and arrows to indicate sequence. This demonstrates how compositional structures guide reading pathways and emphasize salience. Such visual markers help students quickly identify essential information and understand procedural texts more effectively.

Table 6. Relating Learning to Real Life

Description	Student Quotes
Realistic images help learners connect vocabulary to their daily experiences.	- When I see pictures of students or daily life, I feel like, This is about me. (Student 5) - When I see situations that happen in real life, it feels useful and real.‖ (Student 6) - It helps me connect the English words with my own life. (Student 10)

Students showed greater engagement when images reflected familiar contexts, such as school life, friendship, or local culture. According to social semiotic theory, meaning is socially and culturally situated. When textbook visuals represent students' lived realities, learners perceive the material as relevant and authentic. This supports vocabulary retention and deeper comprehension because language is contextualized within recognizable experiences.

Table 7. Impact of Layout Design

Description	Student Quotes
Clean layout, use of color, and cultural relevance support navigation and connection.	- When things are organized nicely, I can focus better and find what I need quickly. (Student 4) - When I see pictures from my culture or places I know, I feel more connected to the lesson. (Student 1)

Layout design significantly influences how students navigate textbook pages. Clear spacing, balanced composition, and consistent color schemes make content easier to follow. In visual grammar, layout contributes to compositional meaning by structuring information flow from left to right or top to bottom. Students reported that organized pages reduced cognitive overload and allowed them to focus on learning objectives. Cultural relevance in images further strengthened identification with the material.

The findings indicate that the integration of multiple modes, visual, textual, symbolic, and spatial, plays a central role in shaping how students interpret meaning in Indonesian EFL textbooks. Consistent with Kress and van Leeuwen's visual grammar framework, the textbooks analyzed demonstrate representational meanings (depiction of

actions and concepts), interactive meanings (viewer engagement through gaze and emotion), and compositional meanings (layout and salience).

Importantly, students do not passively receive visual information. Instead, they actively interpret and integrate multimodal cues to construct meaning. Images help them predict topics, infer vocabulary, and connect learning with personal experiences. Emotional expressions foster motivation, while symbols and layout guide attention and comprehension pathways. This confirms that multimodal elements are not merely decorative but function as pedagogically meaningful resources.

Accordingly, educators and textbook developers should prioritize intentional multimodal design that aligns visual grammar principles with pedagogical objectives. Visual elements should be contextually relevant, emotionally supportive, and structurally coherent. By strategically integrating diverse semiotic modes, EFL textbooks can enhance students' multimodal literacy competence, enabling them to interpret complex texts and communicate effectively in an increasingly multimodal world.

Ultimately, this study reinforces the broader educational goal of equipping senior high school learners with the interpretive skills necessary to navigate contemporary communication landscapes while developing proficiency in English as a foreign language.

4. CONCLUSION

This study demonstrates that multimodal design in the analyzed EFL textbook operates as a structured meaning-making system that shapes how learning content is organized, emphasized, and socially positioned for students, rather than functioning merely as decorative support. By integrating visual grammar analysis with students' interpretive perspectives, the research confirms the theoretical relevance of social semiotic frameworks in textbook studies while highlighting the pedagogical importance of aligning visual, linguistic, and spatial elements with learners' literacy backgrounds. Practically, the findings suggest that textbook developers, curriculum designers, and teachers should treat multimodal elements as strategic instructional resources to enhance meaningful engagement and contextualized learning. However, the study is limited to one Grade X textbook and a small group of students in a single educational setting, which restricts broader generalization. Future research is recommended to examine multiple textbooks across regions, apply mixed or experimental designs to measure learning outcomes, and explore how teachers mediate multimodal materials in classroom practice. Overall, this study contributes to multimodal literacy research by bridging textual analysis and learner interpretation and underscores for educational stakeholders and the wider public the importance of critically designed textbooks in preparing students for increasingly visual and media-rich communication environments.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to the supervisors for their valuable guidance and support throughout this research. Appreciation is also extended to the teachers and students of the participating senior high schools for their cooperation and participation.

REFERENCES

[1] S. V. Aryand and Y. Tiarina, "Multimodality in ELT textbooks: A comparative study between locally and internationally published materials," in *Proceedings of the International Conference on Education and Language (ICEL)*, Medan: Universitas Muhammadiyah Sumatera Utara, 2021, pp. 45–56.

[2] G. Kress, *Multimodality: A social semiotic approach to contemporary communication*. Inggris: Routledge, 2010.

[3] C. Jewitt, *The Routledge handbook of multimodal analysis*. Inggris: Routledg, 2009.

[4] M. Walsh, "Multimodal literacy: What does it mean for classroom practice?," *Aust. J. Lang. Lit.*, vol. 33, no. 3, pp. 211–239, 2010.

[5] N. Trisanti, D. Suherdi, D. Sukyadi, and L. Education, "Multimodality Reflected in EFL Teaching Materials : Indonesian EFL In- Service Teacher ' s Multimodality Literacy Perception," *Lang. Circ. J. Lang. Lit.*, vol. 17, no. 1, pp. 13–22, 2022.

