

# A Multimodal Discourse Analysis of English Textbooks

-- A Comparative Study of Images in Old and New PEP 7A

Haonan Lin \*

School of International Studies, Hangzhou Normal University, Hangzhou, Zhejiang, China

\* Corresponding author Email: linhaonan126@gmail.com

---

**Abstract:** The reform of information technology is gradually blurring the boundaries between different modalities such as text, image, sound and video. Instead, multimodal ways are always chosen to convey information, with no exception of English textbooks. However, multimodal research on English textbooks has shown that there are some problems regarding images used, such as the over simplicity of the content of images, the monotonous forms of images, and the rigid functions of images. The new English Curriculum Standards, however, call for the development of students' multimodal literacy and emphasize attention to multimodal resources in textbooks. In view of the implementation of new PEP 7A junior English textbook, this study adopted Kress and van Leeuwen's Visual Grammar, and took the images as the main research object, to observe and compare the old and new PEP junior English textbooks, to interpret the inter-semiotic relationship between images and text, and to examine how these change align with the new English Curriculum Standards.

**Keywords:** Multimodal Discourse Analysis; English Textbook Analysis; Visual Grammar.

---

## 1. Introduction

English textbooks have incorporated non-verbal resources like images, but the rationale behind their use and the principles guiding their design have received little attention in academic research [1]. Despite the long-standing integration of visual elements, there is no systematic theoretical framework underpinning the design and use of images in textbooks. As information technology evolves, the presentation of modern information has diversified, moving beyond single mediums to encompass multiple modalities, including text, images, and sound. Consequently, multiliteracy, or multimodal literacy, has emerged as a crucial component of communicative competence [2]. Students now need the skills to interpret and communicate effectively using images, highlighting the need to cultivate foreign language learners with strong multiliteracy skills in China's new century [3].

In this evolving context, middle school English textbooks in China serve as multimodal artifacts that combine linguistic symbols with visual resources. Although many domestic textbook authors have made efforts to enhance the use of visual resources—such as increasing the number of color illustrations and expanding their functions—they often lack expertise in effective visual design, leading to various shortcomings [4]. Gao [5] points out several prevalent issues: the content of images is often overly simplistic, their forms are monotonous, and their functions are rigid and limited.

With the introduction of the 2022 English Curriculum Standards, the standards for textbook content have become more demanding. The new Standards emphasize the need for students to develop language proficiency by “recording key information, understanding, and summarizing main discourse content during listening, reading, and viewing” [6]. Additionally, they underscore the complex relationships between elements of discourse, including sentences, paragraphs, headings, and the interplay between text and diagrams [6]. Recognizing that some content and formats in

older textbooks are outdated and fail to meet contemporary teaching and learning needs, new editions have been introduced to address these gaps. This study adopted Kress and van Leeuwen's Visual Grammar, and took the images as the main research object, to observe and compare the old and new PEP junior English textbooks, to interpret the inter-semiotic relationship between images and text, and to examine how these change align with the new English Curriculum Standards.

## 2. Theoretical Framework

### 2.1. Multimodal Discourse Analysis

Compared to traditional discourse, multimodal discourse is more complex, and definitions given by different scholars can be broadly divided into two categories: narrow meaning and broad meaning. The broad definition holds that any ideational unit that uses more than two modalities for communication can be called multimodal discourse, and such ideational units may or may not be written. For example, Kress and van Leeuwen [7] argue that discourse using multiple social sign systems simultaneously is multimodal discourse. Similarly, Zhang [8] states that a communicative unit using multiple senses—such as hearing, sight, and smell—or multiple symbol systems—such as images, language, and typography—or both, can be called multimodal discourse. The narrow definition, on the other hand, is still restricted to the written form, considering multimodal discourse as written discourse that uses one or more other modalities, such as images and symbols, in addition to text. This view is supported by Li [9], who argues that a multimodal discourse is a composite discourse containing both text and other symbol systems such as images and diagrams, or a text that achieves meaning through two or more symbol systems. Yang and Miao [10] also pointed out that usually, in addition to language, images, fonts, formatting, typography, are involved in conveying meaning, thus classifying all such texts as multimodal discourse. Since this study is a comparative

analysis of textbooks and focuses only on paper-based textbooks, it adopts the narrow definition that multimodal discourse is written discourse containing two or more symbol systems. Symbolic systems here include various types, such as text, images, and typography, with text typically being the required symbol system.

Based on this, the focus of multimodal discourse analysis is to extend the analysis of discourse to other social semiotic systems besides text, interpreting all semiotic systems as meaning-generating resources. Xu and Qu [11] point out that to achieve this kind of interpretation, we need to understand the principles of constructing meaning within different semiotic systems and study the meaning expressed by each. In other words, the interaction between symbols is also worth observing, and that multimodal discourse analysis should pay attention to how symbol systems collaborate within the discourse to construct its overall meaning. The image research conducted in this study moves beyond the traditional modality of text to explore the relationship between the two symbol systems of images and text, observing how they construct meaning within the discourse, which is an aspect of multimodal discourse analysis.

## 2.2. Visual Grammar

Visual Grammar, a classic theory for analyzing images in multimodal discourse, is chosen as the framework of this study. It is primarily based on Halliday's view of language, especially his metafunctions of language, which include ideational function (expressing personal experience and inner activities), interpersonal function (communicating identity, status, attitudes, and evaluations), and textual function (organizing and transmitting information) [12]. Zhang and Jia [13] highlight two key aspects: Visual Grammar presupposes the social semiotics and meaning potential of language as proposed by Halliday. Halliday [14] argues that language is a social signifier and a system of meaning potentials, suggesting that language functions as a resource for creating meaning rather than just a set of rules. This view expands the idea that besides language, other symbols, including images, sounds, and colors, also express meaning. Consequently, the social symbolic nature of language has led linguists to explore non-verbal symbols, opening the door to interpreting images [13]. Li [9] argued that while these metafunctions were initially proposed for language, they are not limited to it. Kress and van Leeuwen [7] focused on extending these metafunctions into the visual modality, and built on Halliday's metafunctions to develop Visual Grammar, positing that images, as social symbols, also hold the potential to convey meaning.

Based on Halliday's three metafunctions from Systemic Functional Grammar, Kress & van Leeuwen [7] proposed three corresponding types of meaning in Visual Grammar: representational meaning, interactive meaning, and compositional meaning. Representational meaning refers to the processes and meanings that an image primarily conveys; interactive meaning involves how an image engages with its potential viewers; compositional meaning pertains to how an image structures its elements and aligns with accompanying text [7].

