J-LEE
Journal of Law, Economics, and English
http://ejournal.uhb.ac.id/index.php/J-LEE/issue/archive

REPRESENTATIONAL MEANINGS AND POTENTIAL VISUAL SUPPORT IN *ENGLISH FOR NUSANTARA* GRADE VII: A MULTIMODAL TEXTBOOK ANALYSIS

Muhammad Soali

Fakultas Ilmu Sosial, Universitas Harapan Bangsa

Email: muhammadsoali@uhb.ac.id.com

Aulia Zilmi Kafah

Email: auliakafah@gmail.com

Fakultas Ilmu Sosial, Universitas Harapan Bangsa

Abstract

Visual elements are increasingly integrated into English language textbooks as resources for representing participants, actions, events, and contextual information. However, the pedagogical contribution of such visuals should be interpreted cautiously when the evidence is derived from textbook analysis rather than learner-based data. This study examines the representational meanings realized in visual elements accompanying reading texts in the *English for Nusantara* Grade VII Student Book and explores their potential comprehension-supporting functions. A qualitative document analysis was conducted on 33 visual elements accompanying 13 selected reading texts. The analysis employed Kress and van Leeuwen's (2006) Visual Grammar, particularly the representational meaning system, to identify narrative and conceptual structures and their process types. The visual–verbal relationship was also considered to interpret the potential contribution of the visual elements to reading comprehension. The analysis indicates that the visuals represent everyday activities, interpersonal interactions, descriptive information, and conceptual relationships through narrative and conceptual structures. These representations provide contextual, participant-related, action-related, and organizational cues that may support readers in interpreting the accompanying texts. From a multimedia-learning perspective, such visual–verbal correspondence may provide additional routes for meaning-making; however, the study does not claim that the visuals directly improve learners' comprehension. The findings suggest that purposeful visual representation can function as a potential source of comprehension support in Indonesian EFL textbook materials.

Keywords: representational meaning; visual grammar; visual support; multimodality; EFL textbook; reading comprehension

Abstrak

Elemen visual semakin banyak diintegrasikan ke dalam buku teks bahasa Inggris sebagai sumber untuk merepresentasikan partisipan, tindakan, peristiwa, dan informasi kontekstual. Namun, kontribusi pedagogis dari visual tersebut perlu diinterpretasikan secara hati-hati ketika bukti yang digunakan berasal dari analisis buku teks, bukan dari data yang diperoleh langsung dari peserta didik. Penelitian ini mengkaji makna representasional yang diwujudkan melalui elemen visual yang menyertai teks-teks bacaan dalam *English for Nusantara Grade VII Student Book* serta mengeksplorasi potensi fungsinya dalam mendukung pemahaman bacaan.Analisis dokumen kualitatif dilakukan terhadap 33 elemen visual yang menyertai 13 teks bacaan terpilih. Analisis menggunakan Visual Grammar dari Kress dan van Leeuwen (2006), khususnya sistem makna representasional, untuk mengidentifikasi struktur naratif dan konseptual beserta jenis prosesnya. Hubungan antara visual dan teks verbal juga dipertimbangkan untuk menginterpretasikan potensi kontribusi elemen visual terhadap pemahaman bacaan. Hasil analisis menunjukkan bahwa visual merepresentasikan berbagai aktivitas sehari-hari, interaksi interpersonal, informasi deskriptif, dan hubungan konseptual melalui struktur naratif dan konseptual. Representasi tersebut memberikan petunjuk kontekstual, terkait partisipan, terkait tindakan, dan terkait pengorganisasian informasi yang dapat membantu pembaca dalam menginterpretasikan teks yang menyertainya. Dari perspektif pembelajaran multimedia, kesesuaian antara visual dan teks verbal tersebut dapat menyediakan jalur tambahan dalam proses pembentukan makna. Namun demikian, penelitian ini tidak mengklaim bahwa visual secara langsung meningkatkan pemahaman peserta didik. Temuan penelitian menunjukkan bahwa representasi visual yang dirancang secara tepat dapat berfungsi sebagai sumber potensial untuk mendukung pemahaman dalam materi buku teks EFL di Indonesia.

