Diksi

Diksi (2025), Vol. 33, No. 1, **33-52**
doi.org/10.21831/diksi.v32i2

How word-image relationships in children's picture books adapt across levels: A multimodal analysis

Muhamad Fanni Al Farisi[1]*, Bernadette Kushartanti[1], Dwi Purwanto[2]

AFFILIATION
[1]Universitas Indonesia, Indonesia
[2]University of Colorado Boulder, United States
* Corresponding Author:
muhammadfannia@gmail.com

ARTICLE HISTORY
- Received 25 January 2025
- Accepted 27 March 2025
- Published 31 March 2025

CITATION (APA STYLE)
Al Farisi, M., Kushartanti, B., & Purwanto, D. (2025). How word-image relationships in children's picture books adapt across levels: A multimodal analysis. *Diksi, 33*(1), 33-52. https://doi.org/10.21831/diksi.v33i1.82907

Abstract: Children's picture books apply a levelling system to adjust linguistic complexity according to the reader's developmental stage. Despite having constitutive elements of text and images, their interdependent relationships within this system receive less attention. While the relationships of texts and images in picture books have been extensively studied, there is still a gap in understanding how these relationships adapt across different reading levels. Applying Halliday's systematic functional linguistics (2004), Kress and Leeuwen's multimodal approach (2021), and Nikolajeva and Scott's approach to picture books (2006), this research qualitatively and quantitatively examined thirty-four text-image pairs from two children's picture books at low and high reading levels. This study aimed to reveal whether such an adaptation exists and, if so, to what extent. The findings confirm that an increasing complexity in text-image relationships. The lower-level picture book exhibits three types of text-image relationships: enhancing, complementary, and counterpointing. The higher-level book, however, shows four types, incorporating a contradictory relationship alongside the other three. These findings suggest that the increasing complexity in children's picture books occurs not only in linguistic aspects but also within multimodal relationships between texts and images, highlighting the need for picture book authors and educators to design or select picture books that progressively introduce multimodal complexity.

Keywords: *picture book, text-image relationship, transitivity, multimodal, systemic functional linguistics*

INTRODUCTION

Stories in picture books are built upon an interplay between words and images. According to Bateman (2014), one out of three properties that cannot be neglected when defining picture books is the existence of both verbal texts and images, indicating that any artefacts with only verbal texts or only images cannot be considered picture books. Scholars consistently reinforce that picture books integrate both modes to construct meaning: Serafini (2024) defines picture books as artefacts that enclose "an amalgamation of words and images", Ramos (2020) characterizes them by their "close relationship between texts and images", and Nikolajeva and Al-Yaqout (2015) characterize them by artefacts which consist of "a combination of the verbal and visual modes". This interdependence between verbal texts and images in picture books consequently creates complex relationships which change dynamically across the book (Lewis, 2006). Images do not solely play decorative roles to grab readers' attention, who are likely to be young children, and texts are not always meant to narrate a story.

Verbal text and images play significant and complex roles in the meaning-making process in picture books (Mikkelsen, 2016; Koutsikou, Christidou, Papadopoulou, & Bonoti, 2021). Various frameworks have been proposed to describe the relationships between texts and images in picture books. Callow (2020) identifies two types of relationships, or "image-language intersections" in his term, including "concurrence" and "complementary". A concurrent relationship refers to a relationship in which verbal texts and images represent the same thing or "express the same idea in different ways". A complementary relationship means that verbal texts and images "extend meanings by augmenting or adding new meanings". In line with Callow (2020), Damayanti, Moecharam, and Asyifa (2021) categorize the relationships between verbal texts and images into two types: complementary, where images and verbal texts support or enhance each other, and divergent relationships, where the two modes convey different meanings. Similarly, images in picture books may not only contextualize but also recontextualize verbal texts, as identified in different text-image configurations in two picture books: Coronavirus and My Hero (Shi, 2023). The same categorization of text-image relationships are also used by Yefymenko (2024) in his analysis of The Paper Bag Princess by Munsch and Snow White in New York by French. He categorizes the relationships as "complementary" and interdependency", suggesting how verbal texts and images creates a synergy. Expanding on this, Nikolajeva and Scott (2006) have categorized five possible relationships between verbal texts and images in picture books, including symmetrical, enhancing, complementary, counterpointing, and contradictory. This categorization is based on the ways in which verbal texts and images share or differ in the information they convey. A symmetrical relationship occurs when verbal texts and images present the same information, reinforcing each other. An enhancing relationship happens when verbal texts and images minimally complement each other, adding subtle layers of meaning. A complementary relationship goes further, with verbal texts and images working together in a more integrated and substantial way to provide a fuller narrative. A counterpoint relationship occurs when verbal texts and images offer contrasting perspectives that work together to express meanings neither could convey alone. In some instances, the contrast becomes stronger, with verbal texts and images directly opposing each other, forming a contradictory relationship.