To better interpret metafunctions of Visual Grammar, some key concepts need to be expounded. In terms of the representational function, images can exhibit two types of processes: narrative representation and construct representation. Narrative representation depicts a process of

change, showcasing a developing action or event, while construct representation portrays a stable and generalized static meaning [7]. The interactive function includes subordinate dimensions such as distance, perspective, and modality. Distance relates to the size ratio of elements within the image and their relation to the image border, reflecting the social relationship between the image and the observer. Horizontal perspective denotes the level of involvement and vertical one represents power. Modality relates to the realism of the image. To be more specific, images with high color saturation and detailed backgrounds tend to have more modality compared to those with low saturation and minimal backgrounds [7]. The compositional function includes three subordinate dimensions: information value, salience, and framing. Information value refers to the placement of elements within the image, salience indicates how much an image captures the observer's attention, and framing involves how the image and text are divided. Kress and van Leeuwen [7] emphasize analyzing multimodal discourse holistically rather than separating images and text. Therefore, their compositional function accounts for the coherence between images and text, and its subordinate dimensions can be used to analyze the visual relationship between these two social semiotic systems.

Compared to other multimodal analysis theories, Zhang and Jia [13] highlight several advantages of visual grammar theory. First, its range of applicable research objects is broad, allowing for the analysis of all types of images. Second, visual grammar theory integrates Halliday's three major metafunctions with the inherent characteristics of images, offering a more coherent framework. Each metafunction includes a rich set of analytical dimensions, which provides strong logical structure. Therefore, this approach supports both the qualitative analysis of individual images and the quantitative classification of large amount of image datasets.

## 3. Literature Review

### 3.1. Multimodal Research on Images in English Language Textbooks

In the past decade, multimodal research in English textbook analysis has flourished, with numerous studies on image features emerging both domestically and internationally. However, definitions of images vary across different studies in this field. Yang and Wu [15] define images as visual representational materials in textbooks, including pictures, diagrams, and videos. Chen and Cong [1] view images as non-textual resources perceived visually, specifically distinguishing between static images and dynamic videos, and limit their definition to static non-textual visual resources in text. Postigo and López-Manjón [16] align closely with this definition, considering images in textbooks to be synonymous with illustrations, providing examples such as photographs, drawings, and maps. Most scholars [1, 3, 7, 17] in multimodal analyses focus on static non-textual visual resources, adhering to the second definition.

Abroad, Bezemer and Kress [18] conducted a comparative analysis of English textbooks from the 1930s, 1980s, and 2000s, focusing on images, writing, typography, layout, and the interactions between these elements. They applied various multimodal discourse analysis theories, including visual grammar theory and graphic-textual relations theory. Their findings indicated that English language textbooks have shifted from text-centric layouts to more balanced multimodal

relationships. Tahririan and Sadri [19] used visual grammar theory to analyze images in Iranian high school English textbooks. Their combined quantitative and qualitative analysis revealed that while the images were familiar and relatable to students, there were issues with outdated and stereotypical representations of life. Additionally, the simplistic patterns, dull colors, and lack of graphic detail in the images may undermine students' trust in these visuals. Their study also found that images can serve as meaningful constructs rather than mere decoration. Salnego, Heberle and Balen [20] also applied visual grammar theory to English language textbooks and found that images are integral to the overall meaning and can support language learning as scaffolding tools. Similarly, Elmiana [21] analyzed high school English textbooks in Indonesia using Visual Grammar. Their classified statistics showed that most images reflected students' real-life experiences and local social life. However, issues such as stereotyping and insufficient coverage of modern technology were noted. It is worth mentioning that studies aforementioned focused more on narrative representation and less on construct representation.

In China, research in this field primarily focuses on secondary school English textbooks, with relatively few studies on English textbooks for other educational levels. Chen and Qin [22] analyzed the primary school English textbooks from PEP (People's Education Press) and discovered that multimodal resources—such as dialogue bubbles, word labels, and illustrations—serve as crucial tools for multivoiced interactions. These resources facilitate communication between textbook authors, interlocutors, and readers. Wang and Jiang [3] applied the theory of graphic-textual relations to analyze the multimodal discourse in the story sections of the primary school English textbooks from PEP. Their findings revealed that textbooks across different educational levels exhibit distinct characteristics. The integration of images and text in the story discourse is both reasonable and effective in form and semantics. When fully utilized, these resources can enhance the development of primary school students' multiliteracy skills. Liu and Qu [23] also used Visual Grammar to conduct a comparative analysis of multimodal resources in two sets of university English textbooks. Their qualitative analysis indicated a shift in teaching modes from direct instructional commands to providing teaching resources, reflecting an evolving approach in the textbooks.

In the context of secondary school English textbooks, Qiao [24] analyzed the High School English textbooks using Visual Grammar and graphic-textual relation theory. His quantitative and descriptive analysis revealed that images in the textbooks convey rich meanings and establish an equitable relationship with students. However, he noted that diagrams often accompany text without fully integrating with it, serving primarily as extensions that make the text more intuitive and concrete. Chen and Cong [1] explored the function of image resources in secondary school English textbooks using visual grammar theory. They found that many images lack substantial educational value, are poorly integrated with text, and have insufficiently detailed designs. Luan [25] applied visual grammar theory and intersemiotic complementarity theory to analyze secondary school English textbooks through categorical quantitative statistics. The study found that images effectively recreate the language-use environment, visually represent new vocabulary, and align well with students' cognition. Additionally, images use equal

perspectives and close-ups to foster intimate interactions with students, enhancing their linguistic knowledge. Xiang [26] used visual grammar theory to analyze images in an activity section of a junior high school English textbook, employing categorical and quantitative statistics. The study found that the images are well-matched to students' cognitive levels and support interactive learning. The visual arrangement of images aligns with the overall discourse, enhancing student engagement and understanding.

In conclusion, multimodal analyses of images in English textbooks exhibit the following characteristics. As for theoretical framework, most studies are grounded in the three metafunctions of language from Systemic Functional Grammar, with a focus on Kress and van Leeuwen's Visual Grammar when analyzing images. In term of analysis methods, these studies vary in approach. Some studies classify and count the forms or functions of images in textbooks according to the theoretical framework, while others select representative images for detailed qualitative analysis. These features highlight the diverse methods used to explore how images contribute to the educational value and communication in English textbooks.