Kata kunci: makna representasional; tata bahasa visual; dukungan visual; multimodalitas; buku teks EFL; pemahaman bacaan

INTRODUCTION

English language textbooks increasingly incorporate visual resources alongside written language. Images, illustrations, diagrams, infographics, and other visual elements are now commonly embedded in language-learning materials to represent people, activities, settings, objects, and abstract information. In an EFL context, these resources are particularly relevant because learners may encounter linguistic expressions, cultural contexts, and concepts that are not immediately accessible through written language alone. From a multimodal perspective, visual resources are not simply decorative components but semiotic resources that participate in the construction and interpretation of meaning (Jewitt, 2014; Kress, 2010). Multimodality recognizes that meaning is constructed through the interaction of multiple semiotic modes, including language, image, layout, colour, typography, and spatial organization (Jewitt et al., 2016). In educational materials, the combination of verbal and visual resources may provide learners with different sources of information for interpreting a text. Mayer's (2021) Cognitive Theory of Multimedia Learning similarly proposes that learning materials can involve both

verbal and pictorial representations. However, the presence of visual resources in a textbook does not by itself establish their effectiveness for learning. Their potential contribution needs to be examined in relation to what the visuals represent and how they correspond with the accompanying verbal information. Visual Grammar provides a systematic framework for examining how images construct meaning. Kress and van Leeuwen (2006) distinguish three broad metafunctional dimensions of visual communication: representational, interactive, and compositional meaning. Among these dimensions, representational meaning is particularly relevant to the analysis of textbook visuals because it concerns how participants, actions, events, and conceptual relationships are depicted. Representational structures are broadly divided into narrative and conceptual representations. Narrative representations involve processes and vectors that depict actions, reactions, or interactions, whereas conceptual representations organize participants and information through structures such as classification and analytical relationships. The application of Visual Grammar to EFL textbooks has received increasing attention in recent years. Previous research has demonstrated that textbook images can contain systematic representational structures rather than functioning merely as illustrations. For example, multimodal studies of EFL textbooks have identified narrative and conceptual processes through the application of Kress and van Leeuwen's framework, showing how visual elements contribute to the representation of participants, actions, and relationships. Research on English for Nusantara has also examined multimodal resources from different perspectives. Febraningrum and Suroso (2023), for example, examined the content of the Grade VII textbook in relation to the Merdeka Curriculum and language-learning development, while other recent studies have investigated multimodality and visual representation in English for Nusantara materials. More recent work has examined the interaction between verbal and visual modes in the Grade VII coursebook, indicating that visual and linguistic resources can operate together in the construction of meaning. Recent research also suggests that multimodal resources are increasingly relevant to reading instruction. Daulay and Utami Dewi (2025) examined the use of a multimodal literacy approach in English reading and emphasized the contribution of multimodal resources to reading-related learning. Similarly, Pramestri and Priyana (2025) analyzed multimodal texts in the Grade VII English for Nusantara coursebook by combining Systemic Functional Linguistics for verbal texts with Visual Grammar for visual texts. Their study demonstrates the importance of examining relationships between verbal and visual resources rather than treating textbook images as isolated objects. These studies provide an important foundation for the present research because they demonstrate that English for Nusantara contains multimodal resources that warrant systematic analysis. Nevertheless, a more specific gap remains. Existing research on English for Nusantara has examined the textbook from several perspectives, including curriculum alignment, multimodality, visual representation, and verbal–visual relationships. However, less attention has been given specifically to the representational meanings of visual elements accompanying reading texts and the potential comprehension-supporting functions that can be inferred from these representations. In particular, the

