Scilight

Studies of Applied Linguistics in Asia
https://www.sciltp.com/journals/sala

Article

More than Meets the Eye: An Image-Text Meaning Construction of Contemporary Senior High School English Textbooks in China

Yajing Zhou [1] and Yanyi Liao [1,2,*]

[1] School of Foreign Languages, Guangzhou University of Software, Guangzhou 510990, China
[2] Faculty of Humanities, The Hong Kong Polytechnic University, Hong Kong, China
* Correspondence: yanyi-phoebe.liao@connect.polyu.hk; Tel.: +86-13926130361

How To Cite: Zhou, Y., & Liao, Y. (2026). More than Meets the Eye: An Image-Text Meaning Construction of Contemporary Senior High School English Textbooks in China. *Studies of Applied Linguistics in Asia*, *1*(1), 65–77. https://doi.org/10.53941/sala.2026.100006

Received: 29 May 2025
Revised: 8 December 2025
Accepted: 10 December 2025
Published: 24 March 2026

Abstract: In a globalized era of information technology and multimedia, the content compilation of high school English textbooks has transformed from a single textual modality into a multimodal one. This study aims to conduct a systematic visual grammar analysis of the current People's Education Press (2024) of senior high school English textbooks in China and explore how multimodal resources, such as images and texts, integrate and present knowledge in language textbooks. The study analyzed three compulsory English textbooks, employing Visual Grammar as its analytical framework. Both quantitative and qualitative analyses are used to examine images in the reading sections of senior high school English textbooks. The findings of this study are: (1) the visual design in these textbooks not only effectively complement the language content of reading section through the representative meanings of images but also strengthen the coherence of the readings' plots through narrative structures of the images; (2) the reading section's images mainly aims at enhancing the viewers' sense of participation; (3) a text-centered layout with images playing an assisted role. This study seeks to offer a detailed analysis and reference for enhancing the visual design and multimodal integration of senior high school English textbooks, and to further contribute to the application of visual grammar theory in research on the related field.

Keywords: high school English textbooks; multimodal discourse; visual grammar

1. Introduction

With the rapid development of multimedia technology, the division of labor in society has become increasingly refined, and the forms of social interaction have become increasingly diverse. In this context, the boundaries between various symbols are gradually blurred. Education has entered the postmodern period, and discourse is no longer confined to the traditional textual mode but integrates a variety of modalities, such as drawings, animations, sounds, colors, choreographies, charts, and so on. This multimodality, and the joint participation of multiple modalities in social practice, have prompted us to conduct in-depth research on multimodal discourse. In 1996, Kress and van Leeuwen introduced the theory of Visual Grammar. In 2001, they developed a framework for multimodal discourse that provides a theoretical foundation for related research.

As essential resources for students' learning, English textbooks are key vehicles for achieving teaching goals and enhancing educational effectiveness. They are also essential materials for Chinese schoolchildren learning English. With the continuous enrichment of textbook content, multimedia elements such as visual and auditory resources have been widely introduced. When using these textbooks, students are bound to utilize a variety of senses, such as auditory, visual, and tactile senses, to communicate and learn, which is an essential manifestation of multimodal discourse. What's more, the Compulsory Education's English Curriculum Standards (2019 Edition, 2022 Revised Edition) of China added "viewing" as a

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

language skill (Ministry of Education of the People's Republic of China, 2022) while reading multimodal texts. Therefore, the multimodal design of textbooks has become a significant trend in the development of educational resources today.

In recent years, advances in multimedia technology have driven changes in English teaching materials. The compulsory module of the Senior High School English textbook (published by People's Education Press in 2024, hereafter, PEP) is a typical example of the multimodal trend. It was born in response to the evolving times and the needs of teaching, offering a fresh perspective and innovative resources for English education.

Grounded in Kress and Van Leeuwen's Visual Grammar (2020), the research conducts a multimodal discourse analysis (MDA) of the "Reading and Thinking" section in high school English textbooks published by People's Education Press. The aim is to explain how these multimodal discourses achieve metalanguage functions and modal meanings, revealing how the social symbols in the textbooks construct representational, interactive, and compositional meanings. Such analysis could support potential readers, such as teachers and students, in interpreting multimodal discourse, exploring the layered meanings embedded in teaching materials, raising students' awareness of multimodal literacy, and ultimately enhancing the effectiveness of these materials.

2. Literature Review

Research on multimodality has expanded rapidly over the last two decades, and its development has significantly shaped the way textbooks are examined as pedagogical and semiotic artefacts. This section brings together work on multimodal discourse analysis (MDA) and textbook studies, that is, fields that have often been reviewed separately, to show how they converge in the analysis of English teaching materials and to position the present study within this evolving landscape.

2.1. Multimodality and Multimodal Discourse Analysis

Multimodality emphasises that meaning is produced through the interplay of multiple modes, such as linguistic, visual, auditory, and spatial resources. Among the many approaches, Halliday's Systemic Functional Linguistics (1994; hereafter SFL) and Kress and van Leeuwen's Visual Grammar (1996; 2001; 2020) have been widely applied in applied linguistics, communication studies, and educational research because they offer concrete analytical tools for examining how images and language co-construct meaning. According to their theory, the grammar of visual design could be presented as a set of narrative and conceptual processes (Kress & van Leeuwen, 2020). Moreover, contact, social distance, and attitudes (or perspectives) are three factors in image-text interactions from the viewers' perspective (Kress & van Leeuwen, 2020). Lastly, information value, framing, and salience are the meanings of composition (Kress & van Leeuwen, 2020).

Building on these frameworks, Zhu and Yan (2011) proposed two criteria for identifying multimodal discourse:

(1) whether more than one mode is present, and
(2) whether these modes interact to convey meaning that cannot be accomplished by a single mode alone.

These criteria are directly relevant to the current study, as English textbooks typically organise linguistic and visual modes in integrated layouts, and the analytical focus lies not only on identifying these modes but also on understanding how they jointly contribute to pedagogical meanings.

Research abroad has engaged with similar concerns. Early work explored how images perform ideational and interpersonal functions in school materials (Jewitt, 2009; Souryana, 2012), and later studies investigated how layout, typography, and page design guide learners' reading paths and meaning making (Torres, 2015). These contributions treat textbooks as socially situated semiotic ensembles rather than vessels of neutral knowledge. Compared with these, many Chinese studies have tended to emphasise theoretical articulation, such as mapping visual grammar categories onto textbook examples or discussing how multimodal resources support curriculum goals (Cheng & Cong, 2020; Cheng & Zhang, 2017; Li, 2009; Zhang, 2009a, 2009b; Zhang, 2012; Zhang & Zhang, 2010). While important, some of these studies rely heavily on descriptive analysis and less on linking multimodal patterns to broader educational or ideological implications.

2.2. Textbook Research from a Multimodal Perspective

Textbooks have long been seen as carriers of cultural and linguistic knowledge, but the multimodal turn has brought renewed attention to how semiotic design shapes learning experiences. International research has examined issues such as the representation of gender, ethnicity, and culture (Gheisari & Akbari, 2024; Hidayah, 2025; Prihatiningsih et al., 2021; Xiong et al., 2022), as well as how multimodal ensembles position learners through gaze, social distance, and interaction patterns (Karadayı, 2022; Jauhara et.al., 2021; Smith, 2024; Torres, 2015). Studies have also questioned how visual and linguistic modes may reinforce or challenge particular worldviews, highlighting the ideological role of school materials (Jewitt, 2009; Smith, 2022; Souryana, 2012; Weninger, 2021; Xiong et al., 2022).