Despite advancements in the multimodal analysis of English textbook images, several deficiencies remain. First, although existing studies encompass a broad range of English textbooks, including those for primary and secondary education, there is a notable gap in research on textbooks that adhere to the most recent English Curriculum Standards. Further investigation is needed to analyze the multimodal characteristics of these newer editions. Second, the majority of studies focus on individual versions of English teaching materials, with limited comparative research. Furthermore, current comparative studies predominantly examine different versions of textbooks in use, while longitudinal analyses of temporal changes in textbooks are scarce. There is a need for more comprehensive research that tracks the evolution of multimodal features over time and evaluates their pedagogical significance. These gaps highlight the need for continued and expanded research to fully understand and interpret the evolving multimodal characteristics of English textbooks.

### **3.2. Research on Functions of Images in English Language Textbooks**

Research on the functions of images in English language textbooks has been conducted from two main perspectives: the image-centered perspective and the learner-centered perspective. The image-centered perspective classifies images based on their role within the text. Hunter [27] argues that images in textbooks serve five functions: decorative, reinforcing, interpretive, generalizing, and comparative. The decorative function means that images serve purely decorative purposes and do not interact with the textual information. The reinforcing function means that images partially overlap with and support the textual information. The interpretive function means that images explain or elaborate on the textual information. The generalizing function means that images summarize the textual information. And the comparative function means that images compare different pieces of information. Tang [28] built upon Hunter's categorization framework by integrating the reinforcing, interpretive, and generalizing functions into a single classification, which he referred to as the elaboration and generalization functions. This revised classification

includes reinforcement, generalization, and promotion of text comprehension functions. Chen synthesized previous studies and refined the functions of images into three categories: the decorative function, which is unrelated to the content of the text; the representation function, which explains the content of the text; and the organizational function, which describes the internal relationships within the text.

Studies from a learner-centered perspective classify image functions based on their role for students. Duchastel [29] identified the following image functions: attentional function, which attracts students' attention; orienting function, which focuses students' attention; affective function, which increases students' interest in learning; and cognitive function, which enhances students' learning. Levie and Lentz [30] categorized attentional and orienting functions as attention function. They grouped the functions influencing students' emotional attitudes and enjoyment under affective functions, while functions aiding text comprehension, memorization, and providing additional information were classified as cognitive functions. They also proposed a compensatory function to assist learners with reading difficulties. Song [4] classified image functions into decorative, explanatory, and facilitating functions. The decorative function refers to images that are independent of the text, intended to enhance the visual appeal of textbooks and stimulate interest. The explanatory function describes images that support and depend on the text by visualizing content and presenting complex concepts, principles, and processes in an intuitive way to aid comprehension. The facilitating function includes the sub-functions of expanding content, broadening horizons, and improving readers' abilities.

Existing studies on image functions have some limitations. Some frameworks for categorizing functions are overly detailed, with categories that overlap and are difficult to distinguish [1]. Moreover, these frameworks are primarily theoretical, making them less applicable for textbook writers and frontline teachers in practical textbook writing and teaching. Based on previous research, Chen and Cong [1] proposed classifying the functions of images in English textbooks into three categories: decorating the layout, promoting comprehension, and providing learning content. Decorating the layout refers to images that are not closely integrated with the text. These images do not constitute the main learning content but serve to enhance the aesthetics of the textbook layout or provide a background for learning activities. Although these decorative images play a less direct role in learning, they contribute to the compositional meaning of visual grammar, which is part of the overall textbook layout. Promoting comprehension involves images that support the text by providing essential background information or context, or by hinting at the meaning of the text. These images have a more direct role in learning, helping students understand the content of the text. For example, an image used to illustrate a word helps students grasp its meaning. Such images are substitutable, meaning their information could be conveyed through text. However, omitting comprehension-enhancing images would require significant changes to the textbook, as their absence would necessitate presenting the information in another form. Additionally, these images reflect the interactive meaning of visual grammar, indicating the relationship between images, texts, and readers, and guiding readers on how to engage with the material. Providing learning content refers to images that constitute the main learning content and directly impact the learning process and

outcomes. These images are closely related to the text or learning activities, forming a complete learning experience. For instance, English teaching activities that require students to listen to a recording while observing a picture or completing tasks based on an image depend on these images. Without them, the activities cannot be conducted. Such images are integral to the learning process, helping students integrate visual information to construct meaningful discourse and develop higher-order cognitive skills. They embody all three aspects of Visual Grammar: representational meaning, interactive meaning, and compositional meaning.

Therefore, this study selects the old and new English textbooks from PEP 7A as examples, which are published before and after the issuance of the new English Curriculum Standards, and uses visual grammar theory to take images as the object of multimodal analysis. Based on the framework of image functions proposed by Chen and Cong [1], the study adopts both categorical quantitative statistics and qualitative descriptions to analyze the functions of the images and the relationship between images and learners, in order to decipher the direction of development and changes in the multimodal features of the English textbooks. The main research questions explored in this study are:

RQ1: What are the changes in the functions of images in the new textbook compared to the old one?

RQ2: How does the new textbook reflect the requirements of the new English Curriculum Standards compared to the old one?

## 4. Methodology

This study adopts the textual analysis method and the comparative research method to analyze in depth the changes in the functions of image resources in the new and old textbooks. Among these methods, textual analysis is the primary research approach used in this study. The images in the textbooks are selected for textual analysis to reveal changes in the functions of images in the old and new textbooks through quantitative statistical analysis and qualitative descriptions. To obtain quantitative data, this study labels the images in the textbooks according to the classification proposed by Chen and Cong [1] and counts the number of images with different functions. This approach helps decipher the overall trend of multimodal feature changes by comparing the number of images with different functions in the old and new textbooks. To obtain qualitative data, this study selects multimodal discourses with the same or similar themes in the old and new textbooks, and conducts qualitative descriptive comparative analyses of the images within them. This provides evidence for the quantitative data and supplements details that the quantitative analysis may overlook.

### 4.1. Data of the Present Study

In this study, both the old and new versions of the English textbook from PEP 7A were selected as research objects, and the images in these textbooks were used as the corpus. As a newly published English textbook, the new English textbook from PEP 7A is an excellent choice for studying changes in multimodal features after the release of the new English Curriculum Standards. Moreover, secondary school English textbooks have only been updated with one copy of the first book of the seventh grade, making it the only choice for comparative analysis. At the same time, this version of the textbook is well-known to the researcher, which is conducive

to the smooth progress of the analysis.