Despite the fact that scholars have extensively identified the relationships between verbal texts and images in picture books (Callow, 2020; Damayanti, Moecharam, & Asyifa, 2021; Elmiana & Shen, 2024; Hermawan & Sukyadi, 2017; Shi, 2023; Yefymenko, 2024), little attention has been given to examining how these relationships vary across different levels of picture books. This gap is particularly important because many children's picture books incorporate a levelling system to ensure that their contents are developmentally appropriate for readers with varying reading abilities. Existing studies,

however, typically analyze individual books or small sets without considering how these relationships change with intended reader proficiency. For example, Callow (2020) in his research only studied one single picture book, namely The Watertower by Gary Crew and Steve Woolman. Furthermore, despite the fact that Yefymenko (2024) studied two different picture books, text-image relationships were not compared according to the picture book levels. Similarly, another study by Elmiana and Shen (2024), which involved ten picture books, also disregarded the levels of the picture books in relation to the production of their text-image relationships. Hermawan and Sukyadi (2017), who studied the multimodal aspects of three picture books, also paid no attention to the levels of the picture books. This gap leaves an important question unanswered: Do picture books designed for more advanced readers exhibit increasingly complex text-image relationships?

Examining text-image relationships in picture books at different complexity levels is essential. It will shed light on whether existing picture books designed for higher-level readers show an increasing complexity of text-image relationships compared to those aimed at lower-level readers. This raises a broader question about how the complexity of these relationships contributes to the overall difficulty or accessibility of picture books for readers with different reading levels. Traditional conceptualizations of text complexity, however, do not acknowledge this multimodal feature of picture books (Kelly & Kachorsky, 2022). The conceptualizations of text complexity have so far centered around only two views. In one view, it refers to linguistic aspects (Benjamin, 2011), such as lexical and syntactic complexity (Green, C., Keogh, K., Sun, H., & O'Brien, B., 2024; Green, 2025; Zhao, Zhu, Ruiter, and Chen, 2022). In another view, it refers to the relationships between readers and texts (Morris et. al., 2013). The text complexity in the second conceptualization is seen from, for instance, the reader's background knowledge concerning the topics of a text. This was confirmed by an eye-tracking study that students with high prior knowledge have better reading comprehension when reading a long scientific article compared to students with low prior knowledge (Jian, Yu-Cin, 2022).

Following this call, our research expands the conceptualization of text complexity, which consider the relationships between two modalities i.e. verbal and visual texts as offered in picture books. Understanding the complexity of picture books requires considering the interplay between their verbal and visual elements, as both contribute to the overall meaning-making process. Kümmerling-Meibauer & Meibauer (2013) highlighted that when reading a picture book, readers are required to have three developmental cognitive-related abilities to be able to understand a picture book comprehensively, which include the ability to comprehend texts, the ability to understand images, and the ability to interpret the relationships between those two elements. Building upon this reasoning, the multimodal complexity should also be addressed in the levelling system of picture books.

The interaction between texts and images in picture books does not only influence comprehension but also reading pace. The written narrative propels the reader forward, while the visual images encourage them to slow down and engage with details (Serafini & Reid, 2022).

In response to this, the present study aims to identify the text-image relationships of two picture books at different reading complexity levels. Additionally, this study examines whether the two picture books exhibit increasing complexity in terms of their text-image relationships as their level progresses. Research questions formulated in this study are: 1) Does the increasing complexity emerge in two different-level children's picture books? 2) To what extent do the text-image relationships in two different-level picture books differ? To answer these two research questions, two picture books at different levels were selected. The present study used two children's picture books published by the Indonesian Ministry of Education, Culture, Research, and Technology, which can be legally accessed for free via https://buku.kemdikbud.go.id/. These two picture books were designed following criteria mentioned in a guideline book *Pedoman Perjenjangan Buku Nomor 030/P/2022.* While the guidelines ensure that the books meet developmental standards for verbal texts and images, they do not address the interplay or relationships between these two elements.

METHOD

This study applied both qualitative and quantitative approaches to address the research problems. The application of the mixed approach was motivated by the nature of the present study, which demanded a critical and complex analysis to reveal the existence of an increasing complexity of text-image relationships in two different-level picture books. Such an examination required two steps: identifying the categories of text-image relationships and examining whether their complexity of the text-image relationships increases at a higher picture book level.

Identifying categories of text-image relationships is a complex task, as verbal texts and images in picture books interact in nuanced ways. A qualitative approach, in this regard, is capable of doing so, as endorsed by Creswell and Poth (2018), stating that a qualitative approach is particularly suitable for "a complex, detailed understanding of the issue". This task also relies on a flexible interpretation of multimodal texts. Dörnyei (2007) argue that the qualitative approach is well-suited for a flexible analysis that investigates "the subtle nuances of meaning". However, a purely qualitative approach is inadequate. It is due to the fact that the current study is also demanded to examine whether the complexity of text-image relationships develops in a parallel with increasing picture books levels. This requires quantitative work to compare the distribution of categories of text-image relationships across the picture books. As suggested by Edmonds and Kennedy (2017), quantitative approach is a well-suited approach to address this need.

By integrating two approaches, the present study offers a comprehensive understanding of the research problems. While qualitative analysis was applied to analyse different types of text-image relationships, quantitative analysis was used to quantify the spread of frequency for each category. By integrating these two approaches, this mixed approach ensured that the findings were both loaded with an in-depth interpretation supported by statistical evidence.