Studies of textbooks in China exhibit a similar trajectory but with varying emphases (Li, 2003; Li & Lu, 2012; Li, 2013; Xu & Feng, 2022). Numerous studies focus on aligning linguistic and visual modalities with national curriculum standards or on the adequacy of illustrations in

facilitating language acquisition and instructional activities. Some studies examine narrative coherence across images and texts, while others comment on the aesthetic or motivational value of illustrations. Although these works provide descriptive accounts that are applicable, fewer studies examine the pedagogical metalanguage embedded in multimodal design or how visual-verbal configurations guide learners' interpretation of social meanings.

2.3. Identifying the Gaps

Although multimodal approaches have been widely applied to English textbook research, most existing domestic studies were conducted before the release of the 2024 edition of the English Curriculum Standards and therefore focus on earlier generations of teaching materials. These works have offered valuable insights into how linguistic and visual modes contribute to meaning making. Nevertheless, few have examined whether the revised curriculum's emphasis on "viewing", which as a core component of students' multimodal literacy in English Curriculum Standards, is reflected in the newest set of compulsory textbooks.

The modified Chinese curriculum standards specifically referenced the "*viewing*" skills necessary for students to interpret images, layouts, and other visual resources as components of meaning construction (MOE, 2022). This stimulates inquiries regarding the efficacy of the semiotic signs in the 2024 PEP textbooks in achieving this objective, and whether the visual images correspond with the principles outlined in Visual Grammar (Kress & van Leeuwen, 2020), including representational, interactive, and compositional meanings. Although prior studies have established the efficacy of visual-grammar-based analyses, their implementation in the post-reform textbook series has not been examined.

With this context in mind, the present study is guided by the following research questions:

RQ1: How do linguistic and visual modes interact to construct meaning in the 2024 edition of the compulsory English textbooks?

RQ2: To what extent do the images realise representational, interactive, and compositional meanings consistent with the principles of visual grammar?

By addressing these questions, the study aims to offer an updated account of textbook multimodality in light of the most recent curriculum reform and to consider whether current visual design practices effectively support students' multimodal literacy development.

3. Materials and Methods

3.1. Research Materials and Procedure

This study focuses on the *Reading and Thinking* section in the currently used compulsory English textbooks in senior high schools. These sections were chosen because they contain the most consistent combination of linguistic input and visual support in reading and therefore offer a stable basis for assessing how visual resources contribute to meaning making. Across the three textbooks, each unit includes several reading passages accompanied by illustrative materials. After an initial survey of all visual elements, 25 images were selected for analysis. The selection criteria were:

(1) the image must directly accompany a reading text in *Reading and Thinking*;
(2) it must serve a pedagogic or interpretive function rather than decorative purposes; and
(3) it must present sufficient visual detail to allow meaningful analysis through visual grammar.

These criteria ensured that the self-built corpus represents the types of visuals that students encounter when engaging with the main reading content.

The illustrations in these sections are typically static images designed to support textual interpretation, activate background knowledge, or guide students' understanding of cultural and situational context. All images meeting the above criteria were included rather than being sampled through frequency or thematic categories in that the study aims to explore how visual design contributes to multimodal meaning-making.

3.2. Data Analysis

Following practices in multimodal image studies, the analysis combines descriptive coding with interpretive analysis. The process was carried out in three steps.

Step one: Each image was manually categorized according to its basic features, including type (e.g., photograph, drawing), placement on the page, and relation to the reading text. This stage ensured that the corpus was clearly organised before further analysis.

Step two, a coding scheme based on Kress and van Leeuwen's visual grammar (2020) was applied. The scheme included categories for representational meaning (participants, processes, circumstances), interactive meaning (gaze, distance, angle), and compositional meaning (information value, salience, framing). All the coding was conducted manually and reported in a spreadsheet. No automated image-coding tools were used, as the analytical framework requires interpretive judgement rather than pixel-level annotation. However, coding sheets were used to maintain a systematic process and ensure consistency across images.

Finally, the coded descriptions were interpreted qualitatively to identify how linguistic and visual modes interact and how the visual design relates to curriculum expectations for *viewing* competence of the Chinese New Curriculum Standard. This stage involved comparing

patterns across units and examining how recurring visual choices support or constrain students' multimodal literacy development.

3.3. Ethical Considerations

During the study, only the images in the textbooks were scanned and analyzed; their contents were not copied, altered, or disseminated in any way without authorization. When citing images from textbooks, the source should be clearly indicated, and academic norms and copyright regulations must be strictly followed. Therefore, there is no copyright infringement in this study, and it fully complies with the ethical requirements of academic research.

4. Results

4.1. Overall Descriptive Results

Tables 1–3 below are an overall descriptive summary based on the analytic framework of Visual Grammar (Kress & van Leeuwen, 2020). According to Table 1, the symbolic process in conceptual representation is most prevalent among 25 images, at 48%. Table 2 showed that the selected images reflect interactive meanings mainly through "offer" contact (92%), impersonal distance (76%), and primarily at an eye-level (64%). In an interactive sense, "offering" a contact dominates, indicating that images primarily provide information to the viewer rather than seek interaction. For example, the image of a virtual choir invites viewers to join a global community and shows members singing together, making students feel the warmth of connection. Table 3 showed that information value is mainly conveyed through left-right and top-bottom structures, as in compositional meanings, accounting for 52% and 44% respectively. The image-text structures that go from left to right could make contrasts, while structures that go from top to bottom can draw attention to main objects, that is, the linguistic text itself. This layout aligns with how people read and helps students understand the text, which is centered and supported by images.

Table 1. The summary of Representation Meaning in compulsory textbooks.

The Grammar of Visual Design		
Classification		**Total (%)**
Narrative Representation	Action Process	8 (32%)
	Reactional Process	2 (8%)
	Speech and Mental Process	0 (0%)
	Total	10 (40%)
Conceptual Representation	Classificational Process	3 (12%)
	Analytical Process	0 (0%)
	Symbolic Process	12 (48%)
	Total	15 (60%)

Table 2. The summary of Interactive Meaning in compulsory textbooks.

Interactive Meaning		
Classification		**Total (%)**
Contact	Demand	2 (8%)
	Offer	23 (92%)
Total		25 (100%)
Social Distance	Intimate	2 (8%)
	Social	4 (16%)
	Impersonal	19 (76%)
Total		25 (100%)
Perspective	High Angle	5(20%)
	Eye Angle	16 (64%)
	Low Angle	4 (16%)
Total		25 (100%)

Table 3. The summary of Compositional Meaning in compulsory textbooks.