## 4.2. Data Analysis

### 4.2.1. Quantitative Analysis

To meet the needs of teaching at the beginning stage and during the transition between primary and secondary schools, the English textbook for the first year of Grade 7 includes transitional units. The old textbook has three transitional units and nine formal units, while the new textbook reduces the total number of units, featuring three transitional units and seven formal units to reflect the idea of reducing the teaching burden. Simultaneously, each unit's volume increased from 6

pages to 8 pages, expanding unit capacity, enriching the teaching content, and accommodating large-unit teaching needs. This study quantified the images in both textbooks: the old textbook has a total of 181 images, and the new one has 159 images due to the reduced number of units. The researcher then used Chen & Cong's [1] classification framework for image functions in English textbooks to label and classify images, counted the images separately for each textbook, and calculated the percentage of each image function. Finally, the data of the old and new textbooks were compared to form the following table:

**Table 1.** The functions of images and their percentages

| Image functions    | Layout-decorative function | Comprehension-promoting function | Learning-content-providing function |
|--------------------|----------------------------|----------------------------------|-------------------------------------|
| <b>The old</b>     | 133                        | 31                               | 17                                  |
| <b>percentages</b> | 73.5%                      | 17.1%                            | 9.4%                                |
| <b>The new</b>     | 96                         | 44                               | 19                                  |
| <b>percentages</b> | 60.4%                      | 27.7%                            | 11.9%                               |

As can be seen from Table 1, in the old textbook, there are 133 images performing the function of decorating the layout, accounting for 73.5%; 31 images performing the function of promoting comprehension, accounting for 17.1%; and 17 images performing the function of providing learning content, accounting for 9.4%. The old textbook has more decorative images, which account for 73.5% of the total number of images. More than half of the images in the textbook have little to do with the content of learning, and these images have no direct relevance to the content of learning other than to individual words in the text or background information. Only 9.4% of the images in the old textbook are used as learning content, and there are fewer types of images. Many of these images appear in graphic matching learning activities, such as sorting pictures according to text content or listening to recordings, matching words with images, and making correct or incorrect judgments or choices based on text information.

In contrast, in the new textbook, there are 96 images performing the function of decorating the layout, accounting for 60.4%; 44 images performing the function of promoting comprehension, accounting for 27.7%; and 19 images performing the function of providing learning content, accounting for 11.9%. Due to the difference in the number of units between the old and new textbooks, it is not reasonable to compare the numbers directly, so it is supposed to compare the percentages. The comparison shows that the percentage of images performing the function of decorating the layout has decreased, while the percentages of images performing the functions of promoting comprehension and providing learning content have both risen. This indicates an increase in the number of pictures with substantial learning value, which means that images in English textbooks are no longer merely on the margins of the textbook, assuming the role of decorating and beautifying the layout, but have begun to become a part of the multimodal discourse, working in tandem with the text to express meaning, which needs to be observed and interpreted by the learner.

From the perspective of the new English Curriculum Standards, this overall trend reflects the adjustments and optimization of the design of teaching materials and teaching objectives in line with the new English Curriculum Standards.

It focuses on cultivating core literacy, promoting scientific learning methods, and enhancing education quality, and these concepts directly impact the use of images in the new textbooks. First, the new English Curriculum Standards emphasize the development of students' core literacy, including subject-specific literacy, key competencies, and character development. To achieve this, images in the textbooks are no longer merely decorative; instead, they serve to help students understand content and develop their abilities. For example, the new textbooks have increased the use of mind maps and flow charts, which can enhance students' thinking and inquiry skills, thereby supporting deep learning. At the same time, the reduced use of decorative images helps students focus on meaningful image content and avoid distractions from irrelevant decoration. This focus training enables students to read and process image information more effectively. Secondly, the new English Curriculum Standards advocate for multiple literacies, including the ability to interpret information in various forms such as text, images, diagrams, and symbols. This requires students to learn how to integrate multiple forms of information, combining images, diagrams, and textual content to form a comprehensive understanding of the subject matter. The increase in images that promote comprehension and provide learning content in the new textbooks aligns with its requirement for students to enhance their multiple literacy skills. The new English Curriculum Standards encourage students to make connections across different modalities, such as reading a diagram to explain the content of a text or summarizing a text through a schematic diagram. Such images help students connect modalities, thereby improving their overall literacy skills.

### 4.2.2. Qualitative Analysis

To obtain qualitative data, this study selects multimodal discourses with the same or similar themes in the old and new textbooks, and conducts qualitative descriptive comparative analyses of the images within them. This provides evidence for the quantitative data and supplements details that the quantitative analysis may overlook. The selected comparative discourses are listed in the table below:

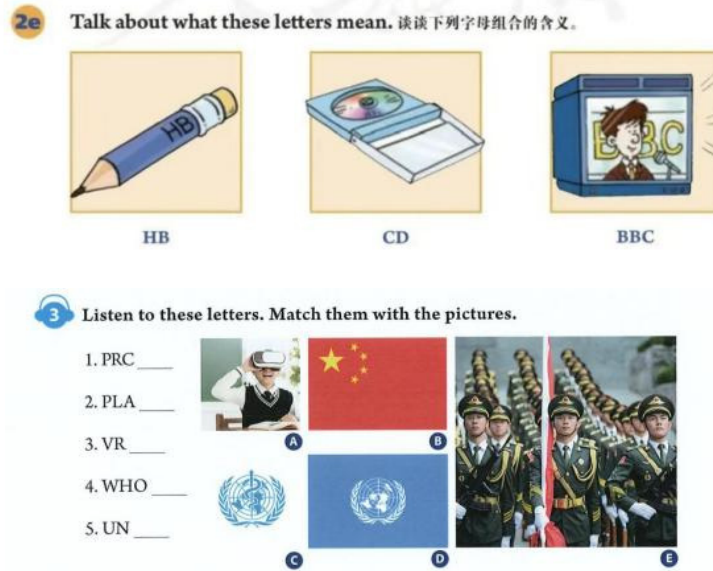
**Table 2.** Discourses chosen for comparative analysis

| The old textbook                                    | The new textbook   |
|---|--|
| STARTER UNIT 1 Good morning! -- 2e(p. S2)           | STARTER UNIT 1 Hello! -- Section A-- Pronunciation 3 (p. 3)      |
| STARTER UNIT 1 Good morning! -- 3c(p. S3)           | STARTER UNIT 1 Hello! -- Section B -- 1b-- Conversation 2 (p. 4) |
| STARTER UNIT 2 What's this in English -- 1a (p. S5) | STARTER UNIT 3 Welcome! -- Section A -- 1a (p. 13)               |
| UNIT 1 My name's Gina -- Section A -- 1a (p. 1)     | UNIT 1 You and Me -- Introduction (p. 19)                        |
| UNIT 2 This is my sister -- Section B -- 2b (p. 11) | UNIT 2 We're Family! -- section B -- 1b (p. 32)                  |

In the following section, the above discourses in pairs will be comparatively analyzed through qualitative descriptive analysis using visual grammar to elucidate in detail how images change functionally.