Thirty-four pairs of verbal texts and images were obtained from two picture books at different levels, which include *Gambar Lucu Mika* (Mika's Cute Pictures), or GLM for short, authored by Tyas Widjati and illustrated by Faza (2022) as well as *Naik-Naik ke Puncak Bukit* (Up the Hill We Go!), or NNPB for short, authored by Sarah Fauzia and illustrated by Alima Nufus (2022). These two picture books are among thirty-six picture books released by the Indonesian Ministry of Education, Culture, Research, and Technology. These picture books are systematized into seven developmental reading categories, which are emergent readers, early readers, intermediate readers, advanced readers, and skilled readers. The emergent-readers level is intended for an audience of children below 8. The early-readers level is designed for readers aged 6 to 10. The intermediate-readers level is aimed at readers aged 10 to 13. The advanced level targets children aged 13 to 15, and the skilled-readers level is designed for readers aged 16 and above. Each level consists of four books, except for the advanced level and the skilled-reader level, which each level has eight books.

The data used in this study were chosen according to two main criteria, including the targeted readers' age and the availability of text-image relationships in the picture books. Considering the current study's objective aimed at studying children's picture books, picture books that are not intended for children were eliminated. According to the *UU Perlindungan Anak* (Indonesian Child Protection Law), readers under the age of 10 are considered as children, while those aged 10 and above are categorized as teenagers according to the law. As a consequence, this study randomly selected two picture books within these following reading levels: the emergent readers and the early readers. Furthermore, the current study only included data that show a relationship between text and images. Consequently, elimination of any pages that posses either text or images only were required. Following these criteria, a total of thirty-four text-image pairs were collected, consisting of thirteen text-image pairs from GLM and twenty-one text-image pairs from NNPB.

Our data analysis was conducted in two phases: qualitative and quantitative phases. In a qualitative phase, the analysis was conducted in both inductively and deductively applying the following theories: Halliday's systemic functional linguistics (2004), Kress and Leeuwen's multimodal approach (2021), and Nikolajeva and Scott's (2006) approach to picture books. In the inductive process, Halliday's transitivity system (2004) and Kress

and Leeuwen's multimodal approach (2021) were used to analyse linguistic and visual features that emerged in both texts and images. This bottom-up process allows more flexible analysis without relying too much on the existing theories or framework, which reveals subtle nuances in the data. Afterwards, in the deductive, top-down process, findings from the inductive phase were deductively classified into categories that are already offered by Nikolajeva and Scott (2006). This inductive approach was necessary, as it helped explore how text and images collaborate to make meaning in the picture books (Fife & Gossner, 2024).

Along with qualitative analysis, this study also applied quantitative approach to gauge the occurrence distribution of all text-image relationship categories in the picture books. After all data were qualitatively classified using Nikolajeva and Scott's framework (2006), each finding was quantified. This numerical data allowed for a clearer comparison of how frequently certain types of text-image relationships appeared in different picture books and at varying complexity levels. Descriptive statistic was used to help describe "general tendencies in the data" (Dörnyei, 2007), highlighting whether higher-level picture books exhibited more diverse or complex text-image relationships.

This study applied Nikolajeva and Scott's categorizations (2006) to classify text-image relationships. The categorization is comprised of five text-image relation categories, including symmetrical, enhancing, complementary, counterpointing, and contradictory. This categorization is based on the similarity the verbal texts and images share, as demonstrated in Figure 1. In symmetrical relationships, texts and images tell the same story. In enhancing relationships, verbal texts add more details to the images, or images help explain the meaning of the verbal texts. When the enhancing relationship becomes very important in the story, the relationship turns to be complementary. Depending on the different information presented, a counterpoint dynamic may occur, where words and images work together to convey meanings that neither can express alone. In an extreme case, there is a contradictory counterpoint, where words and images seem to oppose each other.

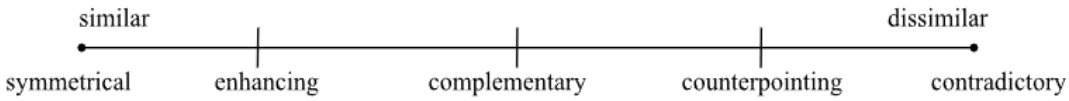

Figure 1. Nikolajeva and Scott's text-image relationships in picture books

RESULTS AND DISCUSSIONS

Results

The results are presented quantitatively to illustrate the distribution of text-image relationships across the picture books. Absolute values with percentages are displayed side by side to help compare each finding. This method allows for a clearer understanding of how different types of text-image relationships are distributed across varying levels of complexity in the books.

Applying Nikolajeva and Scott's approach (2006), the analysis initially included five categories of text-image relationships: symmetrical, enhancing, complementary, counterpointing, and contradictory. However, only four categories were identified in both picture books, indicating that not all relationships are present at every reading level. The absence of one category suggests that certain complexities in text-image interaction emerge only at higher levels. A detailed distribution of the text-image relationships for each category from the two picture books is provided in Table 1 below.

