

How Visual Resources Support Reading Comprehension: A Multimodal Analysis of English for Nusantara for Grade VIII

Sulfaedah Lestari¹

¹STIA Abdul Haris, Indonesia

*Correspondence: ulfhalestary23@gmail.com

ABSTRACT

This study examines how visual resources accompanying reading materials in English for Nusantara for Grade VIII may support reading comprehension. A qualitative textbook content analysis was conducted on all 18 explicit Reading and Your Turn: Reading sections in Chapters 1-5. Meaningful visuals were defined as images or visual compositions that contributed information, organization, genre cues, or task meaning beyond decoration. Thirteen reading sections (72.2%) contained meaningful visual support, producing 14 visual clusters for analysis. Digital or genre interfaces and narrative illustrations were the most frequent types (21.4% each), followed by photograph sets, posters or infographics, and comic strips (14.3% each); a mixed image cue set and a spatial diagram each accounted for 7.1%. Using an operational image-text framework, complementary or integrated relations were most common (42.9%), while elaborative and contextual or enhancing relations each represented 28.6%. The visuals served functions such as activating prior knowledge, visualizing events and referents, guiding genre and layout recognition, organizing procedures, tracking speakers, supporting inference, explaining genre concepts, and mapping spatial information. These features may provide useful contextual and organizational support for reading, but the analysis does not demonstrate actual gains in students' comprehension. The findings suggest that visual resources are most pedagogically relevant when teachers direct attention to how visual and verbal information work together.

ARTICLE HISTORY

Published September 26th 2026



KEYWORDS

Multimodality, Visual Resources, Reading Comprehension, EFL Textbook, English for Nusantara.

ARTICLE LICENCE

© 2026 Universitas Hasanuddin
Under the license CC BY-SA
4.0



1. Introduction

Reading comprehension in a foreign language involves more than recognizing words and sentences. Readers need to connect linguistic information with prior knowledge, follow the organization of a text, identify relationships among ideas, and build a coherent representation of what the text means. In school settings, textbooks shape much of this work because they select texts, place supporting information around them, and guide students through tasks before, during, and after reading. Contemporary English textbooks also communicate through more than written language (Swandayani et al., 2026; Yaumi et al., 2023; Aswad et al., 2023.). Photographs, illustrations, diagrams, comics, posters, icons, and digital-interface designs can participate in meaning-making together with the verbal text.

From a multimodal perspective, visual and verbal modes offer different resources for representing experience and organizing attention. Kress and van Leeuwen (2006) describe visual communication as a structured system of meaning rather than as an optional decoration. Walsh (2010) similarly argues that literacy practices increasingly require readers to move across written, visual, and digital modes. In an English as a foreign language (EFL) textbook, this is important because a learner may meet unfamiliar vocabulary or content while also being expected to interpret what an image, layout, or graphic contributes to the same message.

Research on learning from text and pictures provides a useful cognitive reason for examining these combinations. Carney and Levin (2002) reviewed evidence showing that appropriately designed illustrations can improve learning from text, while also emphasizing that the instructional value of pictures depends on what they do in relation to the text. Eitel et al. (2013) showed that a relevant picture can provide a global structure that readers may use as a scaffold while processing verbal information. Mayer (2021) likewise explains that learning with words and graphics can be effective when the two sources of information are meaningfully coordinated. These findings should not be used to assume that every textbook

picture improves comprehension. Decorative or weakly related images may add little to a reading task. The more useful question is how a visual is connected to the verbal material and what kind of reading work that connection may support.

Image-text relations provide a practical way to investigate this issue. Martinec and Salway (2005) developed a system for describing the relative status and logico-semantic relations between images and text. Their work shows that an image can restate, expand, contextualize, or otherwise interact with verbal information. For a classroom textbook, these relations may be seen when a photograph makes a verbal referent concrete, when a diagram adds spatial information, when a comic combines speech and image into one message, or when a digital interface guides the reader through a blog post or comment thread. The present study uses this general principle in a simplified pedagogical coding scheme so that the analytical categories remain manageable and directly connected to reading tasks.