[6] G. Kress and T. van Leeuwen, *Reading images: The grammar of visual design*. Inggris: Routledge, 2006.

[7] R. R. Mayer, *Cognitive Theory of Multimedia Learning. In: Mayer RE, ed. The Cambridge Handbook of Multimedia Learning.* London: Cambridge University Press, 2014.

[8] A. Paivio, *Mental representations: A dual coding approach*. Inggris: Oxford University Press, 1986.

[9] N. A. Drajati, S. Tan, L., Haryati, D. Rochsantiningsih, and H. Zainnuri, "Investigating English language teachers' beliefs on multimodal literacy: A case study in Indonesian EFL classrooms," *J. Lang. Educ.*, vol. 6, no. 3, pp. 50–63, 2020, doi: https://doi.org/10.17323/jle.2020.10586.

[10] E. Dwi Jayanti and I. L. Damayanti, "Exploring Teachers' Perceptions of Integrating Multimodal Literacy into English Classrooms in Indonesian Primary Education," *Child Educ. J.*, vol. 5, no. 2, pp. 98–109, 2023, doi: 10.33086/cej.v5i2.5240.

[11] N. Hendrawaty, Z. Sakhiyya, S. Wahyuni, and Yulianti, "The multimodal approach in English language teaching: a systematic review," *Proc. Fine Arts, Lit. Lang. Educ.*, pp. 457–492, 2024.

[12] S. Jamilah, N. M. Ismail, and C. Faizah, "Multimodal Analysis of an English Textbook Used for EFL Young Learners," *New Lang. Dimens.*, vol. 5, no. 1, pp. 50–63, Jun. 2024, doi: 10.26740/nld.v5n1.p50-63.

[13] K. Rahikummahtum, J. Nurkamto, and S. Suparno, "The Pedagogical Potential of Visual Images in Indonesian High School English Language Textbooks: A Micro-Multimodal Analysis," *AL-ISHLAH J. Pendidik.*, vol. 14, no. 4, pp. 5979–5990, Sep. 2022, doi: 10.35445/alishlah.v14i4.2171.

[14] A. D. Nugraheni, J. Nurkamto, and K. A. Putra, "Teachers' Multilingualism Belief and Practice in Indonesian EFL Classroom," *AL-ISHLAH J. Pendidik.*, vol. 15, no. 3, pp. 2655–2665, Jul. 2023, doi: 10.35445/alishlah.v15i3.2908.

[15] V. L. L. Cristovão, B. Sanches, and G. Smart, "Environmental discourse in Brazilian English-as-a-foreign-language textbooks: socio-discursive practices and their implications for developing students' critical environmental literacy," *Environ. Educ. Res.*, vol. 28, no. 1, pp. 75–94, Jan. 2022, doi: 10.1080/13504622.2021.2007855.

[16] A. A. K. Ningtiyas, I. Ghozali, and D. A. Nugraheni, "Multicultural Values in the Indonesian EFL Textbook 'English for Nusantara': A critical discourse analysis," *JEAPCO J. English Acad. Prof. Communcation*, vol. 11, no. 2, pp. 121–138, 2025, doi: https://doi.org/10.25047/jeapco.v11i2.5805.

[17] N. K. Denzin and Y. S. Lincoln, *The SAGE handbook of qualitative research*, 4th ed. New York: Sage Publications, Inc, 2011.

[18] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough?," *Field methods*, vol. 18, no. 1, pp. 59–82, Feb. 2006, doi: 10.1177/1525822X05279903.

[19] A. Bashir, A. Aziz, M. Imran, and N. M. Almusharraf, "A Multimodal Discourse Analysis of Visual Representation in English Language Textbooks at the Elementary Level: A Social Semiotic Perspective," *Khazar J. Humanit. Soc. Sci.*, vol. 28, no. 3, pp. 135–153, 2025, doi: 10.5782/2223-2621.1348.

[20] L. Unsworth, "Multimodal reading comprehension: curriculum expectations and large-scale literacy testing practices," *Pedagog. An Int. J.*, vol. 9, no. 1, pp. 26–44, Jan. 2014, doi: 10.1080/1554480X.2014.878968.

[21] Y. Zhang and K. L. O'Halloran, "Multimodal literacy pedagogy," *Vis. Commun.*, vol. 18, no. 2, pp. 197–221, 2019, doi: https://doi.org/10.1177/1470357218759806.

[22] J. Callow, *The shape of text to come: How image and text work*. Austalia: rimary English Teaching Association Australia, 2013.

[23] I. L. Damayanti and Y. Febrianti, "Multimodal literacy in Indonesian EFL textbooks: Representation, design, and pedagogical implications," *Indones. J. Appl. Linguist.*, vol. 13, no. 1, pp. 1–12, 2023.

[24] F. Serafini, *Reading the visual: An introduction to teaching multimodal literacy*. Inggris: Teachers

College Press, 2014.
[25] M. Larsen-Walker, “Can Data Driven Learning address L2 writers’ habitual errors with English linking adverbials?,” *System*, vol. 69, pp. 26–37, Oct. 2017, doi: 10.1016/j.system.2017.08.005.