relationship between the representational structures of visuals and the types of information needed for interpreting reading texts requires further examination. This distinction is important because simply identifying an image as narrative or conceptual does not explain what kind of reading-related support the image potentially provides. The present study therefore focuses on the visual elements accompanying selected reading texts in the Grade VII English for Nusantara Student Book. The textbook is an official English learning resource developed within the context of the Merdeka Curriculum and is designed for learners in Phase D. The selected reading materials contain visual elements that depict everyday activities, interpersonal interactions, physical characteristics, food-related information, learning activities, and other contextual information. These materials provide an appropriate corpus for investigating how representational meaning is constructed visually and how such representations relate to the information presented in the accompanying reading texts. Importantly, this study does not treat visual elements as proven learning interventions. Because the research employs qualitative document analysis and does not involve students, teachers, comprehension tests, interviews, or classroom observations, it cannot determine whether the visuals actually improve learners' reading performance. Instead, the study examines the potential visual support that can be inferred from the observed visual structures and their relationship to the accompanying verbal texts. This distinction addresses an important methodological issue: document analysis can establish what is represented and how the representation may function semiotically, but claims about actual learning effects require empirical evidence from learners. The concept of visual support in this study is therefore understood more cautiously than the traditional concept of pedagogical scaffolding. Classical scaffolding refers to support that is contingent on learners' needs and can be adjusted or gradually withdrawn during interaction (Vygotsky, 1978; Wood et al., 1976). A static textbook image does not possess these interactive characteristics automatically. Accordingly, the present study uses potential visual support to refer to representational cues that may assist readers in locating, connecting, or interpreting information presented in the accompanying text. These cues may include representations of participants, actions, settings, sequences, characteristics, or conceptual relationships. The interpretation of potential visual support is informed by multimedia learning theory, particularly Mayer's (2021) explanation of learning through verbal and pictorial representations. In this study, Mayer's framework is not used as empirical evidence that the textbook visuals reduce cognitive load or improve comprehension. Rather, it provides a theoretical lens for discussing how the correspondence between visual and verbal information may potentially facilitate meaning-making. The interpretation therefore remains grounded in the observed features of the textbook rather than in assumptions about learners' actual cognitive processing. This study addresses the following research questions: RQ1: What types of representational meanings are realized in the visual elements accompanying the selected reading texts in English for Nusantara Grade VII?, RQ2: What potential comprehension-supporting functions can be inferred from the relationships between these visual representations and the accompanying verbal

texts? By addressing these questions, the study contributes to research on multimodal EFL textbook analysis in two ways. First, it provides a systematic account of how representational meaning is realized through narrative and conceptual visual structures in an Indonesian EFL textbook. Second, it moves beyond simple categorization by examining what kinds of comprehension-related support may be inferred from the relationship between visual representations and their accompanying reading texts. The study thus offers a cautious interpretation of the pedagogical potential of textbook visuals while maintaining a clear distinction between observed visual features, theoretically inferred affordances, and empirically demonstrated learning effects.

METHODOLOGY

Research Design

This study employed qualitative document analysis within a multimodal discourse framework to examine the representational meanings of visual elements accompanying selected reading texts in the *English for Nusantara* Grade VII Student Book. Kress and van Leeuwen's (2006) Visual Grammar was used as the primary analytical framework, while Mayer's (2021) Cognitive Theory of Multimedia Learning served as an interpretive lens. The study focuses on observable visual features and their potential comprehension-supporting functions rather than actual learning effects, as no learner-based data were collected.

Corpus and Unit of Analysis

The corpus consisted of visual elements accompanying 13 selected reading texts, comprising 33 visual elements. The individual visual element was the primary unit of analysis. Included visuals were images, illustrations, photographs, diagrams, or comparable pictorial representations directly associated with the selected reading texts. Decorative elements, isolated icons, and visuals used exclusively for other activities were excluded. Multi-panel visuals were treated as separate units only when their panels were visually and functionally distinguishable; otherwise, they were coded as one visual unit.

Analytical Framework

The analysis focused on representational meaning, comprising narrative and conceptual structures. Narrative representations included actional, reactional, speech, and mental processes, while conceptual representations included classification, analytical, and symbolic processes. Categories were based on observable visual features. Complex visuals could receive multiple process codes when more than one process was identifiable within the same Visual ID.

Coding Procedure

Each included visual was assigned a unique Visual ID and coded for representational category, process type, participant roles, and relevant visual

relations. The visual representation was then examined in relation to the accompanying verbal text to identify contextual, participant-related, action-related, descriptive, or organizational correspondences. Finally, potential comprehension-supporting functions were inferred from these visual–verbal relationships. These interpretations indicate potential affordances rather than demonstrated effects on learners.