Table 1. Distribution of text-image relationships in GLM and NNPB

Categories	GLM		NNPB	
	Absolute values	Values in percentages	Absolute values	Values in percentages
Symmetrical	0	0%	0	0%
Enhancing	7	54%	8	38%
Complementary	5	38%	6	29%
Counterpointing	1	8%	5	24%
Contradictory	0	0%	2	9%
Total	13	100%	21	100%

Discussion

The data analysis revealed that the lower-level picture book (GLM) shows different text-image relationships to the higher-level picture book (NNPB), as displayed in Table 1 above. Those numerical data in Table 1 show an increasing complexity from lower-level picture book to the higher-level one. Three types of text-image relationships were found in GLM, which include enhancing, complementary, and counterpointing relationships, while four types were identified in NNPB, namely enhancing, complementary, counterpointing, and contradictory relationships. The absence of contradictory relationships in GLM suggests that simpler picture books rely more on straightforward connections between text and images, whereas more advanced books introduce greater interpretative challenges.

These findings highlight the progression of multimodal complexity in picture books as the level advances. The following sections provide an in-depth analysis of these text-image relationships at both picture book levels. Additionally, recommendations for future research will be presented.

Text-image Relationships in a Lower-level Picture book

In a lower-level picture book, three types of text-image relationships are identified, including enhancing, complementary, and counterpointing relationship. Enhancing and complementary relationships in GLM are found to be more frequent than the same types of relationships shown in NNPB. These two types of relationships demand lower cognitive efforts to actively use imagination since the differing features within verbal and visual texts

are not significant, consequently facilitating easier interpretations. Figure 2 reinforces this idea, showing what is displayed in the image shares high similarity to what is narrated in the verbal text.

Additionally, another differing characteristic shown in the GLM is its inclusion of fewer verbal texts compared to the NNPB. Simpler linguistic and visual structures in picture books help young readers comprehend the picture books better (Hu and Qiu, 2020; Karatza, 2020). This aligns with findings from studies on vocabulary, metacognitive knowledge, and task orientation as predictors of narrative picture book comprehension, which emphasize that younger readers benefit from simpler text-image relationships (Lepola, Kajamies, Laakkonen, & Niemi, 2020). A study by Haris, Febrianti, and Yannuar (2023) suggests that complementary and enhancing relationships, in their term "collocation", help children facilitate comprehensions. A study examining the eye movements of 115 children aged 3 to 6 years found that younger children spent time more on images than they did on verbal texts (Li, Martin, Baogen, & Xiaomei, 2020). This finding indicates that young children rely very much on images to comprehend a story better. This study also found that as the children grow older, they acquire a capability to focus on more informative aspects such as verbal texts and participants—important parts of the pictures which include information related to people, places, or things in the story— to comprehend a story better. To explore how verbal text-image relationships manifest in the GLM, the following sections provides a in-depth descriptive analysis.

Enhancing Relationship

As shown in Figure 2, the example data taken from the GLM picture book shows a minimal enhancement, implying that it does not significantly influence the narrative. It is shown that the image adds details to the verbal texts. In the verbal texts, there are only three participants mentioned, namely Kak Gio, Mika, and *gambar lucunya* [its funny pictures]. Meanwhile, the image shows richer represented participants: Kak Gio, Mika, ornamental plants, funny pictures, and a huge smartphone.

The image adds information to the verbal text concerning the skin and hair colours of the main participants, Kak Gio and Mia, which may help elicit their racial identity. Additionally, the verbal texts do not provide circumstantial elements. The represented participants, ornamental plants, expand the textual meaning concerning the settings where the activity takes a place, which is home. These informational details are totally absent in the verbal texts. Furthermore, the image provides spatial relations that clarify the positioning of the characters and objects within the scene. For example, the smartphone's size in relation to the other elements may suggest its significance in the interaction. The inclusion of additional objects, such as the plants, also enriches the visual representation, creating a more immersive depiction of the environment that the verbal text alone does not convey.

Figure 2. The GLM's example data of an enhancing relationship

Complementary Relationship

Figure 3 shows the example data of complementary relationship in the GLM picture book. It is marked by the existence of participant-process-circumstance configuration in the image that significantly complements the verbal text. For instance, the existence of the represented participant, Kak Gio, seems to be neglected in the verbal text; however, the image shows it with a vector, marked by a black arrow, indicating a process of pointing at.

This participant-process configuration complements the circumstantial element *sendiri saja* [by herself] in the verbal text. It reveals that Mika is not really drawing by herself; there is an indirect involvement from another participant. The image introduces an additional participant, Kak Gio, who is sitting next to Mika while engaged with a phone. This detail is absent in the verbal text, which only focuses on Mika's independent activity. Kak Gio's relaxed posture and attention toward the phone contrast with Mika's engagement in drawing, visually reinforcing the theme of independence implied by the text. The visual elements enhance the reader's understanding by depicting Mika's solitary effort, not as complete isolation, but within a shared space where attention is divided. Additionally, the illustration

provides visual cues that suggest the setting is a home environment. The presence of a patterned floor, a small table, and Mika sitting on the ground indicate an informal and comfortable indoor space, which the text does not mention. These details help the reader understand where the activity is taking place, reinforcing the domestic atmosphere.