Taking STARTER UNIT 1 Good morning! -- 2e from the

old textbook (p. S2) and STARTER UNIT 1 Hello! -- Section A -- Pronunciation 3 from the new textbook (p. 3) (Figure 1) as examples, both texts are part of transition units designed to review the English alphabet through acronyms.



**Figure 1.** Images from the old textbook and the new one

In the old textbook, the acronyms used are HB, CD, and BBC, while the new textbook presents acronyms such as PRC, PLA, VR, WHO, and UN, accompanied by corresponding images. In terms of image forms, the three images in the old textbook are all drawings or cartoon pictures, whereas the five images in the new textbook are photographs, which enhances authenticity. The shift from cartoon drawings to photographs is more pronounced in similar discourses between the old and new textbooks, especially in introduction picture of the units. For example, the old textbook UNIT 1 My name's Gina -- Section A -- 1a (p. 1) uses cartoon drawings and dialogue bubbles to depict self-introductions and greetings, while the new textbook UNIT 1 You and Me -- Introduction (p. 19) presents a horizontal view of three students introducing themselves to each other. This change in perspective will be analyzed in more detail below. Comparing the functions of the images, in the old textbook, the activity required is 'Talk about what these letters mean,' where the images serve to facilitate comprehension by suggesting the meanings of abbreviations and reproducing the representational function of the visual grammar. However, the images themselves are not the primary content of learning but play a supportive role. For example, students can understand the meaning of CD by linking the image with the text. However, the substitutability of such images means the activity could still be conducted even without them. In contrast, in the new textbook, the activity requires 'Listen to these letters. Match them with the pictures,' which means the images function as content providers. The images are closely related to the learning activity, forming an integral part of the learning content.

Students need to analyze the images and match them with the appropriate abbreviations, requiring deeper engagement.

Notably, besides the shift towards higher-order functions, the images have also become more contemporary. In the old textbook, examples like HB and CD are used, but the new textbook incorporates images like VR, reflecting the latest technological developments and broadening students' knowledge. Additionally, the inclusion of Chinese elements such as PRC and PLA strengthens national awareness and pride. Through repeated exposure to national and military symbols, students' identification with language and cultural symbols is enhanced, ultimately boosting their national pride and sense of belonging.

The shift to higher-order functions of images can also be seen in the comparison of the old textbook UNIT 2 This is my sister -- section B -- 2b (p. 11) and the new textbook UNIT 1 We're Family! -- section B -- 1b (p.32) (Figure 2).

Two family photographs from the old textbook serve primarily as decorative images, enhancing the aesthetic and visual appeal of the material. The illustrations are arranged in a left-right structure, with the family photo on the left and the corresponding text on the right. According to the information value emphasized by the compositional function of visual grammar, the left-right structure typically represents a 'known-new information' relationship, where elements on the left are seen as known or background information, and those on the right are regarded as new or key content. This arrangement emphasizes the decorative function of the family photos, positioning them as supplementary rather than essential. The associated learning activity, 'Read about

Jenny's family and circle the names,' is also not directly related to the pictures; students can complete the task with no reference of the images. This suggests that, although the images relate to the text, they do not significantly contribute to the content or learning activity and primarily serve to decorate the page. In contrast, the images in the new textbook, while similar in content to those in the old textbook, have undergone a substantial functional shift. The activity, 'Read the text and label the people in the photo,' requires students to integrate visual and textual information, constructing a

comprehensive understanding to match names with the correct individuals in the photo. Here, the image is integral to the learning activity, providing key content; without the images, the task would not be possible. Additionally, the top-down structure of the images in the new textbook allows key content to be placed in a visually prominent position, guiding students to focus on the parts of the image necessary for completing the task. This shift demonstrates a change in the function of images from merely decorating the layout to actively contributing to the learning content.



Figure 2. Images from the old textbook and the new one

Both of the above examples illustrate the shift of images from a lower-order decorative function to a higher-order function of providing learning content, aligning with the new English Curriculum Standards for multiliteracies. However, through qualitative descriptive analysis, it was found that although the functions of some images had not changed, the graphic-textual relationship was optimized in two main aspects. Firstly, English textbooks often present dialogues and short texts accompanied by images that reflect their contexts. However, some images do not precisely match the text's context or lack contextual authenticity. For example, in the old textbook STARTER UNIT 1 Good morning! -- 2e (p. S3) and the new textbook STARTER UNIT 1 Hello! -- Section B -- 1b -- Conversation 2 (p. 4) (Figure 3), the relationship between images and text has been refined to improve the contextual relevance and authenticity of the visual elements, enhancing their alignment with the dialogue or narrative presented in the text.

conveys the library setting, it deviates from actual behavioral norms. In real libraries, people are expected to maintain silence, and conversations, especially between students, should be kept brief and restrained. And greetings should be quiet and minimal. The depicted dialogue between two students in the library does not adhere to these norms, undermining the authenticity of the situation and the educational intent of promoting appropriate behavior. In contrast, the new textbook presents the same dialogue in a more contextually appropriate setting. The conversation takes place at a stairway between classes, where two students meet, greet each other, and, prompted by the ringing bell, politely say goodbye. This design reflects daily life more accurately and aligns with public expectations of appropriate behavior. It carefully considers students' behavioral education by presenting realistic scenarios, subtly teaching students how to interact appropriately in different contexts. The open environment of the stairway allows for moderate conversation, while the bell signals the need to end the exchange promptly, mirroring real campus life and guiding students toward proper behavior.

In the old textbook, the image depicts two students having a dialogue in a library, as indicated by the three rows of bookshelves in the background. While this scene visually



Figure 3. Images from the old textbook and the new one

The library scenario in the old textbook, although educationally intended, misrepresents real-life norms and may inadvertently mislead students. In contrast, the scene transformation in the new textbook enhances the fluency and coherence of the story, aligns with real-life behavior, and

makes the dialogue feel more natural within its specific context. This adjustment not only increases the authenticity and emotional resonance of the images but also more effectively conveys correct behavioral norms, guiding students to interact appropriately in suitable contexts. These

contextual changes reflect the humanization of the new teaching materials and their precise alignment with educational objectives, enhancing the overall learning experience and promoting correct social behavior among students.