Studies of Indonesian EFL textbooks have already shown that images deserve systematic attention. Elmiana (2019), for example, analyzed visual images in Indonesian senior high school EFL textbooks and reported that visual design could support engagement with exercises and emphasize concepts expressed verbally. Jauhara et al. (2021) examined visual and verbal meaning in an EFL textbook and showed that intermodal relations and page layout can influence the reading path. These studies demonstrate that textbook visuals are part of the pedagogical message, not merely page decoration.

Recent work has also examined English for Nusantara, the textbook analyzed in the present study. Mahanani et al. (2025) investigated multimodality in Chapter 2 of the Grade VIII book through visual grammar. Yunikaputeri and Farida (2025) examined viewing skills in the same Grade VIII textbook and found that visual-literacy activities occur across the five main chapters. Nisa and Maknun (2026) evaluated the book's reading materials across Chapters 1-5 using textbook-evaluation criteria. These studies provide useful accounts of multimodality, viewing, and reading material. However, among the studies located for this review, the visual resources attached specifically to the book's explicit Reading sections have not been systematically mapped across Chapters 1-5 in terms of visual type, visual-verbal relation, pedagogical function, and possible comprehension support.

This focus is also consistent with the design of the textbook itself. English for Nusantara presents Reading and Viewing as related but distinct learning sections. The book explains that Reading exposes students to written texts and literal, inferential, and interpretive questions, while Viewing presents information through pictures, diagrams, tables, infographics, and other forms and trains students to make meaning by combining verbal and visual resources (Damayanti et al., 2022). A closer analysis of visuals within the Reading sections can therefore show how multimodality is built into tasks that are explicitly presented as reading work.

The purpose of this study is to identify the meaningful visual resources that accompany the explicit Reading sections of English for Nusantara for Grade VIII and to explain how those resources relate to the verbal text and reading task. The analysis addresses three questions: (1) What types of visual resources are used in the reading materials? (2) What functions do these visual resources perform in relation to the reading texts? and (3) How do these visual resources potentially support students' reading comprehension? Because this is a textbook analysis rather than a study of learners, the third question concerns pedagogical affordances visible in the material. It does not test whether students actually understand a text better because a visual is present.

2. Methodology

This study used a qualitative, document-based content analysis supported by simple descriptive counts (Fitriana & Wirza, 2021; Schreier, 2012; Suwandani & Hidayati, 2025). Document analysis is appropriate when the object of study is a stable written or visual record that can be examined systematically and repeatedly (Bowen, 2009). The analysis focused on the relationship between verbal reading material and visual resources rather than on students' performance. Frequencies and percentages were used only to describe the distribution of coded visual clusters; they were not treated as evidence of learning effects.

The primary data source was English for Nusantara untuk SMP/MTs Kelas VIII, first published in 2022 by Pusat Perbukuan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Damayanti et al., 2022). The uploaded digital edition contains the front matter, Chapter 0, and five main instructional chapters. An initial full-book scan was conducted before coding. Chapter 0 mainly prepares language resources for later chapters and does not use the repeated Reading/Your Turn: Reading section structure found in Chapters 1-5. The screening therefore covered every explicit Reading and Your Turn: Reading section in the five main chapters. Eighteen such sections were identified: three in Chapter 1, three in Chapter 2, three in Chapter 3, five in Chapter 4, and four in Chapter 5.

Following established principles of segmentation and unitizing in qualitative content analysis (Krippendorff, 2018; Schreier, 2012), the unit of analysis was operationally defined as a visual cluster associated with an explicit Reading section. Rather than treating every individual image as a separate unit, the analysis considered how visual elements were organized and functioned together within the multimodal composition of the textbook page (Bateman, 2008). A visual cluster was therefore defined as one bounded and pedagogically coherent visual composition or set that worked as a single reading resource. For example, the three parade photographs on page 41 were counted as one cluster because students were asked to interpret the set together, while the elephant preview on page 107 and the later story illustrations on pages 109–110 were counted as two clusters because they performed different tasks. This operational rule maintained a consistent level of analysis and prevented frequencies from being inflated simply because a single textbook task contained several small images.