Eliminating either the image or the verbal text will lead to misunderstandings, showing a highly dependent complementary relationship between the verbal and the visual text. This interdependence suggests that meaning-making in the GLM picture book relies on the synergy between the two modes of communication, rather than one functioning independently of the other.

Figure 3. The GLM's example data of a complementary relationship

Counterpointing Relationship

As shown in Figure 4, the counterpointing relationship is reflected in the way the image and the verbal text provide alternative information. The visual text displays a narrative representation, which is characterized by two eyeline vectors: one directed toward the smartphone and the other pointing to the unknown point. Mika's eyeline vector toward the smart-

phone suggests that she is focused on the emojis, which indicate interest. On the contrary, the verbal text shows disagreement with the image, stating that Mika does not find the pictures funny [*Gambar ini tidak lucu*]. This conflicting scene highlights a counterpointing relationship, where the verbal and visual elements do not directly support each other, but rather produce a more complex understanding. It implies that Mika's emotional response is different from what was expected although the smartphones attracted her attention. This, consequently, creates multiple interpretations.

This relationship between the image and verbal text makes interpretation more complicated, which demands the viewer/reader to manage conflicting elements. The visual element on its own might indicate enthusiasm or excitement; the verbal one, however, evidently says otherwise. This counterpoint shows an implied complexity of human emotions, where outward gestures does not always agree with inward emotions. In this context, paying attention does not indicate enthusiasm or enjoyment. By pointing out this difference, the interaction between text and image encourages readers to think critically about how meaning is built in multimodal storytelling, especially when the two modes show counterpointing perspectives. Furthermore, this contrast emphasizes the different layers of meaning between what is seen and what is felt, making the relationship between words and pictures more interesting. In doing so, the narrative not only improves character development but also challenges readers to connect different modes of meaning within the story.

Figure 4. The GLM's example data of a counterpointing relationship

Text-Image Relationships in A Higher-level Picture book

More varied text-image relationships are identified in a higher-level picture book. The study found four different types of relationships, including enhancing, complementary, counterpointing, and contradictory relationships. Nikolajeva and Scott (2006) argue that the greater the divergence between verbal text and images in conveying the same ideas, the more space is created for readers' imagination, potentially demanding a higher level of cognitive engagement from readers.

In a higher-level picture book (NNPB), where counterpointing (24%) and contradictory (9%) relationships between text and image are more prevalent, the interaction between the two elements creates a more cognitively demanding reading experience. This interplay requires readers to actively navigate between explicit textual information and the implied or contrasting meanings in the visuals. This implies that the NNPB is indeed ideal for readers who have higher cognitive abilities that allow them to use them to understand the stories in the picture books. A previous study by Pike, Barnes, and Barron (2010) suggested that the complexity of text-image relationships determine children's comprehension. In addition, another study suggested that the ability of comprehending multimodal elements in picture books develops more effectively in older children (Shimek, 2021). The following elaboration shows how the relationships between verbal and visual elements in the higher-level picture book (NNPB).

Enhancing Relationship

The sample data in Figure 5 shows an enhancing relationship in the NNPB picture book. While the verbal text narrates the story, the visual text contributes to creating a dramatic scene. Such a scene is realized in action vectors marked by the black arrows, as shown in Figure 5. Furthermore, the red zigzag lines serve as a visual cue for Sabit's discomfort. These lines are often used in illustrations to represent pain, noise, or intense emotions. Their sharp, uneven shape visually reinforces the overwhelming sensory experience, emphasizing the idea that Sabit is struggling with external stimuli. The combination of verbal and visual elements enhances the story by making Sabit's emotions more vivid and immediate for the readers.

Additionally, the image effectively reinforces the text's message that Sabit needs a calmer environment. His closed eyes, tense posture, and the way he covers his ears visually depict his discomfort, aligning with the verbal text's claim that he requires a peaceful setting [*Sabit perlu suasana tenang*]. The presence of Bulan reaching out to him further supports the text's implication that she is trying to help by suggesting they go back home [*Bulan mengajak Sabit pulang*]. This interaction creates a clear cause-and-effect relationship between Sabit's distress and Bulan's action, making the narrative more dynamic. Instead of merely illustrating the words, the image deepens the reader's understanding of Sabit's emotional state. By enhanc-

ing the verbal text with expressive imagery, the visual elements strengthen the storytelling, which make the scene more emotionally engaging.

Figure 5. The NNPB's example data of an enhancing relationship

Complementary Relationship

Figure 6 shows how verbal and visual texts complement each other to create meanings. The verbal text gives clear details that the visual text does not directly show, while the visual text adds extra information that helps explain the verbal text. One example is the phrase *dengan segera* [quickly] in the verbal text, which describes how an action is done. This detail about speed is missing from the image, so without the words, readers or viewers might not realize the urgency of the situation. In addition, the verbal text implies that Sabit is recovering from a previous event that requires him to get up quickly. This could indicate that he has fallen, been startled, or suddenly realized something important. However, the image does not visually show what led to this moment. There are no clear action lines, impact marks, or distressed expressions that directly show a prior fall or incident, leaving the cause open to interpretation.