Misalignment between the content of pictures and text is a typical manifestation of inappropriate graphic configurations, particularly when images reflect only part of the information conveyed in the text. This misalignment can affect learners' overall understanding of the textbook content. For instance, considering the old textbook STARTER UNIT 2 What's this in English -- 1a (p. S5) and the new textbook STARTER UNIT 3 Welcome! -- Section A -- 1a (p. 13) (Figure 4), both learning activities involve practicing the sentence patterns "what's this/that?", but the graphic configuration issues are

distinctly different between the two examples. Textbook dialogues often contain multiple turns of speech, but a static image can usually represent only one turn, leading to mismatches between the image and the dialogue. In the old textbook, the dialogue features two interlocutors and two turns of speech. The image shows both speakers facing the reader, reflecting only the first turn (A: "What's this in English?" B: "It's an orange.") with the boy pointing at an orange on the table. However, for the second turn (A: "What's that in English?" B: "It's a jacket."), the positioning is inaccurate—the speakers should face the jacket on the bed rather than having their backs to it. This mismatch can easily cause learners to misinterpret the dialogue and form incorrect associations with the language context.



Figure 4. Images from the old textbook and the new one

In contrast, the new textbook demonstrates a more reasonable alignment of graphic and textual information. The image includes three interlocutors and two turns of speech, making each turn of speech clear. The third person, positioned behind the other two, can observe their gestures and responds accurately, ensuring that each turn of speech is fully represented. This improvement minimizes the risk of images reflecting only part of the textual information, enhancing the balance between graphics and text and improving learners' comprehension of the scenario and authentic language use.

Furthermore, the old textbook's lack of equivalence between images and text not only disrupts understanding but also weakens the role of images as a language supporting tool. Learners may only associate the first turn of speech with the image, lacking visual cues for subsequent dialogue, which affects the fluency and accuracy of language practice. The limitation of static images is their inability to dynamically present multiple interactions, and without appropriate visual accompaniment, important information can be lost or misrepresented. The new textbook addresses these issues by optimizing graphic configurations, ensuring that each image corresponds closely with the relevant speech turn. This enhancement not only improves the teaching effect by aligning visual and textual elements but also enriches the learning experience, reinforcing the textbook's effectiveness as an educational resource.

Last, the comparison between the old and new textbooks

also reveals an increased use of the horizontal perspective, a key dimension of the visual grammar's interactive function. This perspective significantly impacts the interaction between the reader and the image and the way information is conveyed. Unlike the top-down perspective, the horizontal perspective mirrors the everyday human viewpoint, placing the reader on the same level as the depicted scene. This alignment reduces the perceived distance between the reader and the image, fostering a sense of immersion and establishing a more equal and interactive relationship. As a result, readers find it easier to engage with the scene, enhancing their sense of intimacy and empathy. In contrast, the top-down perspective, often associated with a sense of authority or distance, can create a feeling of detachment or subordination. This may conflict with the psychological needs of secondary school students, who generally respond better to visuals that feel relatable and inclusive. The horizontal perspective, free from a sense of hierarchy or dominance, aligns more closely with the perceptual comfort and cognitive needs of this age group. For instance, in the old textbook UNIT 1 My Name's Gina -- Section A -- 1a (p. 1) (Figure 5), images of classmates introducing themselves are presented from a top-down angle, creating a sense of distance and alienation that can diminish students' engagement. This perspective suggests a condescending view, subtly positioning the observer above the participants in the scene, thus weakening the learner's connection and sense of participation.

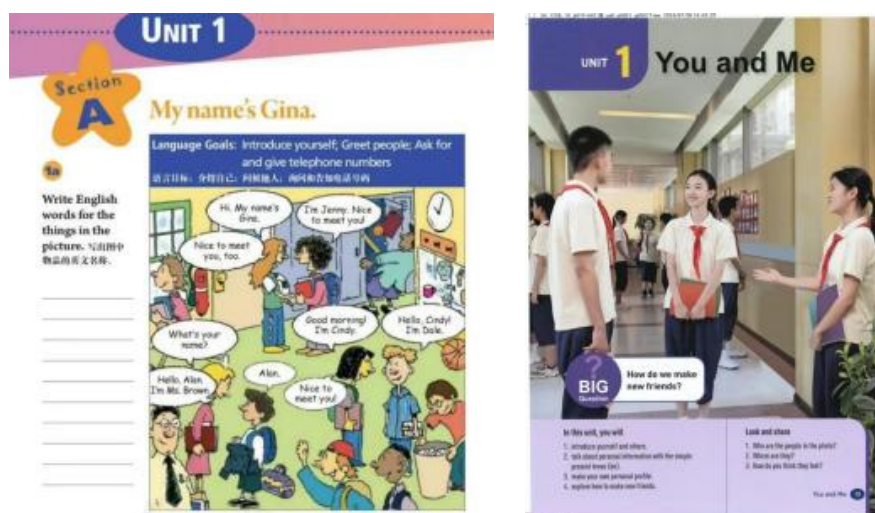


Figure 5. Images from the old textbook and the new one

Conversely, the new textbook UNIT 1 You and Me -- (p. 19) (Figure 5) uses a horizontal perspective, depicting three students introducing themselves on equal footing. This change in viewpoint enhances the reader's immersion, making the scene feel more realistic and relatable. The horizontal perspective reproduces the learning environment more authentically, aligning closely with students' everyday experiences. As a result, the content appears more relevant and accessible, helping students connect the material to their own lives, which enhances comprehension and retention. This shift not only improves the learning experience by fostering a sense of closeness and engagement but also supports a more realistic representation of social interactions, reinforcing the textbook's educational intent. By aligning visual perspectives with the cognitive and emotional needs of learners, the new textbook effectively enhances the authenticity and relatability of the learning content, making it more memorable and impactful for students.

## 5. Results and Discussions

### 5.1. Image Functions

Quantitative data show that compared to the old textbook, more images in the new textbook performing high-order functions, such as promoting comprehension, and providing learning content, which is also manifested by the qualitative comparative analyses. This indicates an increase in the number of pictures with substantial learning value, which means that images in English textbooks are no longer merely on the margins of the textbook, assuming the role of decorating and beautifying the layout, but have begun to become a part of the multimodal discourse, working in tandem with the text to express meaning. In other words, This reflects the fact that the images in the new textbook are more concerned with presenting holistic concepts in a multimodal discourse, rather than capturing an individual process in a multimodal discourse.