Compositional Meaning		
Classification		**Total (%)**
Information value	Left and Right	13 (52%)
	Top and Bottom	11 (44%)
	Center and Margin	1 (4%)
Total		25 (100%)
Framing	Weak Framing	11 (44%)
	Actual Framing	14 (56%)
Total		25 (100%)

4.2. Representational Meaning

According to the theory of Visual Grammar (Kress & van Leeuwen, 2020), representational meaning is categorized into narrative and conceptual representations. To examine the features of representational meaning construction in images, statistics on narrative and conceptual representation in the PEP for senior high school textbooks are presented and discussed. Tables 4 and 5 present the detailed statistics.

As shown in Tables 4 and 5, images of conceptual representation are more prominent than narrative representation. The prevalence of action processes and symbolic processes in the three compulsory textbooks can be attributed to their effectiveness in promoting student engagement and understanding. To teach content that is more relevant to students' lives, the action process helps them grasp the text more visually by depicting dynamic human interactions. Symbolic processes convey deeper cultural or thematic meanings, such as festivals or human landscapes, enriching the text's contextual framework and developing students' cultural awareness and critical thinking. Thus, all these processes embodied in textbooks can be effective in improving comprehension and educational outcomes.

Table 4. The statistics of the image of narrative representational (NR).

NR / Textbook Volume	Action Process	Reactional Process	Speech and Mental Process	Total
Compulsory I	1	0	0	1
Compulsory II	5	0	0	5
Compulsory III	2	2	0	4
Total (%)	8 (32%)	2 (8%)	0	10 (40%)

Table 5. The statistics of images of conceptual representation (CR).

CR / Textbook Volume	Classificational Process	Analytical Process	Symbolic Process	Total
Compulsory I	1	0	7	8
Compulsory II	2	0	0	2
Compulsory III	0	0	5	5
Total (%)	3 (12%)	0 (0%)	12 (48%)	15 (60%)

Symbolic processes in conceptual representation account for 48%. Unlike narrative representation images, conceptual representation images focus on classifying, analyzing, and symbolizing participants. This helps deepen the viewer's understanding of the participants in the images. For instance, the following Figure 1 is selected from Unit 1 of Compulsory II. 2 images were shown alongside the reading titled "*From problems to Solutions*". The upper image shows what a real "dam" looks like in Egypt, as mentioned in the reading material; the picture supports the reading content symbolically by visually representing the concept of a "dam" and its cultural relics. Therefore, it is considered a symbolic process of conceptual representation function. The lower image shows a scene from the Aswan Dam project, where travellers are visiting the temples mentioned in the text. It also symbolizes the specific dam mentioned in the text, which might not be so familiar to the Chinese young students. These pictures visualize the significant contribution of the Aswan Dam project to the preservation of cultural heritage.

As in Figure 2, a map of Peru on the right side helps explain the country's geography and history by using visual elements that match the left-sided text. It uses different colors and symbols to represent the three central regions, namely the coastal region, the Andes, and the Amazon rainforest. In this map, the yellowish areas symbolized "*the narrow, dry coastal land*", and greenish areas represented the "*Amazon rainforest*" in the reading text. Mountains are labeled with icons to help readers easily locate "*the Andes*". This visual presentation is directly linked to the reading text, helping students understand Peru's natural features and historical sites more clearly and deepening readers' understanding of the text.

Figure 3 also originated from the same unit in Compulsory I. The image depicts tourist culture and landscapes that symbolize different countries. These symbolic processes allow readers to make immediate associations and prompt them to recall more about ancient Inca culture, centuries-old Spanish culture, or other wonderful cities full of fantastic architecture. For senior high school students, these images shorten the distance to unfamiliar places in the real world and to the language itself. In this case, readers might not only acquire knowledge from linguistic resources, such as reading texts, but also visualize this knowledge through real maps and images.

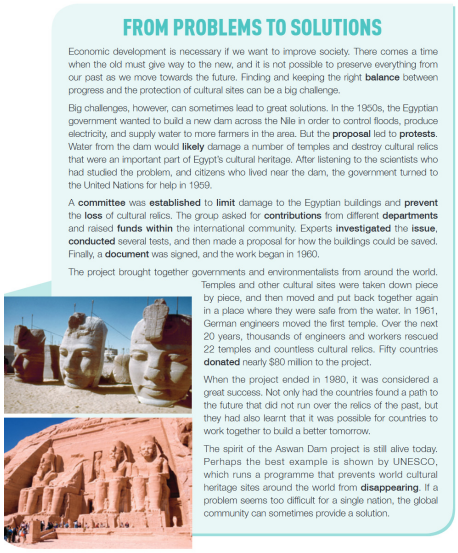


Figure 1. Examples of Conceptual Meanings from Unit 1, Compulsory II.

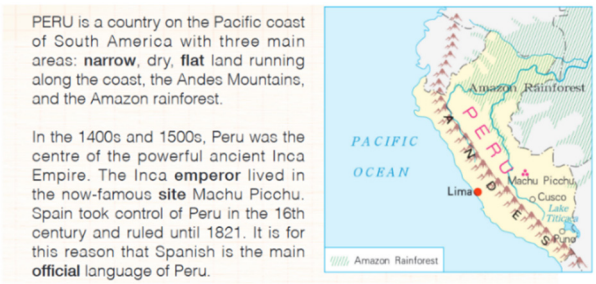


Figure 2. A Map showing Symbolic Process. Extracted from Unit 2, Compulsory I.

As Table 4 shows, action processes and reactional processes are the core processes of narrative representation. The action process accounts for 32% and ranks second, while the reactional process accounts for 8% of the total. Figure 4 shows a doctor looking at a crying baby in an incubator, smiling, with the baby as the target, and the doctor's reaching towards the baby is the response vector. Readers might easily figure out that the doctor's expressions and movements convey emotions such as caring and relief. The eye contact between the doctor and the baby highlights their interaction, from which students can understand the doctor's role as caregiver, as the reading title suggests, *Mother of the Thousand Babies*, and the baby's vulnerability.

To sum up, the "*Reading and Thinking*" section of the compulsory textbooks tends to use images as complements to the linguistic text, most of which depict actional (32%) and symbolic (48%) processes. Therefore, the representational meaning of images can be conveyed through narrative and conceptual representation, enabling readers to better understand and grasp the linguistic texts' central ideas and correct values. Without those images, texts are only texts with limited imagination. With the assistance of images, such as symbolic real pictures and maps, fantasies are brought to life.

Figure 3. An Image of Cusco. Extracted from Unit 2, Compulsory I.

Figure 4. Example of Narrative Meaning from Unit 2, Compulsory III.

4.3. Interactive Meaning

The interactive meaning is categorized into three dimensions: contact, social distance, and perspective. Contact includes demand (eye contact) and offer (a stance) for the readers. Social distance in the interactive function can be further categorized into intimate, social, and impersonal distance to the readers. Perspective can be viewed as high, eye, and low angles. Tables 6, 7, and 8 show the descriptive statistics, respectively.

As shown in Table 6, among the interactive functions of contact with the readers, images "offering a stance" are dominant, accounting for 92%. In such images, which offer "something" to readers, the creatures do not make direct eye contact. The primary interactive function of these pictures was to provide viewers with information, as shown in Figure 5. The text presents "*a virtual choir*", and the real picture underneath extends this invitation to the readers to join the global community. Moreover, this real choir picture "offered" a scene of how different members sing at the same time, that is, a symbolic concept of "*choir*". This type of picture did not invite readers to interact; instead, it offered information.