Following a multimodal view of textbook pages as configurations of verbal, visual, and layout resources (Bateman, 2008; Weninger & Kiss, 2013), a visual was included when it contributed information, organization, genre cues, reference, sequence, stance, or task meaning connected to the reading material. Included resources therefore covered photographs, narrative illustrations, posters and infographics, comics, digital-interface representations, mixed image cues, and diagrams. Visuals serving only a decorative role were excluded, consistent with previous EFL textbook research that distinguishes informative and illustrative images from decorative visuals that do not contribute information to the associated text or learning task (Elmiana, 2019). Generic section icons, page ornaments, borders, background textures, and character drawings with no clear informational relation to the reading text were therefore excluded. Conventional tables used mainly to arrange verbal information were also excluded from the visual-resource count so that the analysis remained focused on image-based or multimodally integrated resources. These criteria were applied consistently before calculating frequencies.

Data collection proceeded in four steps. First, the complete PDF was searched page by page to identify the 18 Reading openings and their associated reading pages. Second, each section was inspected visually, including adjacent pages when the visual or verbal material continued across a page boundary. Third, an analysis sheet recorded chapter, unit, page number, reading activity, visual type, description of the visual, relation to verbal information, primary pedagogical function, and a cautious interpretation of possible comprehension support. Fourth, all codes were verified against the original page images to confirm inclusion and exclusion decisions, page references, and category assignments.

The analytical framework combined deductive and inductive coding. Visual type was coded from the form that students encountered on the page, and the type categories were refined after the full dataset had been reviewed. Visual-verbal relations were informed by Martinec and Salway's (2005) account of image-text relations but simplified into three operational categories suited to textbook reading: elaborative, when the visual concretized or re-expressed information also present in the verbal material; contextual/enhancing, when it added setting, genre, topical, emotional, or spatial information that framed the text; and complementary/integrated, when visual and verbal resources jointly formed the message or task and could not be interpreted adequately as independent parallel elements. These are operational categories for this study rather than a reproduction of the complete Martinec and Salway system.

Pedagogical function was coded according to what the visual was positioned to help the reader do in the immediate task. The final primary-function categories were prediction/schema activation, referential or event visualization, genre/layout orientation, procedural sequencing, dialogue/speaker tracking, persuasive or critical interpretation, genre/concept explanation, and spatial/conceptual mapping. One primary function was assigned to each cluster for descriptive counting, although a cluster could support more than one process. Interpretations of possible comprehension support were then compared with relevant multimodal and multimedia-learning literature (Carney & Levin, 2002; Eitel et al., 2013; Kress & van Leeuwen, 2006; Mayer, 2021).

All coding was conducted by a single researcher, and no inter-rater or intra-rater reliability statistic was calculated. This is a limitation of the study, particularly because the distinction between contextual / enhancing and complementary/integrated relations requires interpretive judgment. To make this judgment as consistent and transparent as possible, an explicit decision rule was applied: a visual was coded as complementary/integrated only when the reading task or passage could not be completed or adequately understood without the visual (e.g., questions that refer directly to poster imagery, or speech that is attributed to speakers only through comic panels); when the verbal text remained understandable on its own and the visual mainly supplied setting, topic, genre, or spatial framing, it was coded as contextual/enhancing. Borderline cases were rechecked against the original page and task instruction before a final code was assigned. To allow readers to evaluate each decision, the complete coding of all 14 clusters is reported in Table 3. Consistent with this limitation, descriptive claims in the Findings refer to directly observable textbook features, while claims about reading comprehension are phrased as potential support rather than measured effects.