### 5.2. Graphic-textual Relations

The comparative analyses of graphic-textual configurations reveal two main issues in the old textbook: first, the pictures often fail to match the context of the text or lack situational authenticity; second, there is a significant asymmetry between the graphic and textual information. Specifically, this study found that the images in the old textbook primarily support the text, resulting in a graphic-text

relationship where the text is dominant, and the images merely serve as supplementary visual aids. To addressing these shortcomings, the new textbook has improved the integration of images and text, creating more balanced and rational configurations. The images and texts now exhibit a closer relationship, reflecting a deliberate effort to align the visual and verbal elements more effectively. This shift marks a significant improvement, as images are no longer secondary elements that only serve to enhance the text but are becoming essential components that contribute to the overall meaning.

In conclusion, the images in the new textbook are beginning to show new features. In terms of composition, images in the new textbook are more frequently positioned in the left and upper parts of the discourse layout, closer to the focal point of the page. This strategic placement draws more attention to the images, making them integral to the reader's engagement with the material. The positioning closer to the center of the discourse helps establish a more balanced and interactive relationship between text and images. As for the status between images and text, although text still holds a relatively dominant position, images have moved beyond the mere role of marginal supplements to the text. Instead, they have started to complement the text by conveying core information. This shift reflects a more sophisticated approach where images and texts work in tandem to provide a comprehensive understanding of the content. The images are no longer confined to reinforcing the text but are critical to the reader's comprehension of the topic. Last, in the new textbooks, the logical-semantic relationship between images and text has evolved. The proportion of images that serve as illustrative examples has increased, also indicating a shift towards images that express the core information of the discourse rather than merely setting the scene. This change enhances the instructional value of the images, allowing them to carry significant content and contribute directly to the learning objectives. These developments demonstrate that the new textbook is moving towards a more integrated and meaningful use of images, aligning with modern pedagogical standards that emphasize multimodal learning. By rationalizing the relationship between images and text, the new textbooks effectively enhance students' engagement, comprehension, and retention of the material, thereby creating a more dynamic and interactive learning experience.

### 5.3. Multimodal Literacy

The new English Curriculum Standards emphasize

multimodal literacy, and one of the features that distinguish this competence from traditional literacy is its orientation towards resources beyond textual ones, interpreting symbolic systems such as images, typography, and sound as well. This interpretation is not done separately but focuses on the overall meaning that the different semiotic systems construct together [2]. In the process of meaning construction, each symbolic system does not work independently; they depend on each other, and sometimes some symbolic systems cannot even express meaning independently without the help of others [2]. Therefore, the images and words in the new textbooks tend to be more holistic in their composition, guiding students to consciously pay attention to the relationship between different symbol systems, integrate the meanings expressed by these systems, and seek the overall meaning of the discourse after recognizing the regularity of each symbol system, thereby cultivating their multimodal literacy skills.

Furthermore, the new English Curriculum Standards insist on integrating Chinese elements into English language teaching and require the cultivation of comprehensive talents with international vision and cultural confidence. Comparative analyses also show that the new textbook responds to the call to incorporate more Chinese elements, which not only enriches the learning content but also realizes the organic combination of language learning and cultural inheritance. It provides students with a window to display and learn Chinese culture and cultivates cross-cultural communicative competence.

## **6. Conclusion**

### **6.1. Pedagogical Implications of the Present Study**

English teachers, as leaders of the English language curriculum, bear the responsibility of organizing and structuring the content of the curriculum and helping students improve their core literacy in the English language. Images are an important part of the multimodal features of textbooks. Therefore, teachers should recognize that images are a key resource for developing students' multimodal literacy skills and help students improve their ability to interpret and gain more discourse knowledge in line with the new English Curriculum Standards. Based on the findings, this study puts forward three suggestions to help frontline English teachers better navigate the new multimodal features in the newly implemented textbook.

From the results and analyses of this study, it is clear that the functions of image resources in the new textbooks are undergoing fundamental changes. Generally, images in English textbooks are no longer merely decorative; they are beginning to become part of the multimodal discourse, expressing meanings in collaboration with the text, which must be observed and interpreted by learners. To better adapt to the change in the status of multimodal resources, English teachers should first actively change their teaching concepts and start paying attention to multimodal resources such as images. Specifically, teachers should not focus solely on the textual content of the textbook or extract only linguistic information for teaching. Instead, they should recognize that multimodal resources are also part of the teaching content, observe them in their lesson planning, and incorporate them according to their characteristics in the teaching process. In the long run, teachers will be able to present multimodal resources more effectively, and students will recognize the

importance of multimodal resources, developing the habit of observing and understanding them, such as images and typography, during reading.

Additionally, the use of multimodal resources, such as images, in the classroom holds significant potential. When used properly, they can not only promote the development of students' multimodal literacy skills but also help implement the core literacy of the English subject. Therefore, English teachers should think strategically about the use of images, allowing them to contribute to developing moral values rather than serving as mere formalities. Specifically, teachers should focus on three aspects when using images to implement the core qualities of the English subject. First, regarding language proficiency, core literacy emphasizes the need for students to develop the five skills of listening, speaking, reading, viewing, and writing, and to acquire the ability to apply these skills to understand and express meaning in real life. Therefore, English teachers should actively guide students to observe and interpret images and develop their visual literacy skills during English lessons. Teachers should also pay attention to the relationship between images and texts, using images to assist students in text reading and develop their language skills holistically. Secondly, regarding thinking skills, core literacy emphasizes students' logical, critical, and creative thinking qualities. Thus, when English teachers use images in classroom activities, they should pay special attention to the conceptual nature of images and guide students to engage in generalizing, synthesizing, evaluating, and other thinking activities by asking questions about images. Thirdly, in terms of learning ability, core literacy expects students to exercise their initiative and consciously self-regulate their English learning. Therefore, in the process of using images, English teachers should explore the connection between images and students' real-life experiences, focus on features of images that can capture students' attention, and use the natural affinity between images and learners to encourage independent learning.

### **6.2. Limitations and Future Suggestions**

This study certainly has some limitations. Firstly, this study focuses solely on the English textbook from PEP 7A. However, other textbooks, including the junior middle school English textbooks from the Foreign Language Teaching and Research Press, have also been updated during this period. Consequently, the perspective of this study is somewhat limited, as it only addresses the changes in the textbook from PEP 7A and interprets these changes. Secondly, due to time and resource constraints, this study concentrated on images as a significant multimodal resource when examining the multimodal features of the new textbooks. However, according to existing literature on the multimodal aspects of textbooks, other features, such as fonts and typography in the discourse of English textbooks, are equally meaningful, and differences in these aspects can also reflect changes in the teaching concepts of the new textbooks.