Table 6. Statistics on images of "contact".

Contact / Textbook Volume	Demand	Offer	Total
Compulsory I	1	8	9
Compulsory II	1	6	7
Compulsory III	0	9	9
Total (%)	2 (8%)	23 (92%)	25 (100%)

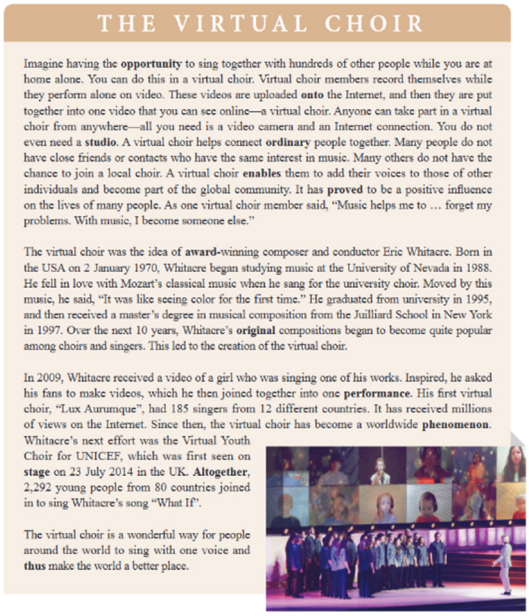
Table 7. The Statistics of images of "social distance" (SD).

Textbook Volume \ SD	Intimate	Social	Impersonal	Total
Compulsory I	1	1	7	9
Compulsory II	0	2	5	7
Compulsory III	1	1	7	9
Total (%)	2 (8%)	4 (16%)	19 (76%)	25 (100%)

Table 8. Statistics on images of "perspectives".

Textbook Volume \ Perspective	High Angle	Eye Angle	Low Angle	Total
Compulsory I	1	8	0	9
Compulsory II	1	4	2	7
Compulsory III	3	4	2	9
Total	5 (20%)	16(64%)	4 (16%)	25 (100%)

In contrast, as in Figure 6, the image shows a male student smiling and looking directly at the readers. This direct gaze creates a sense of demanding interaction in visual grammar, as he appears to actively seek the reader's attention or empathy. Combined with the text, readers can easily assume the male student's name from the first phrase, "*My name is Adam*". In the reading text, Adam introduces himself and shares his experiences as a freshman. This visual image invites viewers to imagine this male student in his role as a college student (wearing a handbag in his left hand while holding some books in his right) and connect with his story. The image and text work together to create an engaging atmosphere, making the readers feel included in Adam's journey. This combination effectively conveys that Adam is open and willing to share his challenges and hopes for the future, as most readers are likely senior high school students.



THE VIRTUAL CHOIR

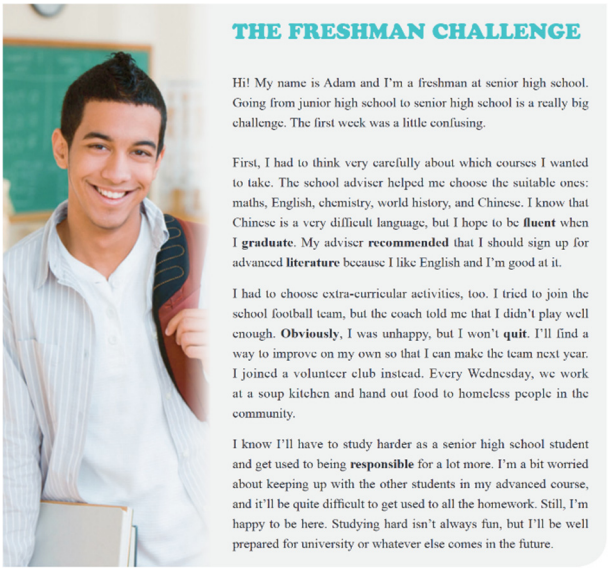
Imagine having the **opportunity** to sing together with hundreds of other people while you are at home alone. You can do this in a virtual choir. Virtual choir members record themselves while they perform alone on video. These videos are uploaded **onto** the Internet, and then they are put together into one video that you can see online—a virtual choir. Anyone can take part in a virtual choir from anywhere—all you need is a video camera and an Internet connection. You do not even need a **studio**. A virtual choir helps connect **ordinary** people together. Many people do not have close friends or contacts who have the same interest in music. Many others do not have the chance to join a local choir. A virtual choir **enables** them to add their voices to those of other individuals and become part of the global community. It has **proved** to be a positive influence on the lives of many people. As one virtual choir member said, "Music helps me to … forget my problems. With music, I become someone else."

The virtual choir was the idea of **award**-winning composer and conductor Eric Whitacre. Born in the USA on 2 January 1970, Whitacre began studying music at the University of Nevada in 1988. He fell in love with Mozart's classical music when he sang for the university choir. Moved by this music, he said, "It was like seeing color for the first time." He graduated from university in 1995, and then received a master's degree in musical composition from the Juilliard School in New York in 1997. Over the next 10 years, Whitacre's **original** compositions began to become quite popular among choirs and singers. This led to the creation of the virtual choir.

In 2009, Whitacre received a video of a girl who was singing one of his works. Inspired, he asked his fans to make videos, which he then joined together into one **performance**. His first virtual choir, "Lux Aurumque", had 185 singers from 12 different countries. It has received millions of views on the Internet. Since then, the virtual choir has become a worldwide **phenomenon**. Whitacre's next effort was the Virtual Youth Choir for UNICEF, which was first seen on **stage** on 23 July 2014 in the UK. **Altogether**, 2,292 young people from 80 countries joined in to sing Whitacre's song "What If".

The virtual choir is a wonderful way for people around the world to sing with one voice and **thus** make the world a better place.

Figure 5. An Image of "Offer" Information. Extracted from Unit 5, Compulsory II.



THE FRESHMAN CHALLENGE

Hi! My name is Adam and I'm a freshman at senior high school. Going from junior high school to senior high school is a really big challenge. The first week was a little confusing.

First, I had to think very carefully about which courses I wanted to take. The school adviser helped me choose the suitable ones: maths, English, chemistry, world history, and Chinese. I know that Chinese is a very difficult language, but I hope to be **fluent** when I **graduate**. My adviser **recommended** that I should sign up for advanced **literature** because I like English and I'm good at it.

I had to choose extra-curricular activities, too. I tried to join the school football team, but the coach told me that I didn't play well enough. **Obviously**, I was unhappy, but I won't **quit**. I'll find a way to improve on my own so that I can make the team next year. I joined a volunteer club instead. Every Wednesday, we work at a soup kitchen and hand out food to homeless people in the community.