Coding Decision Rules

A visual was coded as narrative when it represented an observable action, reaction, communication, or mental process, and as conceptual when it primarily organized participants, characteristics, or relationships. Actional processes involved vectors connecting actors and goals; reactional processes involved gaze or visual attention; speech processes represented communication; and mental processes represented internal thought. Classification concerned grouping by shared characteristics, analytical processes represented part–whole or possessive relationships, and symbolic processes conveyed identity or significance. Multiple processes were permitted within one visual when supported by the visual evidence. Interpretations were limited to potential comprehension support; actual comprehension, cognitive-load reduction, and learning effectiveness were not inferred.

Trustworthiness

To ensure analytical consistency, the coding used predefined categories and decision rules, and each visual was recorded under a unique Visual ID. The coding records were reviewed to identify inconsistencies and maintain an audit trail linking interpretations to the visual evidence. Because independent double-coding was not conducted, no inter-coder agreement statistic was calculated.

Trustworthiness of the Analysis

Analytical consistency was addressed through systematic application of predefined coding categories and decision rules. The coding records were reviewed repeatedly to identify inconsistencies in categorization and to maintain alignment between the visual evidence and the analytical interpretation. An audit trail was maintained by linking each coded visual to its Visual ID and textbook location.

To ensure the trustworthiness of the analysis, the coding process was conducted systematically using predefined analytical categories derived from Kress and van Leeuwen's (2006) Visual Grammar. Each visual element was assigned a unique Visual ID and examined according to its representational structure, process type, participant roles, and representational function. The coding decisions were documented in a coding table to maintain an audit trail and to ensure consistency across the analyzed visual elements. The researcher also reviewed the coding records repeatedly to identify inconsistencies and to ensure that each interpretation

remained grounded in observable visual features. Particular attention was given to complex visuals containing more than one representational process, such as actional, reactional, speech, or analytical processes. Rather than forcing each visual into a single category, multiple process types were retained when they were simultaneously identifiable within the same visual. This procedure was intended to strengthen the transparency and consistency of the qualitative interpretation. Because the study did not employ independent double-coding, no inter-coder agreement statistic was calculated. The analysis therefore emphasizes an explicit audit trail, predefined decision rules, and repeated review of the coding records as strategies for maintaining analytical trustworthiness.

Interpretation of Potential Visual Support

The term potential visual support is used in this study instead of treating textbook visuals as demonstrated scaffolds. Classical scaffolding refers to support that is contingent on learners' needs, responsive to interaction, and potentially reduced as learners become more capable (Vygotsky, 1978; Wood et al., 1976). Static textbook visuals do not automatically possess these interactive and adaptive characteristics.

Accordingly, potential visual support in this study refers to representational cues embedded in the textbook that may assist readers in locating, connecting, or interpreting information in the accompanying reading text. Such cues may concern participants, actions, settings, sequences, characteristics, or conceptual relationships. The use of this term allows the analysis to discuss the pedagogical potential of visual representations without claiming that the textbook images function as empirically verified scaffolds.

Mayer's (2021) Cognitive Theory of Multimedia Learning was used to interpret this potential relationship between visual and verbal information. The theory provides a basis for discussing how learners may potentially construct meaning through coordinated verbal and pictorial representations. However, because no learner data were collected, the present study does not use Mayer's theory to establish that the visual elements reduce cognitive load or improve comprehension. Instead, the theory provides a conceptual lens for interpreting how the observed visual–verbal correspondences may afford additional routes for meaning-making.

Thus, the methodological distinction adopted in this study is threefold: (1) observable visual structures are established through Visual Grammar analysis; (2) potential comprehension-supporting functions are inferred from visual–verbal relationships; and (3) actual learning effects remain outside the evidential scope of the study.

RESULTS AND DISCUSSION

Overview of the Visual Representations

The analysis of the coded visual elements identified both narrative and conceptual representations in the selected reading materials of the *English for Nusantara* Grade VII Student Book. The coding records show that the visual elements do not operate through a single representational pattern. Instead, they combine representations of actions, reactions, verbal interactions, mental processes, classifications, and analytical relationships. In several cases, a single visual contains more than one representational process. For example, V1 combines an action process with an analytical process, while V4 combines action, speech, and analytical processes.