3. Results and Discussion

3.1. Distribution and Types of Visual Resources

The first research question asked what types of meaningful visual resources occur in the Reading materials. Across Chapters 1-5, 13 of the 18 explicit Reading sections contained at least one visual cluster that met the inclusion criteria. This represents 72.2% of the screened sections. The five sections not counted as meaningfully visual were page 54, where the large character illustration does not provide specific information needed for the recount; page 116, where a conventional table organizes verbal story elements; page 261, where a character illustration is decorative in relation to the soccer narrative; page 273, where the activity is a phrase-sorting organizer; and page 274, where the story is presented as continuous verbal text. This distinction is important because counting every graphic mark as a visual resource would give a misleading picture of how visuals actually participate in reading.

The distribution was uneven across chapters. Every Reading section in Chapters 3 and 4 included meaningful visual support, whereas only one of four Reading sections in Chapter 5 did so. Chapters 1 and 2 each contained meaningful visuals in two of three Reading sections. Table 1 summarizes the section-level distribution.

Table 1. Distribution of Reading sections with meaningful visual support

Chapter	Reading sections screened	Sections with meaningful visuals	Coverage (%)
Chapter 1	3	2	66.7
Chapter 2	3	2	66.7
Chapter 3	3	3	100.0
Chapter 4	5	5	100.0
Chapter 5	4	1	25.0
Total	18	13	72.2

The 13 included sections produced 14 visual clusters because the Elephant and Friends Reading sequence contained both a separate prediction image on page 107 and a later sequence of story illustrations on pages 109-110. As shown in Table 2, digital or genre interfaces and narrative illustrations were the most frequent types, with three clusters each (21.4%). Photograph sets, posters or infographics, and comic strips each occurred twice (14.3%). One mixed image cue set and one spatial diagram each accounted for 7.1%.

These types show that the book does not rely on a single visual format. Some visuals are familiar textbook illustrations, but others imitate communication outside a conventional printed passage. Page 50 frames 'A School Parade' as a blog post; page 152 presents a social-media post and comments; page 159 presents a comment thread inside a phone-like interface. In these cases, layout is not merely decoration because it signals the kind of text students are expected to read. The two poster-based clusters on pages 166 and 215 also ask readers to interpret a designed public message, while the comic on page 205 uses panels and speech bubbles to organize dialogue.

Table 2. Types of meaningful visual clusters in the Reading sections

Visual type	Frequency	Percentage	Representative page(s)
Digital/genre interface	3	21.4%	50, 152, 159
Narrative illustration	3	21.4%	91, 107, 109-110
Photograph set	2	14.3%	41, 221
Poster/infographic	2	14.3%	166, 215
Comic strip	2	14.3%	205, 226
Mixed image cue set	1	7.1%	209
Spatial diagram	1	7.1%	256

Visual type	Frequency	Percentage	Representative page(s)
Total	14	100%	-

3.2. Visual-Verbal Relationships and Pedagogical Functions

The second research question concerns what the visuals do in relation to the reading text. Six of the 14 clusters (42.9%) were coded as complementary/integrated, four (28.6%) as elaborative, and four (28.6%) as contextual/enhancing. The largest category therefore consists of cases where the visual is part of the message or task itself rather than a picture placed beside an otherwise complete passage.

The parade activity on pages 41-42 is a clear elaborative example. Students first match three photographs to the labels 'Indonesian heroes' costumes parade,' 'decorated bicycles parade,' and 'traditional music instruments parade.' On the following page, Andre's recount mentions people wearing heroes' costumes, traditional instruments such as angklung, suling, and rebana, and children riding decorated bicycles. The photographs and verbal text therefore refer to the same event categories in different modes. The pictures may give concrete referents before students meet the same content in a longer recount.

Narrative illustrations work similarly in Chapter 2. On page 91, three small illustrations show the Ugly Duckling with chickens, a dog, and a cow before or alongside the final part of the story on page 92. The sequence on pages 109-110 depicts the elephant meeting forest animals and later confronting the tiger. These visuals do not replace the narrative, but they re-present characters and key events. In contrast, the single elephant image on page 107 is better understood as contextual/enhancing because it is placed with a story outline and a prompt asking students to predict the elephant's feeling before reading the full story.