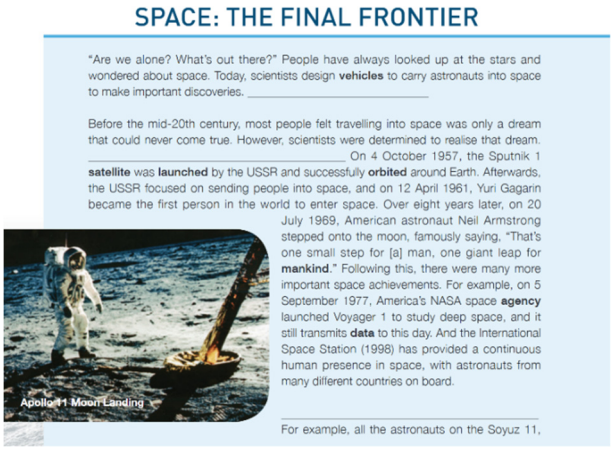
I know I'll have to study harder as a senior high school student and get used to being **responsible** for a lot more. I'm a bit worried about keeping up with the other students in my advanced course, and it'll be quite difficult to get used to all the homework. Still, I'm happy to be here. Studying hard isn't always fun, but I'll be well prepared for university or whatever else comes in the future.

Figure 6. An Image of "Demanding" Interaction. Extracted from Unit 1, Compulsory I.

As shown in Table 7, impersonal distance has the highest proportion at 76%, followed by social distance at 16%, while intimate distance is relatively rare, appearing only in the three compulsory textbooks. Figure 7 documents the final frontier, an image of an astronaut on the moon. This picture's impersonal distance from the astronaut to Apollo 11 (as shown) provides information about the continued presence of humans on the moon. This fosters a desire to explore and learn about the unknown, enhances the spirit of adventure, and inspires a learning attitude.

The least social distance shown in the textbooks could also be intimate. From Figure 8, an elderly female holds a baby and is surrounded by children. Her gentle expression and close physical contact with the baby show the intimate distance. This intimate distance is enhanced by the toddler's relaxed, happy expressions, suggesting a warm, trusting relationship. The image aligns with the

text, which describes Dr. Lin's deep care for patients and her dedication to women and children. The text mentions that she delivered over 50,000 babies and was known as the *"mother of ten thousand babies"*, highlighting her role as a caregiver. The combination of the image and text reinforces the idea that Dr. Lin was not just a medical professional but also a compassionate figure deeply connected to those she helped. This visual and textual pairing effectively conveys her warmth and dedication, making her legacy more personal and relatable to the readers.



SPACE: THE FINAL FRONTIER

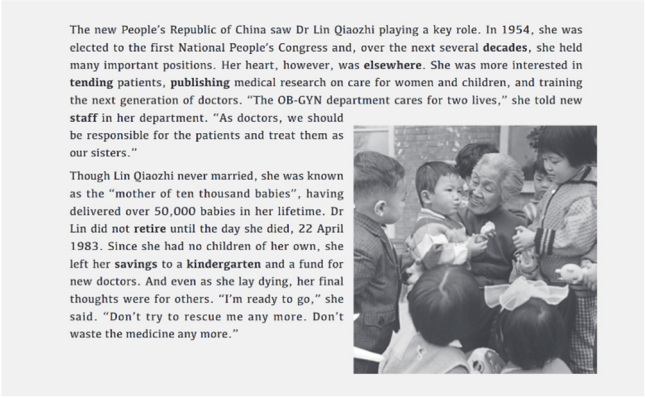
"Are we alone? What's out there?" People have always looked up at the stars and wondered about space. Today, scientists design **vehicles** to carry astronauts into space to make important discoveries. _____

Before the mid-20th century, most people felt travelling into space was only a dream that could never come true. However, scientists were determined to realise that dream. _____ On 4 October 1957, the Sputnik 1 **satellite** was **launched** by the USSR and successfully **orbited** around Earth. Afterwards, the USSR focused on sending people into space, and on 12 April 1961, Yuri Gagarin became the first person in the world to enter space. Over eight years later, on 20 July 1969, American astronaut Neil Armstrong stepped onto the moon, famously saying, "That's one small step for [a] man, one giant leap for **mankind**." Following this, there were many more important space achievements. For example, on 5 September 1977, America's NASA space **agency** launched Voyager 1 to study deep space, and it still transmits **data** to this day. And the International Space Station (1998) has provided a continuous human presence in space, with astronauts from many different countries on board.

Apollo 11 Moon Landing

For example, all the astronauts on the Soyuz 11,

Figure 7. An Image of Impersonal Distance. Extracted from Unit 4, Compulsory III.



The new People's Republic of China saw Dr Lin Qiaozhi playing a key role. In 1954, she was elected to the first National People's Congress and, over the next several **decades**, she held many important positions. Her heart, however, was **elsewhere**. She was more interested in **tending** patients, **publishing** medical research on care for women and children, and training the next generation of doctors. "The OB-GYN department cares for two lives," she told new **staff** in her department. "As doctors, we should be responsible for the patients and treat them as our sisters."

Though Lin Qiaozhi never married, she was known as the "mother of ten thousand babies", having delivered over 50,000 babies in her lifetime. Dr Lin did not **retire** until the day she died, 22 April 1983. Since she had no children of her own, she left her **savings** to a **kindergarten** and a fund for new doctors. And even as she lay dying, her final thoughts were for others. "I'm ready to go," she said. "Don't try to rescue me any more. Don't waste the medicine any more."

Figure 8. An Image of Intimate Distance Extracted from Unit 2, Compulsory III.

As shown in Table 8, among the "perspective" interactive meanings in the three compulsory textbooks, the "eye angle" images occupy an absolutely dominant position, accounting for 64%, while the high angle accounts for 20% and the low angle accounts for 16%.

Generally, when "eye angle" is adopted for filming, the readers and the pictures are at the same eye angle so that the viewers can integrate into the image's moment. The eye-level angle primarily refers to the width of the human eye. If the image in the textbook is presented at eye level, it can shorten the interaction distance between the reader and the object (image), demonstrating an immersive interactive significance. As shown in Figure 9, the presentation of the statues and sculptures was from an eye-level perspective. When reading the text and looking at the pictures, readers might feel closer to the "dams", as if they were in Egypt.

Figure 9. An Eye-angle Image. Extracted from Unit 1, Compulsory II.

In contrast, high-angle pictures enable readers to view the overall situation or the construction of things and events from a high vantage point, as in a map (see Figure 10, also see Figure 2).

WHAT'S IN A NAME?

The United Kingdom, Great Britain, Britain, England—many people are confused by what these different names mean. So what is the difference between them, if any? Getting to know a little bit about British history will help you solve this puzzle.

In the 16th century, the **nearby** country of Wales was joined to the Kingdom of England. Later, in the 18th century, the country Scotland was joined to create the Kingdom of Great Britain. In the 19th century, the Kingdom of Ireland was added to create the United Kingdom of Great Britain and Ireland. Finally, in the 20th century, the southern part of Ireland broke away from the UK, which resulted in the full name we have today: the United Kingdom of Great Britain and Northern Ireland. Most people just use the shortened name: "the United Kingdom" or "the UK". People from the UK are called "British", which means the UK is also often referred to as Britain or Great Britain.

ATLANTIC OCEAN
NORTH SEA
SCOTLAND
Edinburgh
NORTHERN IRELAND
Belfast
IRISH SEA
ENGLAND
Cambridge
Oxford
London
WALES
Cardiff
Bath
Hastings
ENGLISH CHANNEL

The four countries that **belong** to the United Kingdom work together in some areas. They use the same flag, known as the Union Jack, as well as share the same currency and **military defence**. However, they also have some differences. For example, England, Wales, Scotland, and Northern Ireland all have different education systems and **legal** systems. They also have their own traditions, like their own national days and national dishes. And they even have their own football teams for competitions like the World Cup!