At the same time, the visual text contributes additional meaning that words alone might not fully convey. While verbal text provides the narrative structure, images enhance understanding by adding context, emotion, and detail. In this case, the image features an extended hand reaching for a flower, which helps clarify the scene's significance. The verbal text may suggest urgency or struggle, but the visual element reveals a deeper layer of meaning. Specifically, it indicates that the main character, Sabit, is not entirely alone in this difficult moment. The presence of an extra participant in the visual text subtly alters the viewer's perception of the story, suggesting that someone else might be involved. This interplay between text and image was categorized as having a complementary relationship, where

both elements work together to shape meaning, influence interpretation, and enhance the emotional depth of the narrative.

Figure 6. The NNPB's example data of a complementary relationship

Counterpointing Relationship

The participant-process-circumstance configurations in the verbal text and the image in Figure 7 show a counterpointing relationship. Most of the process types emerging in the verbal text are characterized as mental processes. In the multimodal approach, this type of process is equivalent to a narrative representation. However, the visual text does not show any clear actions. There are no action lines or movement indicators (vectors) that suggest something is happening. According to Kress and Leeuwen's visual grammar, this kind of image is considered a static conceptual image. Instead of showing an event, it presents a more general or symbolic meaning. The difference between the mental processes in the verbal text and the lack of action in the image creates a contrast, emphasizing the counterpointing relationship between the two modes.

Additionally, the counterpointing relationship is realized in the contrast

between the verbal text's emphasis on preparation and protection and the image's static representation of the characters. The verbal text conveys urgency through the phrase *bersiap-siap dengan segera* [preparing quickly], which implies an immediate action. However, the visual text lacks dynamic elements, as the characters stand still, with no clear movement shown. The represented participants stand upright with relaxed postures, and their facial expressions do not indicate any rush. This lack of visual dynamism creates a contrast between what is described in words and what is depicted in the image. This indicates the counterpointing relationship of the two modes.

Figure 7. The NNPB's example data of a counterpointing relationship

Contradictory Relationship

Figure 8 presents a contradictory relationship between the verbal and visual texts. As shown in Figure 8, the process mentioned in the verbal text contradicts with the one presented in the image. According to the text, the participant, Sabit, is mentioned to echo the other participant, Bulan. Such an activity is found in these two processes in the verbal text: *menirukannya* [copied it] and *melakukan ekolalia* [did echolalia] – echolalia, by definition, is a medical condition where an individual uncontrollably repeats the words that somebody else has just said. These two processes indicate Sabit's ac-

tive participation in response to Bulan's utterances. In contrast, the image shows a conflicting narrative. In the image, the activity of talking is absent, which is visually represented by the Sabit's closed mouth. In addition to this, while it is mentioned verbally that Bulan says *tenang... tenang* [calm down... calm down] to comfort Sabit, Bulan visually shows no processes, which indicates silence rather than speech.

Furthermore, contradiction performed by the other characters, Ayah and Bulan, make the story even more difficult to comprehend. It is verbally mentioned that those two characters sighed in relief, as mentioned in this following line *Ayah dan Bulan pun bernapas lega* [Father and Bulan sighed in relief], their gestures do not coherently demonstrate that they are at peace. The father's stiff postures suggest that he is concerned, and Bulan's hand gesture suggests engagement rather than visible comfort. These visual details oppose the verbal statement that the condition has become stable, highlighting the contradiction between the verbal texts and images. This interplay invites readers to critically question how much Sabit is actually participating in the verbal interaction, whether his echolalia is occurring internally rather than outwardly, and whether the sense of resolution described in the text is really reflected in the image.

Figure 8. The NNPB's example data of a contradictory relationship

Recommendations for Future Studies

This study contributes to the field of picture book research, especially on its developmental aspects of text-image relationships, which, to our knowledge, remains underexplored. For a more comprehensive understanding, however, further exploration within this research topic needs to be undertaken. Future studies are encouraged to include a larger sample of picture books to compare since the current study focuses only two picture books. Including bigger dataset will hopefully lead to more comprehensive and insightful results. Additionally, future studies are encouraged to carefully select the picture books by setting a limitation to picture books that only share the same narrative point of view. The picture books that were included in the current study have different narrative perspectives: one uses a first-person point of view and another one uses a third-person point of view. By restricting the picture book criteria, more accurate findings can be achieved.

Moreover, future research is also encouraged to investigate how children at different age and reading levels interpret and integrate text-image relationships in picture books. A study revealed that children aged 5-6 years old make an effort to make sense of the interplay between verbal and visual information, often treating them as two unrelated elements rather than integrated ones (Wang, Su, & Zheng, 2023). This finding implies that multimodal literacy skills are developmentally acquired and may need explicit instructional interventions. To gain deeper insights into young readers' ability to recognize and process different multimodal relationships (enhancing, complementary, counterpointing, contradictory), future studies could employ eye-tracking or think-aloud protocols. These methods would help researchers examine real-time cognitive processing and engagement, shedding light on the mechanisms underlying children's multimodal literacy development. Additionally, longitudinal studies could track how children's ability to integrate visual and textual information evolves over time, providing a more nuanced understanding of how multimodal literacy skills emerge and progress. Studies in this field could also help educators understand better on effective instructional strategies that consequently improve children's ability to comprehend the interdependent relationships of text and images in picture books, which ultimately facilitate greater engagement.