Several Chapter 3 and Chapter 4 visuals are more strongly integrated with the verbal material. On page 152, the social-media layout identifies the post, user accounts, reaction icons, and comments as one reading object. Page 159 similarly presents an old-books problem followed by comments from different users inside a phone-like interface. Page 166 combines numbered handwashing illustrations with short verbal instructions, so order is carried by both visual position and language. Page 205 presents the reading dialogue as a comic, where character images, speech bubbles, and panel progression help identify who is speaking and what situation the dialogue refers to.

The anti-littering task on page 215 is also integrated because the reading questions require students to inspect poster imagery and slogans rather than read a separate passage. Students are asked why a swimmer looks angry and what appears on a man's head. Meaning must therefore be inferred from the combination of visual depiction and verbal campaign design. Page 226 uses a comic exchange to explain the idea of factual recounts: the speech bubbles state that previously read texts are factual stories, are based on fact-based events or incidents, and commonly end with an evaluation or emotional response. Here, the comic format turns a genre explanation into a visual-verbal dialogue.

Contextual/enhancing relations appear when a visual adds a frame that the verbal passage later develops. On page 209, two turtle images introduce the 'No More Plastic' topic before students read a factual rescue story on page 210 about scientists removing a plastic straw from a sea turtle's nostril. On page 256, a soccer-field diagram asks students to place player positions before they read a story in which Mirza struggles as a goalkeeper and later develops in another position. The diagram therefore adds a spatial model that can support the terminology used in the narrative.

Table 3 summarizes the 14 visual clusters, their operational visual-verbal relation, their primary pedagogical function, and the kind of comprehension support suggested by the page design. The final column is an analytical interpretation, not evidence from learner testing.

Table 3. Visual-verbal relations, primary functions, and potential comprehension support

Page(s)	Visual resource	Relation	Primary function	Potential support
41-42	Three parade photographs	Elaborative	Prediction/schema activation	Concrete references for people, instruments, and decorated bicycles
50	Blog-style frame	Contextual/enhancing	Genre/layout orientation	Signals blog genre and reading path

Page(s)	Visual resource	Relation	Primary function	Potential support
91-92	Ugly Duckling illustration set	Elaborative	Referential/event visualization	Anchors characters and successive encounters
107	Elephant image	predictionContextual/enhancing	Prediction/schema activation	Supports anticipation of character feeling and story problem
109-110	Elephant illustrations	storyElaborative	Referential/event visualization	Makes characters and major events visible
152	Social-media interface	postComplementary/integrated	Genre/layout orientation	Organizes post, users, reactions, and comments
159	Phone/comment-thread interface	Complementary/integrated	Genre/layout orientation	Helps track commenters and alternative suggestions
166	Handwashing infographic	Complementary/integrated	Procedural sequencing	Makes ordered actions visible through numbering and images
205	Comic-style dialogue	Complementary/integrated	Dialogue/speaker tracking	Links speech turns to characters and situation
209-210	Turtle photo/cartoon cues	Contextual/enhancing	Prediction/schema activation	Frames the plastic-harm problem before the factual story
215	Two anti-littering posters	Complementary/integrated	Persuasive/critical interpretation	Requires inference from image, slogan, and campaign purpose
221-222	Beach-cleanup photographs	Elaborative	Referential/event visualization	Provides concrete context for the following cleanup recount
226	Factual-recount comic	Complementary/integrated	Genre/concept explanation	Presents genre features as a structured dialogue
256-257	Soccer-field diagram	spatialContextual/enhancing	Spatial/conceptual mapping	Builds a spatial model for player-position vocabulary

3.3. Potential Support for Reading Comprehension

The third research question asks how these resources may support reading comprehension. The clearest pattern is that visual support often appears before or within the verbal reading task rather than after it. This positioning matters. A

prereading image can activate a topic or provide an initial model, while an image embedded in a passage can help maintain reference and organize information as reading proceeds. The textbook uses both patterns.