The United Kingdom has a long and interesting history. Exploring it can help you understand much more about the country and its traditions. Almost everywhere you go in the UK, you will be **surrounded** by **evidence** of four different groups of people who took over at different times throughout history. The first group, the Romans, came in the first century. They built the first towns and good roads. Next, the Anglo-Saxons arrived in the fifth century. They introduced the beginnings of the English language, and changed the

Figure 10. A High-angle Image. Extracted from Unit 4, Compulsory II.

The last viewing angle is low, occupying 16%. An example is shown in Figure 11, which depicts a graffiti wall in the San Francisco Mission District. The graffiti on the wall looks colorful and vibrant. The low-angle perspective of the image emphasizes the wall's height and grandeur, making the graffiti appear more imposing in the scene. This visual approach aligns with the text, which describes the Mission District as a center of art and culture, specifically noting the influence of the "*Mission School" art movement*. The low-angle shot highlights graffiti's cultural significance, giving viewers a sense of being there and

allowing them to experience the district's lively atmosphere. This combination of visuals and text effectively shows the energy and creativity of the Mission District as an artistic center.

Figure 11. A Low-angle Image. Extracted from Unit 3, Compulsory III.

4.4. Compositional Meaning

Compositional meanings are expressed through three aspects: salience, information value, and framing. According to Kress and van Leeuwen (2020), salience refers to the extent to which an element in a picture attracts the viewer's attention, based on its size and color intensity. This is not an objective measure but depends on the visual principles that create diversity in the image. Therefore, this study only involves analyzing the information value and the framing of compositional meaning. Given most individuals' left-to-right, top-to-bottom reading habits, Kress and van Leeuwen divide the page layout into two parts: top and bottom, left and right, which allows viewers to judge the information value of the pictures. In terms of information value, the distribution of pictures on the top conveys ideal information, on the bottom real information, on the left known information, and on the right, new information (Kress & van Leeuwen, 2020).

As shown in Table 9, among the images of information value in textbooks, the left-right composition accounted for 52%, the top-bottom composition for 44% in second place, and the center and margin had less information.

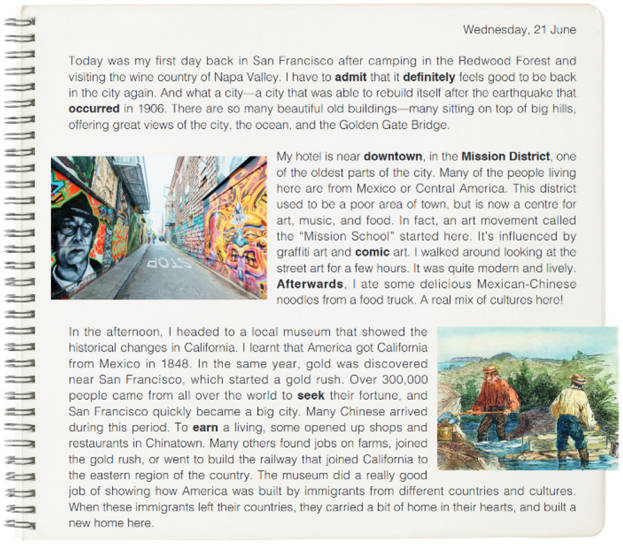
As shown in Figure 12, which is from Unit 3 of Compulsory III. The image and texts are distributed on the left and right, creating a stark contrast and revealing opposing messages. Students were able to perceive the difference between the two images, which helped them better understand the text's structure and focus. However, this study shows that not all images in the left-right follow the pattern of constructing meaning from the known to the unknown, indicating that the explanation of the left-right structure in the visual grammar theory still needs further exploration.

Figure 13 demonstrates another structure: the top-bottom. The character at the top of the image is usually an important element of the text, and this image suggests that the text is mainly centered on the Tibetan antelope to develop the story. Students can glance briefly and deepen their understanding of the text's main character. The title of this article is "*A Day in the Clouds*". This combination of text and images at the top and bottom precisely responds to the content of this article, allowing students to immerse themselves in the idea that the top is the key animals, Tibetan antelopes, and the bottom is the background image of this article.

Although the center-margin structure rarely appears in the three compulsory textbooks, the image in Unit 3 of Compulsory I remains a typical example (see Figure 14). It features colorful silhouettes of athletes in dynamic poses at the center, with a simple background. This use of center and margin in visual grammar positions the athletes as the primary focus, immediately drawing the viewer's attention. The vibrant colors and dynamic movements of the athletes in the center highlight their importance, while the plain background keeps the focus on them. The text below the image introduces Lang Ping and Michael Jordan as "*Living Legends*" of Sport, aligning with the visual emphasis on athletes. The combination of the image and text reinforces the theme that these individuals are central figures in sports, known for their achievements and influence. This visual and textual pairing effectively communicates the significance of the athletes, making it clear to viewers why they are considered legends.

Table 9. Statistics of images of information value (IV).

IV / Textbook Volume	Left and Right	Top and Bottom	Center and Margin	Total
Compulsory I	3	5	1	9
Compulsory II	1	6	0	7
Compulsory III	9	0	0	9
Total (%)	13 (52%)	11 (44%)	1 (4%)	25 (100%)



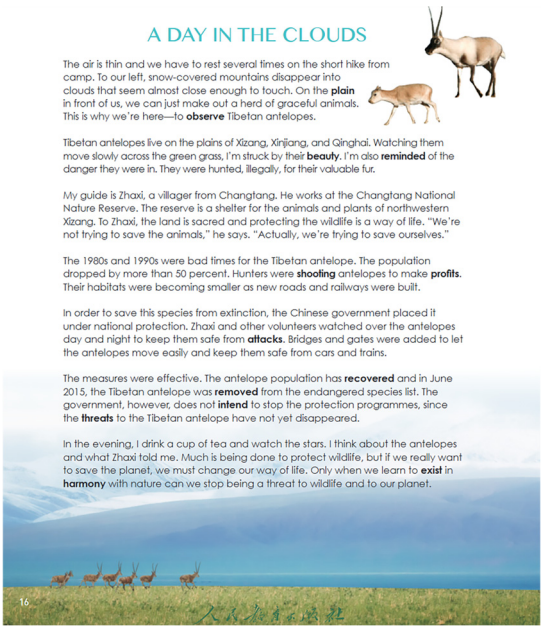
Wednesday, 21 June

Today was my first day back in San Francisco after camping in the Redwood Forest and visiting the wine country of Napa Valley. I have to **admit** that it **definitely** feels good to be back in the city again. And what a city—a city that was able to rebuild itself after the earthquake that **occurred** in 1906. There are so many beautiful old buildings—many sitting on top of big hills, offering great views of the city, the ocean, and the Golden Gate Bridge.

My hotel is near **downtown**, in the **Mission District**, one of the oldest parts of the city. Many of the people living here are from Mexico or Central America. This district used to be a poor area of town, but is now a centre for art, music, and food. In fact, an art movement called the "Mission School" started here. It's influenced by graffiti art and **comic** art. I walked around looking at the street art for a few hours. It was quite modern and lively. **Afterwards**, I ate some delicious Mexican-Chinese noodles from a food truck. A real mix of cultures here!