The presence of multiple processes within the same visual is important for the interpretation of the findings. A complex visual cannot always be adequately described by assigning only one category to it. V1, for instance, represents students participating in hobby-related activities through action processes while also presenting Galang through an analytical relationship involving his appearance and associated attributes. Similarly, V4 depicts dining activities and verbal interaction while simultaneously associating particular meals and drinks with the depicted participants.

This finding suggests that the representational meanings of textbook visuals are often layered. Narrative processes contribute to the representation of what participants are doing, whereas conceptual processes can provide information about who the participants are or what attributes are associated with them. Consequently, the visual meaning cannot always be reduced to a simple distinction between "narrative" and "conceptual." Instead, the two structures may operate together within the same visual composition.

From the perspective of Visual Grammar, this pattern demonstrates the usefulness of examining both the process and participant relationships represented in textbook visuals. The findings are also consistent with the reviewer's recommendation that complex images should be allowed to contain multiple processes rather than being forced into mutually exclusive categories.

Narrative Representations: Action, Reactional, Speech, and Mental Processes

Action Processes

Action processes constitute an important pattern in the coded visuals. V1, V3, V4, V5, and V7 contain actional representations. These visuals depict participants engaged in recognizable activities, including sports, dining, shopping preparation, household activities, and other everyday actions. V3 provides a relatively straightforward example. Made is represented as the actor, while the basketball functions as the goal of the action. The visual therefore constructs meaning through a transactional action process and represents participation in a

sports activity. V7 similarly represents family members participating in household activities, with cleaning tools and household objects functioning as goals. The visual consequently represents both physical activity and family cooperation.

The potential comprehension-supporting function of such representations lies primarily in their ability to provide concrete visual cues for actions and participants. When an action represented in an image corresponds to an action described in the accompanying reading text, readers may have an additional visual reference for interpreting the event. This interpretation should, however, remain cautious. The analysis establishes that the visual depicts a particular action and that such a depiction may provide an additional cue; it does not establish that learners actually understand the action more successfully.

This distinction is important in relation to Mayer's (2021) Cognitive Theory of Multimedia Learning. The theory provides a useful conceptual basis for discussing how verbal and pictorial representations may work together, but the present document analysis does not provide learner evidence demonstrating improved comprehension or reduced cognitive processing demands. The visual–verbal relationship is therefore interpreted as a potential affordance, rather than as an empirically demonstrated learning effect.

Reactional Processes

Reactional processes appear in V8, V9, and V12. In V8, students are represented as reactors who visually attend to signs or stickers. In V9, the teacher and students are represented as reactors, with one another functioning as the phenomenon. V12 represents Monita's attention toward a laptop and learning materials.

These visuals construct meaning through visual attention and gaze. Rather than simply showing what participants are doing physically, reactional representations indicate what participants are looking at or responding to. In V8, for example, the combination of action and reactional processes represents both students' physical actions and their visual attention toward instructional signs.

The potential comprehension-supporting function of reactional representations can therefore be understood as an attention-related cue. A reader may use the represented gaze direction or visual focus to identify the object, person, or activity that is relevant within the depicted situation. In V9, the mutual gaze between teacher and students is accompanied by classroom dialogue, creating a visual representation of interpersonal interaction.

However, the finding should not be interpreted as evidence that the visuals cause learners to attend more effectively. The study does not observe learner behavior or eye movements. It only establishes that the visuals represent attention through reactional structures. Thus, the strongest defensible interpretation is that

these representations potentially provide contextual cues concerning who is attending to whom or to what.

Speech Processes

Speech processes are represented in V4, V9, and V10. V4 represents dialogue between Monita and Galang through speech bubbles, while V9 depicts classroom dialogue between the teacher and students. V10 represents digital communication through written messages exchanged between Pipit and Monita using a smartphone.

These representations extend the meaning of the visuals beyond physical action. In V4, the speech process is integrated with an action process because the characters are simultaneously represented as dining and communicating. In V10, communication is represented through a digital interface rather than conventional speech balloons, but the exchange still functions as a visual representation of verbal communication.