First, several visuals may support prediction and schema activation. The parade photographs on page 41 provide cultural and event cues before the recount on page 42. The elephant image and story-outline boxes on page 107 ask students to predict a feeling from the problem that no animal wants to be the elephant's friend. The sea-turtle images on page 209 ask students to infer why the animals look sad and why one holds a 'No More Plastic' sign before the rescue story appears. These are not simply illustrations of information already read. They position visual interpretation as part of preparation for the verbal text. This use is compatible with Eitel et al.'s (2013) finding that a relevant picture can provide a structure that supports later text processing, although the present textbook analysis cannot establish that the same cognitive effect occurs for these students.

Second, elaborative illustrations and photographs may help readers maintain reference to people, objects, and events. The Ugly Duckling and Elephant and Friends illustrations place visible characters beside episodes that contain multiple participants and changes in action. The beach-cleanup photographs on page 221 create a concrete scene before the factual recount on page 222 describes Melati and Isabel Wijsen, volunteers, beaches, rivers, streets, and collected plastic. Carney and Levin (2002) emphasize that text illustrations are most useful when they are relevant to the information learners are expected to process. The examples in this dataset fit that relevance condition because the pictured content is directly taken up by the reading activity or passage.

Third, the digital-interface examples may support genre recognition and navigation. A blog, a social-media post, and a comment thread have reading paths that differ from a continuous printed paragraph. User names, profile placement, comment organization, icons, and framing show where one contribution ends and another begins. On page 159, for example, students encounter several responses to the same problem of unused books; the visual organization helps separate the voices and makes the text resemble a familiar digital exchange. Jauhara et al. (2021) note that multimodal page layout can shape how readers move through textbook content. The current findings extend that point to explicit Reading sections in which the layout itself represents a communicative genre.

Fourth, some resources may reduce the need to construct temporal or spatial relations from words alone. The handwashing poster on page 166 combines numbered images and short instructions, making a procedure visible as an ordered sequence. The soccer diagram on page 256 provides a field layout before students meet goalkeeper, defender, midfielder, and forward vocabulary and before they read Mirza's soccer story. Mayer's (2021) work on multimedia learning suggests that coordinated words and graphics can support meaning construction when each mode contributes relevant information. Again, the textbook examples provide the design condition for such support, not proof of improved learning.

Fifth, the posters and comics create opportunities for inferential and interpretive reading rather than only literal retrieval. The anti-littering task on page 215 requires students to infer emotion, identify a symbolic placement of trash, and consider the campaign message. The page 205 comic requires the reader to connect speakers, expressions, and the topic of plastic in the ocean. This is consistent with the textbook's own statement that Reading activities include literal, inferential, and interpretive questions (Damayanti et al., 2022). It also reflects a broader view of multimodal literacy in which readers coordinate verbal and visual evidence to construct meaning (Kress & van Leeuwen, 2006; Walsh, 2010).

The concentration of meaningful visuals in Chapters 3 and 4 is another observable pattern. Both chapters use social-media posts, posters, environmental photographs, and comics as part of the Reading tasks. It would be too strong to claim from document analysis alone that the chapters were intentionally designed to increase multimodal complexity, but the distribution shows that environmental and action-oriented topics in these chapters are frequently represented through public and digital communication forms. This pattern also connects with Yunikaputeri and Farida's (2025) observation that visual-literacy work in the book is not limited to sections labelled Viewing.

The findings are related to, but narrower than, earlier analyses of the same textbook. Mahanani et al. (2025) examined multimodal features in Chapter 2 through representational, interactive, and compositional meaning. The present study instead tracks every explicit Reading section across Chapters 1-5 and asks what the visuals do in relation to those reading texts. Nisa and Maknun (2026) evaluated the quality and structure of reading materials across the same five chapters; the present analysis adds a multimodal layer by showing that 13 of 18 Reading sections contain visual resources that perform identifiable pedagogical functions. In this sense, reading-material evaluation and visual analysis address different parts of the same textbook design.