In the afternoon, I headed to a local museum that showed the historical changes in California. I learnt that America got California from Mexico in 1848. In the same year, gold was discovered near San Francisco, which started a gold rush. Over 300,000 people came from all over the world to **seek** their fortune, and San Francisco quickly became a big city. Many Chinese arrived during this period. To **earn** a living, some opened up shops and restaurants in Chinatown. Many others found jobs on farms, joined the gold rush, or went to build the railway that joined California to the eastern region of the country. The museum did a really good job of showing how America was built by immigrants from different countries and cultures. When these immigrants left their countries, they carried a bit of home in their hearts, and built a new home here.

Figure 12. A Left-Right Image-Text Example. Extracted from Unit 3, Compulsory III.



A DAY IN THE CLOUDS

The air is thin and we have to rest several times on the short hike from camp. To our left, snow-covered mountains disappear into clouds that seem almost close enough to touch. On the **plain** in front of us, we can just make out a herd of graceful animals. This is why we're here—to **observe** Tibetan antelopes.

Tibetan antelopes live on the plains of Xizang, Xinjiang, and Qinghai. Watching them move slowly across the green grass, I'm struck by their **beauty**. I'm also **reminded** of the danger they were in. They were hunted, illegally, for their valuable fur.

My guide is Zhaxi, a villager from Changtang. He works at the Changtang National Nature Reserve. The reserve is a shelter for the animals and plants of northwestern Xizang. To Zhaxi, the land is sacred and protecting the wildlife is a way of life. "We're not trying to save the animals," he says. "Actually, we're trying to save ourselves."

The 1980s and 1990s were bad times for the Tibetan antelope. The population dropped by more than 50 percent. Hunters were **shooting** antelopes to make **profits**. Their habitats were becoming smaller as new roads and railways were built.

In order to save this species from extinction, the Chinese government placed it under national protection. Zhaxi and other volunteers watched over the antelopes day and night to keep them safe from **attacks**. Bridges and gates were added to let the antelopes move easily and keep them safe from cars and trains.

The measures were effective. The antelope population has **recovered** and in June 2015, the Tibetan antelope was **removed** from the endangered species list. The government, however, does not **intend** to stop the protection programmes, since the **threats** to the Tibetan antelope have not yet disappeared.

In the evening, I drink a cup of tea and watch the stars. I think about the antelopes and what Zhaxi told me. Much is being done to protect wildlife, but if we really want to save the planet, we must change our way of life. Only when we learn to **exist** in **harmony** with nature can we stop being a threat to wildlife and to our planet.

16

Figure 13. A Top-Bottom Image-Text Example. Extracted from Unit 2, Compulsory II.

Since the main components of the textbook are images and texts, the framing discussed here refers to the distinction between them, which is primarily realized through virtual and actual frame lines. As shown in Table 10, the virtual framing line has a higher percentage (56%) than the actual segmentation line (44%). That is because using tangible frame lines to distinguish elements in images facilitates students' better understanding of textual semantics, and textual illustrations are generally closely related to the text's content. This effect is achieved through comparisons of color, size, and position, which facilitate students' intuitive understanding of textual content.

Make predictions

Look at the pictures and think about the content before reading. This helps you to relate what you read to what you already know and to understand the new text.

Help us choose some "Living Legends of Sports". They must be **athletes** who are **masters** in their sports and also set good examples for others. Here are our first two choices.

LIVING LEGENDS

Lang Ping

As a player, Lang Ping brought **honour** and glory to her country. As a coach, she led the China women's volleyball team to **medals** at world championships and the Olympics. As a person, Lang Ping is loved by fans at home and abroad. When the Chinese team was preparing for the 2015 World Cup, her determination was tested. The team that Lang Ping had built was falling **apart**. One of the best players had been injured, and the team **captain** had to leave because of heart problems. Losing two important players was a big challenge, but Lang Ping did not lose heart. She had faced difficulties before, and she knew that her young players could win if they worked together as a team. Two weeks later, they were world **champions**! Then in 2016, Lang Ping led her volleyball team to Olympic gold in Brazil.

Michael Jordan

When Michael Jordan's feet left the ground, time seemed to stand still. The player who became known as "Air Jordan" changed basketball with his graceful moves and jumps. Jordan's skills were impressive, but the mental **strength** that he showed made him unique. In the final seconds of a game, Jordan always seemed to find a way to win. Jordan says that the secret to his success is learning from his **failures**. "I can accept failure; everyone fails at something. But I can't accept not trying." Losing games taught him to practise harder and never give up. In life, Jordan has learnt to share his success with others. The Boys and Girls Club which he started in Chicago has been helping young people since 1996.

Send your suggestions for "Living Legends of Sports" to LLS@sports.net.

Figure 14. An Example of Central-Margin Extracted from Unit 3, Compulsory I.

As shown in Figure 15, calligraphy scrolls are used as borders to modify the image of calligraphy, creating a peaceful, elegant atmosphere, as if the students were in the presence of the artist, feeling the depth of calligraphy's art.

On the contrary, the use of weak framing also allows the viewer's attention to focus on the content, or the images, such as antelopes themselves in the previous Figure 13, as there are no strong borders or distractions. The antelopes are positioned dynamically, moving across the frame, creating a sense of freedom and a natural habitat. Weak framing scaffolds students' focus more on the story of protecting the antelopes' habitat.

Table 10. Statistics on images of "Framing".

Framing / Textbook Volume	Actual Frame Line	Virtual Dividing Line	Total
Compulsory I	7	2	9
Compulsory II	4	3	7
Compulsory III	0	9	9
Total	11(44%)	14 (56%)	25(100%)

Figure 15. An Example of an Actual Frame Image. Extracted from Unit 5, Compulsory I.

5. Discussion

The descriptive analysis of this study portrayed the interplay between linguistic and visual modalities in the PEP senior high school English textbooks. In response to RQ1, the analysis indicates that meaning-making is predominantly text-driven, with images primarily serving to reinforce or symbolize the presented information. This focus on text is not just a design choice; it also shows a deeper ideology about teaching. It implies that the compilers still see language input as the primary way to learn, while visual elements serve primarily as scaffolding. Such a structure can be interpreted in two ways: on one hand, it may reinforce traditional logocentrism; on the other hand, given the reading-focused demands of high school English, it may represent a pragmatic strategy aligned with students' immediate learning needs.

In relation to RQ2, the analysis of representational, interactive, and compositional meanings shows that the images realize the meta-functions described in Visual Grammar but in a relatively conservative manner. Representative meanings of these images in the selected textbooks are more symbolically displayed, meaning the images help situate students from abstract text to concrete scenarios that mirror the textual content. Narrative images appear less frequently, suggesting less use of visuals to convey storyline processes. Among these images, the interactive meanings between the image and the audience rely heavily on "offer" contact and mostly display impersonal social distance. These design choices limited interpersonal engagement and positioned images as information generators rather than social interactors. Compositional designs further reinforce the subordinate role of visual "aides": text occupies central positions, while images are placed left-to-right or top-to-bottom, visually signaling their supportive functions. In combination, these meanings suggest a stable multimodal pattern in which images enhance but seldom contest the textual mode.