CONCLUSION

A levelling system is used in children's picture books as a way to address the needs of young readers at different stages of reading development. This system helps them get exposed with texts at an appropriate level and stimulates them to be engaged with more complex texts as their reading skills become more advanced. Although many studies have sought to understand how linguistic aspects are accommodated within this system, few studies have been conducted to examine how they consider multimodal features of picture books. This defies the fact that picture books are composed

of verbal and visual elements. Understanding how text-image relationships work across different complexity levels is important because it clarifies how multimodal complexity is integrated into the picture books. Exploring this topic provides insights into how picture books facilitate young readers' comprehension and reading engagement.

The present study found that two picture books at different complexity levels, "Gambar Lucu Mika" (Mika's Cute Pictures) written by Tyas Widjati and illustrated by Faza and "Naik-Naik ke Puncak Bukit" (Up the Hill We Go!) written by Sarah Fauzia and illustrated by Alima Nufus, demonstrated a growing complexity as the levels progress. A higher-level picture book was found to have more diverse types of text-image relationships compared to the lower-level one. The study found that the lower-level one shows only three types of text-image relationships, which include enhancing, complementary, and counterpointing relationship. The higher-level one, however, was identified to have four, including enhancing, complementary, counterpointing, and contradictory relationships. These findings imply that higher-level picture books are developed to engage young readers with more cognitively demanding multimodal interactions. On the other hand, more simplified text-image relationships in a lower-level picture book are developed to help early readers, who might not be able to comprehend the interdependent relationships of the two modes, in engaging with the picture book without being overwhelmed with complex multimodal cues.

ACKNOWLEDGMENTS

We would like to thank the Book Directorate of the Ministry of Primary and Secondary Education of the Republic of Indonesia for giving permission to use their sources. Simultaneously, we also appreciate our colleagues and reviewers for providing meaningful and comprehensive feedback to ensure the quality of this article.

REFERENCES

Al-Yaqout, G., & Nikolajva, M. (2015). Re-conceptualising picturebook theory in the digital age. *Barnelitterært Forskningstidsskrift*, *6*(1), 1-7. https://doi.org/10.3402/blft.v6.26971

Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2022). *Pedoman perjenjangan buku*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. https://buku.kemdikbud.go.id/

Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, *24*(4), 63–88. https://doi.org/10.1007/s10648-011-9181-8

Callow, J. (2020). Visual and verbal intersections in picture books–multimodal assessment for middle years students. *Language and Education*, *34*(2), 115–134. https://doi.org/10.1080/09500782.2019.1689996

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry research design: Choosing among five approaches (fourth edition)*. SAGE

Damayanti, I. L., Moecharam, N. Y., & Asyifa, F. (2021). Unfolding layers of meanings: Visual-verbal relations in Just Ask—a children's picture book. *Indonesian Journal of Applied Linguistics*, *11*(2), 372–381. https://doi.org/10.17509/ijal.v11i2.39195

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed*

methodologies. Oxford University Press

Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods*. SAGE

Elmiana, D. S., & Shen, S. (2024). Let pictures explain the world: Text–image narration in picturebooks. *Journal of Research in Childhood Education, 38*(4), 1-14. https://doi.org/10.1080/02568543.2024.2367406

Fauzia, S., & Nufus, A. (2022). *Naik-naik ke puncak bukit*. Pusat Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. https://buku.kemdikbud.go.id/katalog/naik-naik-ke-puncak-bukit

Fife, S. T., & Gossner, J. D. (2024). Deductive qualitative analysis: Evaluating, expanding, and refining theory. *International Journal of Qualitative Methods, 23,* 1-10. https://doi.org/10.1177/16094069241244856

Green, C. (2025). Multi-word expressions occur infrequently in picturebooks: Implications for early vocabulary instruction. *Literacy Research and Instruction*, *4*(2), 1-18 https://doi.org/10.1080/19388071.2025.2469068

Green, C., Keogh, K., Sun, H., & O'Brien, B. (2024). The children's picture books lexicon (CPB-Lex): A large-scale lexical database from children's picture books. *Behavior Research Methods*, *56*(5), 4504–4521. https://doi.org/10.3758/s13428-023-02198-y

Halliday, M. A. K., Matthiessen, C. M. I. M. (2014*). Halliday's introduction to functional grammar: Fourth edition*. Routledge

Haris, N. A., Febrianti, Y., & Yannuar, N. (2023). Exploring augmentation of meaning through intersemiotic complementarity in children comic book series. *Indonesian Journal of Applied Linguistics*, *12*(3), 739–751. https://doi.org/10.17509/ijal.v12i3.39951

Hu, Y., & Qiu, Q. (2020). A study on verbal and image relations in multimodal texts from the perspective of intersemiotic complementarity. *Canadian Social Science*, *16*(10), 50–56. http://doi.org/10.3968/11938