For classroom practice, the results suggest that teachers should not assume that students will automatically use the visual information provided on the page. Previous research has shown that learners may devote relatively little attention to textbook illustrations and may not spontaneously integrate verbal and pictorial information during reading (Hannus & Hyönä, 1999; Leng et al., 2024). A photograph can be ignored, a comic can be read only for its words, and a diagram can remain disconnected from the narrative unless attention is directed to the relation between modes. This issue is also relevant to second-language reading. In an experimental study with bilingual readers, tighter picture-word integration supported better comprehension in L2 reading than a more independent presentation of verbal and visual information (Yum et al., 2021). Simple teacher prompts can make the relation more explicit: asking what information appears only in the picture, what the text repeats, what the layout tells the reader about genre, or whether the image changes an initial prediction. Such questions preserve the reading focus while treating the visual resource as evidence that can be interpreted.

At the same time, the analytical limits should remain clear. The study identifies affordances in one textbook and uses one researcher's coding. It does not measure eye movement, vocabulary learning, recall, inference accuracy, motivation, or reading-test performance. A visual classified as potentially supportive may not be noticed or interpreted as intended by every learner. The results therefore describe how the textbook is designed to support reading, not what students actually gain from that design.

4. Conclusion

This study analyzed all 18 explicit Reading and Your Turn: Reading sections in Chapters 1-5 of English for Nusantara for Grade VIII. Thirteen sections (72.2%) contained meaningful visual support, producing 14 visual clusters. The resources were varied: digital or genre interfaces and narrative illustrations were the most frequent, followed by photograph sets, posters or infographics, comic strips, a mixed image cue set, and a spatial diagram. The first research question can therefore be answered by noting that visual support in the book extends beyond conventional illustration and includes several forms that imitate public, digital, procedural, and spatial communication.

For the second research question, the visuals performed three broad kinds of relations with verbal material. Complementary/integrated relations were the most common, while elaborative and contextual/enhancing relations occurred equally often. Their primary functions included activating prior knowledge, visualizing referents and events, orienting readers to genre and layout, sequencing procedures, tracking speakers, supporting critical interpretation, explaining a genre concept, and mapping spatial information. These functions were observable in the way the textbook placed and used visuals within the reading tasks.

For the third research question, the analysis indicates several ways the resources may support comprehension. They can provide contextual cues before reading, make participants and events concrete, organize digital or procedural reading paths, add spatial information, and invite inferential interpretation across modes. These claims concern potential support only. Because no learners were observed or tested, the study cannot conclude that the visuals caused better comprehension.

Pedagogically, the findings support an explicit approach to visual-verbal reading. Teachers can ask students to compare what is communicated visually and verbally, identify information available in only one mode, explain how a picture or layout changes their prediction, and justify interpretations with evidence from both modes. Future research could test these textbook affordances with students through comprehension tasks, interviews, classroom observation, or eye-tracking, and could compare visual-verbal reading design across grade levels or different Indonesian EFL textbooks. A second independent coder would also strengthen future content analyses of visual function.

References

- Aswad, M., Rahman, F., Said, I. M., Hamuddin, B., & Nurchalis, N. F. (2019). A software to increase English learning outcomes: An acceleration model of English as the second language. *The Asian EFL Journal*, 26(6.2), 157.
- Bateman, J. A. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan. <https://doi.org/10.1057/9780230582323> Springer Link

- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/10.3316/QRJ0902027>
- Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5-26. <https://doi.org/10.1023/A:1013176309260>
- Damayanti, I. L., Nurlaelawati, I., Hutasuhut, M. L., Febrianti, Y., Rahmadhani, R., & Devianty, G. A. (2022). *English for Nusantara untuk SMP/MTs Kelas VIII*. Pusat Perbukuan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Eitel, A., Scheiter, K., Schöler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48-63. <https://doi.org/10.1016/j.learninstruc.2013.05.002>
- Elmiana, D. S. (2019). Pedagogical representation of visual images in EFL textbooks: A multimodal perspective. *Pedagogy, Culture & Society*, 27(4), 613-628. <https://doi.org/10.1080/14681366.2019.1569550>
- Fitriana, W., & Wirza, Y. (2021). An analysis of multimodal text in EFL textbook of secondary school in Indonesia in assisting students' text understanding. In Proceedings of the Thirteenth Conference on Applied Linguistics (CONAPLIN 2020) (pp. 627–632). Atlantis Press. <https://doi.org/10.2991/assehr.k.210427.095>
- Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, 24(2), 95–123. <https://doi.org/10.1006/ceps.1998.0987>
- Jauhara, D., Emilia, E., & Lukmana, I. (2021). Re-contextualising 'Greeting': A multimodal analysis in an EFL textbook. *Indonesian Journal of Functional Linguistics*, 1(1), 15-24. <https://doi.org/10.17509/ijfsl.v1i1.32621>
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications. [SAGE Publications](https://www.sagepub.com/publications)
- Leng, X., Wang, F., Mayer, R. E., & Zhao, T. (2024). How to train students to engage in text-picture integration for multimedia lessons. *British Journal of Educational Technology*, 55(3), 1167–1188. <https://doi.org/10.1111/bjet.13419>
- Mahanani, D. H., Wahyuni, S., & Sakhiyya, Z. (2025). The application of multimodal approach in the 'English for Nusantara' coursebook for 8th graders. *Language Circle: Journal of Language and Literature*, 19(2s), 231-247. <https://doi.org/10.15294/lc.v19i2s.26825>
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337-371. <https://doi.org/10.1177/1470357205055928>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Nisa, S. R. F., & Maknun, L. (2026). A content analysis of reading material in the English for Nusantara textbook. *Esteem Journal of English Education Study Programme*, 9(2), 1324-1334. <https://doi.org/10.31851/7zhd1n78>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications. Google Buku
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications. [SAGE Publications](https://www.sagepub.com/publications)
- Suwandani, F., & Hidayati, S. (2025). Conceptual representation of culture in EFL textbook images: A multimodal analysis. *Dinamika Ilmu*, 25(1), 99–122. <https://doi.org/10.21093/di.v25i1.10098>
- Swandayani, N. E., Rahman, F., Sulastri, T., & Sosrohadi, S. (2026). The Use of Language Styles in Multimodal Texts on the Facebook Account of HTA. *International Journal of Research and Innovation in Applied Science*, 11(1), 447-456.

- Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *The Australian Journal of Language and Literacy*, 33(3), 211-239. <https://doi.org/10.1007/BF03651836>
- Weninger, C., & Kiss, T. (2013). Culture in English as a foreign language (EFL) textbooks: A semiotic approach. *TESOL Quarterly*, 47(4), 694–716. <https://doi.org/10.1002/tesq.87>
- Yaumi, M. T. A. H., Rahman, F., & Sahib, H. (2023). Exploring WhatsApp as Teaching and Learning Activities during Covid-19/New Normal era: A Semiotic Technology Analysis. *International Journal of Current Science Research and Review*, 6(12), 7627-7634.
- Yum, Y. N., Cohn, N., & Lau, W. K.-W. (2021). Effects of picture-word integration on reading visual narratives in L1 and L2. *Learning and Instruction*, 71, Article 101397. <https://doi.org/10.1016/j.learninstruc.2020.101397>
- Yunikaputeri, A. N., & Farida, A. N. (2025). Viewing skill in an English textbook for 8th grade. *Bookchapter English Language Teaching, Literature, and Translation*, 1(1), 1-29. <https://doi.org/10.15294/eltlt.v1i1.284>