The potential comprehension-supporting function of these visuals is particularly related to identifying speakers, communicative exchanges, and interactional context. Speech bubbles and digital message interfaces can make the relationship between participants and their utterances visually explicit. For readers working with dialogue-based texts, this may provide an additional contextual cue for identifying who communicates with whom.

Again, this finding should not be interpreted as evidence that the visual representation improves learners' comprehension of dialogue. Rather, the analysis indicates that the visual structure itself provides information about communicative participants and their interaction. Such information can potentially complement the verbal content of the reading text.

Mental Processes

Mental representation is identified in V13. Andre is represented as the senser, while study plans, reminders, and ideas shown in thought bubbles function as the phenomenon. The visual therefore represents internal thinking and planning through a mental process. V13 is analytically important because mental processes cannot be represented through physical action alone. The thought bubbles provide a visual convention for making otherwise invisible mental activity visible. This representation potentially assists readers in connecting the depicted character with ideas, intentions, or plans represented in the accompanying text.

The potential support here is therefore primarily interpretive and contextual. The visual may help readers identify that the character is engaged in thinking or planning rather than merely performing a physical activity. Nevertheless, because the mental content is represented through a conventional visual device, the

interpretation depends partly on readers' familiarity with that convention. The study therefore describes the visual affordance rather than claiming a measurable cognitive benefit.

Conceptual Representations: Classification and Analytical Processes

In addition to narrative representations, the coding records identify conceptual representations through classification and analytical processes. V1 contains an analytical representation alongside its dominant action process, while V2 is classified as conceptual through a classification process. V4 and V5 also combine narrative and analytical representations, whereas V6 and V11 are primarily analytical.

Classification Processes

V2 provides the clearest example of classification. The visual represents a group of teenagers as a superordinate category and distinguishes them according to physical characteristics such as hair type, hijab, height, and the use of crutches.

The visual consequently organizes participants according to shared and contrasting attributes. Its potential comprehension-supporting function is related to organizing descriptive information. Rather than requiring readers to rely exclusively on verbal descriptions, the visual makes differences among participants visually observable. This type of representation may be particularly relevant when reading materials introduce people through descriptions of physical characteristics. The visual provides a referential frame through which textual descriptions may be interpreted. However, the analysis does not establish that learners necessarily use the image in this way; it identifies only a plausible relationship between the visual organization and the information represented in the text.

Analytical Processes

Analytical processes are found in V1, V4, V5, V6, and V11. These representations organize relationships between entities and their associated attributes or components. V6 is a particularly clear example. The visual presents food items associated with information about Pecel, including its origin, dressing, ingredients, and taste. The representation therefore organizes descriptive information through an analytical relationship. V11 similarly presents study tips through an infographic structure containing categorized information and information blocks. The visual organizes several pieces of information into a structured composition rather than depicting an unfolding event.

The potential comprehension-supporting function of analytical representations can therefore be described as information organization. Such visuals may provide readers with a visual structure for relating an entity to its attributes, components, or associated information. This is potentially useful for

descriptive and informational reading passages in which comprehension involves recognizing characteristics and relationships rather than following a sequence of events.

The finding also demonstrates why conceptual visuals should not be regarded merely as decorative illustrations. Their representational structure can encode relationships among pieces of information. Nevertheless, the present study stops at this observation and theoretical interpretation; it does not claim that the organization necessarily produces better learning outcomes.

Image–Text Relationships and Potential Comprehension Support

Across the coded visuals, the representational structures frequently provide information that corresponds to the situations, participants, actions, or characteristics represented in the reading materials. The relationship is particularly visible in visuals depicting everyday activities. Sports, dining, shopping, household activities, classroom interaction, digital communication, and learning activities are represented through recognizable participants and actions.

Based on these patterns, three potential comprehension-supporting functions can be identified. First, visuals can provide participant cues. V1, V2, V4, V8, and V9 visually distinguish the people involved in the represented situation. Such representations may help readers connect names or descriptions in the verbal text with particular depicted participants. Second, visuals can provide action and event cues. V3 represents basketball activity, V5 represents shopping preparation, and V7 represents household activities. These images make the represented actions visually available alongside the verbal information. Third, visuals can provide descriptive and organizational cues. V2 organizes physical characteristics of teenagers, V6 organizes information about Pecel, and V11 organizes study-related information.