Jian, Y. C. (2022). Using an eye tracker to examine the effect of prior knowledge on reading processes while reading a printed scientific text with multiple representations. *International Journal of Science Education*, *44*(8), 1209–1229. https://doi.org/10.1080/09500693.2022.2072013

Karatza, S. (2020). Multimodal literacy and language testing: Visual and intersemiotic literacy indicators of reading comprehension texts. *Journal of Visual Literacy*, *39*(3–4), 220–255. https://doi.org/10.1080/1051144X.2020.1826222

Kelly, L. B., & Kachorsky, D. (2022). Text complexity and picturebooks: Learning from multimodal analysis and children's discussion. *Reading and Writing Quarterly*, *38*(1), 33–50. https://doi.org/10.1080/10573569.2021.1907636

Koutsikou, M., Christidou, V., Papadopoulou, M., & Bonoti, F. (2021). Interpersonal meaning: Verbal text–image relations in multimodal science texts for young children. *Education Sciences*, *11*(5), 245. https://doi.org/10.3390/educsci11050245

Kress, G., & Leeuwen, v. T. (2021). *Reading images: The grammar of visual design*. Routledge

Li, L., Martin, C., Baogen, L., & Xiaomei, G. (2019). Moving eyes on pictures following visual grammar benefits meaning making: Evidence from the independent reading of Chinese preschool children. *Journal of Chinese Writing Systems, 4*(1), 57-70. https://doi.org/10.1177/2513850219886109 (Original work published 2020)

Lewis, D. (2006) *Reading contemporary picture books: Picturing text*. Routledge

Lepola, J., Kajamies, A., Laakkonen, E., & Niemi, P. (2020). Vocabulary, metacognitive knowledge and task orientation as predictors of narrative picture book comprehension: From preschool to grade 3. *Reading and Writing*, *33*, 1351–1373. https://doi.org/10.1007/s11145-019-10010-7

Mikkelsen, M. (2016). Picture books as crossover literature: A study of how readers of different ages perceive iconotext and themes in picture books. [Unpublished Master thesis]. Western Norway University of Applied Sciences

Morris, D., Trathen, W., Frye, E. M., Kucan, L., Ward, D., Schlagal, R., & Hendrix, M. (2013). The role of reading rate in the informal assessment of reading ability. *Literacy Research and Instruction*, *52*(1), 52–64. https://doi.org/10.1080/19388071.2012.702188

Nikolajeva, M., & Scott, C. (2006). *How picture books work*. Routledge

Pemerintah Indonesia. (2014). *Undang-undang (UU) Nomor 35 Tahun 2014 tentang Perubahan atas Undang-Undang Nomor 23 Tahun 2002 Tentang Perlindungan Anak*. https://peraturan.bpk.go.id/Details/38723/uu-no-35-tahun-2014

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, *105*(3), 243–255. https://doi.org/10.1016/j.jecp.2009.10.006

Ramos, A. M. (2020). Picturebook format: Beyond the relationship between words and pictures an overview of the portuguese editorial panorama. In *Libri et Liberi*, *9*(1), 61–74. Croatian Association of Researchers in Children's Literature. https://doi.org/10.21066/CARCL.LIBRI.2020.1.4

Serafini, F. (2024). The complex relationship of words and images in picture books. *Journal of Visual Literacy, 43*(3), 233–249. https://doi.org/10.1080/1051144X.2024.2394338

Serafini, F. (2023). How multimodality matters in children's literature scholarship. *Australian Journal of Language and Literacy*, *46*(3), 245-256. https://doi.org/10.1007/s44020-023-00046-2

Serafini, F., & Reid, S. F. (2022). Analyzing picturebooks: Semiotic, literary, and artistic frameworks. *Visual Communication, 23*(2), 330-350. https://doi.org/10.1177/14703572211069623

Shi, D. (2023). Intermodality and visual literacy: Exploring visual-verbal instantiation in children's picture books on coronavirus. *Journal of Visual Literacy*, *42*(3), 183–209. https://doi.org/10.1080/1051144X.2023.2258741

Shimek, C. (2021). Recursive readings and reckonings: Kindergarteners' multimodal transactions with a nonfiction picturebook. *English Teaching: Practice & Critique*, *20*(2), 149-162. https://doi.org/10.1108/ETPC-07-2020-0068

Wang, D., Su, M., & Zheng, Y. (2023). An empirical study on the reading response to picture books of children aged 5–6. *Frontiers in Psychology*, *14, 1-10.* https://doi.org/10.3389/fpsyg.2023.1099875

Widjati, T., & Faza. (2022). *Gambar lucu Mika.* Pusat Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. https://buku.kemdikbud.go.id/katalog/gambar-lucu-mika

Yefymenko, V. (2024). Multimodal text-image synergy in representing interpersonal relations in picture books. *Cognition, Communication, Discourse*, *28*, 102–109. https://doi.org/10.26565/2218-2926-2024-28-07

Zhao, J., Zhu, M., de Ruiter, L., & Chen, S. (2022). Linguistic indicators for text complexity in picture books for young Chinese children learning English as a foreign language. *Frontiers in Education*, *7, 1-11.* https://doi.org/10.3389/feduc.2022.758736