Therefore, during this period of textbook renewal, when the features of new junior middle school English textbooks and the differences between old and new versions are receiving attention from researchers, this study offers the following two suggestions for future multimodal research on English textbooks based on its findings. First, expand the scope of research objects. This study chose the old and new textbooks from PEP 7A as its focus, recognizing that this version is widely used. However, future research should broaden its

perspective to include other versions of junior middle school English textbooks, such as those from the Foreign Language Teaching and Research Press, to uncover changes that have uniformly occurred in the multimodal characteristics of the new textbooks. Researchers could also compare and categorize these multimodal features by examining different versions of textbooks. Secondly, focus on a broader range of multimodal features. While this study concentrated on comparing images within the multimodal discourse of the old and new textbooks, existing literature suggests that other elements, such as fonts, typography, and layout, are also multimodal features worthy of analysis. Future research could combine these multiple multimodal features to better understand the teaching philosophies embedded in the textbooks.

## References

- [1] X.-T. Cheng and L. Cong, "Design and use of visual resources in English textbook writing," *Curriculum, Teaching Material and Method*, no. 8, pp. 78–85, 2020.
- [2] Z.-L. Hu, "Multimodality in social semiotic studies," *Language Teaching and Linguistic Studies*, no. 1, pp. 1–10, 2007.
- [3] R.-H. Wang and G.-Y. Jiang, "Multimodal foreign language teaching: Integrating visual and textual resources—A case study of the story section in the PEP primary English textbook," *Basic Education*, no. 3, pp. 84–90, 2015.
- [4] Z.-S. Song, "A cognitive psychological study of textbook illustrations," *Journal of Beijing Normal University (Social Sciences)*, no. 6, pp. 24–28, 2005.
- [5] S.-H. Gao, "A preliminary study on textbook illustrations," *Educational Research and Experiment*, no. 3, pp. 36–40, 72, 2000.
- [6] Ministry of Education of the People's Republic of China, *English Curriculum Standards for Compulsory Education*, Beijing: Beijing Normal University Press, 2022.
- [7] G. Kress and T. Van Leeuwen, *Reading Images: The Grammar of Visual Design*, 3rd ed., London: Routledge, 2020.
- [8] D.-L. Zhang, "Exploring an integrated framework of multimodal discourse analysis," *Chinese Foreign Languages*, no. 1, pp. 24–30, 2009.
- [9] Z.-Z. Li, "A social semiotic analysis of multimodal discourse," *Foreign Language Research*, no. 5, pp. 1–8, 80, 2003.
- [10] Z.-C. Yang and X.-W. Miao, "Review of research on the layout structure of multimodal discourse," *Journal of Foreign Languages*, no. 6, pp. 24–29, 2017.
- [11] G.-H. Xu and H. Qu, "Multimodal discourse analysis—A new perspective in the information age," *Shandong Foreign Language Teaching*, no. 2, pp. 3–7, 2009.
- [12] Z.-L. Hu, Y.-S. Zhu, D.-L. Zhang, and Z.-Z. Li, *An Introduction to Systemic Functional Linguistics*, Beijing: Peking University Press, 2005.
- [13] J.-Y. Zhang and P.-P. Jia, "Some reflections on visual grammar," *Contemporary Foreign Languages Studies*, no. 3, pp. 38–42, 160, 2012.
- [14] M. A. K. Halliday, *Language as Social Semiotic: The Social Interpretation of Language and Meaning*, London: Edward Arnold, 1978.
- [15] L.-N. Yang and Z.-M. Wu, "An analysis of the use of images in listening tasks of junior high school English textbooks," *Foreign Language Teaching in Schools (Middle School Edition)*, no. 12, pp. 13–17, 2018.
- [16] Y. Postigo and A. López-Manjón, "Images in biology: Are instructional standards used in textbook image design?," *International Journal of Science Education*, vol. 41, no. 2, pp. 210–229, 2019.
- [17] R. Martinec and A. Salway, "A system for image–text relations in new (and old) media," *Visual Communication*, vol. 4, no. 3, pp. 337–371, 2005.
- [18] J. Bezemer and G. Kress, "Visualizing English: A social semiotic history of a school subject," *Visual Communication*, vol. 8, no. 3, pp. 247–262, 2009.
- [19] M. H. Tahririan and E. Sadri, "Analysis of images in Iranian high school EFL course books," *Iranian Journal of Applied Linguistics*, vol. 16, no. 2, pp. 137–160, 2013.
- [20] N. Salbego, V. M. Heberle, and M. G. S. da Silva Balen, "A visual analysis of English textbooks: Multimodal scaffolded learning," *Calidoscópio*, vol. 13, no. 1, pp. 5–13, 2015.
- [21] D. S. Elmiana, "Pedagogical representation of visual images in EFL textbooks: A multimodal perspective," *Pedagogy, Culture & Society*, vol. 27, no. 4, pp. 613–628, 2019.
- [22] Y.-M. Chen and X.-Y. Qin, "Interpersonal meaning and polyphony of multimodal signs in textbook discourse," *Foreign Languages and Teaching*, no. 12, pp. 15–18, 2007.
- [23] X. Liu and D. Qu, "Exploring the multimodality of EFL textbooks for Chinese college students: A comparative study," *RELC Journal*, vol. 45, no. 2, pp. 135–150, 2014.
- [24] W.-J. Qiao, "A multimodal discourse analysis of the PEP senior high school English textbook," M.A. thesis, Shaanxi Normal University, 2016.
- [25] X.-X. Luan, "A study of visual images in middle school English textbooks based on multimodal discourse analysis theory," M.A. thesis, Central China Normal University, 2020.
- [26] R. Xiang, "A multimodal study of image texts in junior high school English textbooks based on visual grammar theory," M.A. thesis, Southwest University, 2021.
- [27] B. Hunter, A. Crismore, and P. D. Pearson, "Visual displays in basal readers and social studies textbooks," in *The Psychology of Illustration: Volume 2: Instructional Issues*, J. W. Brown, Ed. New York, NY: Springer US, 1987, pp. 116–135.
- [28] G. M. Tang, "Textbook illustrations: A cross-cultural study and its implications for teachers of language minority students," *The Journal of Educational Issues of Language Minority Students\**, vol. 13, no. 2, pp. 175–194, 1994.
- [29] P. Duchastel, "Textbook illustration: research and instructional design," in *Educational Media Yearbook, 1980a*, J. W. Brown, Ed. New York: R. R. Bowker, 1980, pp. 58–63.
- [30] W. H. Levie and R. Lentz, "Effects of text illustrations: A review of research," *Educational Communication and Technology Journal*, vol. 30, no. 4, pp. 195–232, 1982.