These functions can be interpreted through Mayer's Cognitive Theory of Multimedia Learning, which emphasizes the relationship between verbal and pictorial representations. However, in the present study, the theory functions only as an interpretive framework. The document analysis cannot establish whether learners actually integrate the two modes, whether cognitive load is reduced, or whether reading comprehension improves. The reviewer similarly emphasizes that such learner-level claims require empirical evidence such as comprehension tasks, interviews, think-aloud protocols, or classroom observation.

Discussion

The findings indicate that the visual elements analyzed in the coding sheet represent meaning through a combination of narrative and conceptual structures. Narrative processes make actions, reactions, communication, and mental activity visible, whereas conceptual processes organize participants and information according to characteristics or part–whole relationships. This pattern supports the

use of Kress and van Leeuwen's representational framework for examining textbook visuals because the images contain systematic relationships among participants, processes, and attributes rather than functioning only as isolated illustrations.

An important contribution of the analysis is the identification of mixed representational structures within individual visuals. V1 combines action and analytical processes, V4 combines action, speech, and analytical processes, and V5 combines action and analytical processes. This finding indicates that a visual may simultaneously represent an event and provide conceptual information about the participants or objects involved in that event. Such complexity supports the methodological decision to allow multiple process codes within a single visual.

The findings also suggest that the potential value of textbook visuals lies partly in their correspondence with information represented in the reading materials. Actional visuals can provide cues concerning what participants do; reactional visuals can indicate visual attention and interaction; speech representations can clarify communicative relationships; and conceptual representations can organize descriptive information. These functions correspond to different types of information that readers may need to identify or connect when interpreting a reading passage.

However, the findings do not demonstrate that these visuals improve reading comprehension. This distinction is essential because the present research analyzes textbook documents rather than learner performance. The study can establish the representational features of the visuals and infer their potential relationship with the accompanying verbal information, but it cannot establish actual comprehension outcomes. This limitation also means that terms such as *effective scaffold*, *improved comprehension*, *reduced cognitive effort*, or *facilitated learning* should not be used as empirical conclusions.

Accordingly, the more defensible interpretation is that the analyzed visuals offer potential comprehension-supporting affordances. Their possible contribution is located in the information they make visually available and in the relationships they establish with the accompanying verbal texts. This interpretation preserves the pedagogical relevance of the findings while maintaining a clear distinction between textual evidence and learner-based evidence.

Overall, the findings answer RQ1 by showing that the coded visual elements realize both narrative and conceptual representational meanings, including actional, reactional, speech, mental, classification, and analytical processes. They address RQ2 by indicating that these representations potentially support comprehension through participant cues, action and event cues, communicative cues, and descriptive or organizational cues. These conclusions remain limited to the observed visual–verbal relationships and should not be interpreted as evidence of actual learning effects.

CONCLUSION

This study found that the visual elements accompanying the selected reading texts in the *English for Nusantara* Grade VII Student Book realize both narrative and conceptual meanings through actional, reactional, speech, mental, classification, and analytical processes. These visual representations potentially support readers by providing participant, action, contextual, communicative, and descriptive cues that correspond to information in the accompanying texts. However, because this study employed qualitative document analysis without learner-based data, the findings do not demonstrate actual improvement in reading comprehension or cognitive processing. The study therefore concludes that the analyzed visuals have potential comprehension-supporting functions, while further empirical research involving learners is needed to determine their actual effects on reading comprehension.

REFERENCES

Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Jewitt, C. (Ed.). (2014). *The Routledge handbook of multimodal analysis* (2nd ed.). Routledge.

Jewitt, C., Bezemer, J., & O'Halloran, K. L. (2016). *Introducing multimodality*. Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia. (2022). *English for Nusantara Grade VII Student's Book*. Center for Curriculum and Book Development.

Nunan, D. (2003). *Practical English language teaching*. McGraw-Hill.

O'Halloran, K. L. (2004). *Multimodal discourse analysis: Systemic functional perspectives*. Continuum.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique, 5*(1), 55–76.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy, 33*(3), 211–239.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x