The recent study by Qi (2024) highlights that the Yilin EFL series provides greater insight into textbook analysis. The Yilin series presents more engaging, direct visuals, such as cartoons and expressive images. In contrast, the current study reveals a more nuanced visual approach in the PEP series, emphasizing realistic, informative imagery. The graphics of the PEP series do not aim to captivate viewers' attention; instead, they facilitate reading comprehension and are particularly focused on conveying the content's abstract or possibly unfamiliar meaning. These trends demonstrate that Chinese EFL textbooks differ not only in the genres of images employed but also in their conceptualization of multimodal literacy. The PEP series appears to be based on the premise that multimodality should clarify themes and concepts, whereas the Yilin series emphasizes emotional engagement and authenticity across contexts. Incorporating this comparison into the discussion situates the current findings within the framework of existing multimodal research rather than distinguishing them from analogous studies.

The results also add to the ongoing discussions in MDA and textbook design. The study empirically validates the enduring significance of SFL-based multimodal frameworks in EFL curriculum research by illustrating that the meta-functional categories of visual grammar remain analytically effective for newly revised textbooks. Additionally, although previous work has investigated earlier versions of the PEP textbooks or concentrated on junior-secondary resources, analyses of the 2024 revised curriculum materials remain scarce. The present study, therefore, offers an updated account of how multimodal principles are enacted after curriculum reform, particularly in relation to the new emphasis on "*viewing*" as a core language skill. The limited multimodal design identified in the present research findings has prompted additional inquiry into the potential for future versions of EFL compulsory textbooks to adopt a more equal multimodal engagement.

Overall, this study's results demonstrate that although the PEP series textbooks' image-text structures maintain coherence and pedagogical clarity, they may inadequately leverage the interactive and interpretive potential of images themselves. A more profound integration of linguistic and visual modalities, surpassing a mere supportive role, could enhance students' analytical, interpretative, and cross-cultural *viewing* competencies, which the new curriculum seeks to promote. Research on other in-use textbook series could and would reveal a broader spectrum of image genres, explore diverse layout techniques, or use graphics that encourage interpretation rather than solely depicting text to improve students' multimodal literacy. These ideas index possible avenues for

future research on the construction of image-text meaning in textbooks.

6. Pedagogical Implications

Although the present study focuses on describing how images and written text work together in the newly released compulsory English textbook, the findings also offer several insights for classroom pedagogical practices. One clear implication is that visual materials in the selected textbooks are no longer merely decorative. Many of the images carry representational and compositional cues that help guide learners' attention and frame the reading task. Teachers may therefore benefit from drawing students' attention to these visual choices, for instance, by encouraging them to discuss what information is highlighted, how objects or subjects are positioned in the images, or how images relate to the surrounding text. Such discussions can scaffold students' development of the "*viewing*" competence emphasised in the new curriculum standard.

Another implication concerns the integration of reading and *viewing*. Since many images subtly prepare students for the themes and perspectives developed in the written passages, teachers can make fuller use of these resources by designing pre-reading or prediction activities or tasks based on the visuals. This approach may assist students in entering the written passages with clearer expectations and start developing awareness of multimodal interpretation, which is increasingly essential for digital literacy in the contemporary era.

Finally, the analysis suggests that not all visual elements fully realise the meanings outlined in the visual grammar framework. This offers space for teachers and textbook users to introduce supplementary tasks, such as comparing similar visual representations from new(s) media or online platforms, to help students understand alternative means of structuring meaning visually. In this way, textbook images serve as a starting point for broader multimodal literacy rather than as the sole material students rely on.

7. Limitations

This study examines the multimodal discourse analysis of the "*Reading and Thinking*" section of the senior high school English textbooks. It does not cover other sections of these textbooks or compare multimodality usage across different textbooks due to space limitations. The multimodal discourse analysis of the other sections may also help teachers use the textbooks better. Additionally, analyzing differences in multimodal use across existing high school textbooks may help teachers select more suitable textbooks for high school students and enhance their comprehensive literacy; therefore, this type of research still needs more attention from scholars and teachers.

Author Contributions

Y.Z. & Y.L.: conceptualization, methodology, software; Y.Z.: data curation, writing, original draft preparation; Y.L.: supervision, writing, reviewing, and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

All materials analyzed in this study are publicly published senior high school English textbooks from People's Education Press (2024 edition). These textbooks can be obtained through standard academic or commercial channels (e.g., university libraries, the publisher's website, or major book retailers). Specific titles and ISBNs of the three volumes we examined are listed below; readers may refer directly to these editions for verification: Senior High School English (Book 1). People's Education Press. ISBN: 978-7-107-32012-3; Senior High School English (Book 2). People's Education Press. ISBN: 978-7-107-32013-0; Senior High School English (Book 3). People's Education Press. ISBN: 978-7-107-32014-7. No proprietary or restricted data were used.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

Cheng, R. L., & Zhang, D. L. (2017). Multimodal Discourse Analysis in China: Taking Journal Articles as an Example. *Foreign Languages in China*, *14*(3), 9–11.

Cheng, X. T., & Cong, L. (2020). The Design and Use of Imagery Resources in the Writing of English textbooks. *Curriculum, Teaching Material and Method*, *40*(8), 78–85.

Gheisari, M., & Akbari, O. (2024). A comparison of cultural representation and ideologies in the multimodal discourses of textbooks used in public and private Iranian contexts: A cross-textual study. *Social Semiotics*, *34*(2), 289–309.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed., pp. 108–116). Edward Arnold.

Hidayah, N. (2025). Visionary Leadership and Its Impact on the Quality of Education in Islamic Educational Institutions. *Al-Hayat: Journal of Islamic Education*, *9*(1), 142–158.

Jauhara, D., Emilia, E., & Lukmana, I. (2021). Re-contextualising 'Greeting': A multimodal analysis in an EFL textbook. *Indonesian Journal of Functional Linguistics*, *1*(1), 15–24.

Jewitt, C. (2009). *Handbook of multimodal analysis* (pp. 57–68). Routledge.

Karadayı, H. (2022). *A Multimodal Analysis of Pictures in an ELT Coursebook: Teenwise 9th Grade Student's Book* [Master's Thesis, Maltepe University]

Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design* (pp. 177–187). Routledge.

Kress, G., & van Leeuwen, T. (2001). Multimodal text: The modes and media of contemporary communication. *College Composition and Communication*, *54*(2), 142.

Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (2nd ed., pp. 201–265). Routledge.

Li, Q. Q. (2009). *A multimodal discourse analysis of textbook discourse* [Master's Thesis, Xiamen University].

Li, Z. Z. (2003). Social semiotic analysis of multimodal discourse. *Foreign Languages Research*, *5*, 1–8.

Li, Z. Z., & Lu, D. Y. (2012). Multimodal semiotics: Theoretical foundation, research approaches and prospects. *Foreign Languages Research*, (2), 1–8.

Ministry of Education of the People's Republic of China. (2022). *Compulsory education English curriculum standards* (2022 ed.). Beijing Normal University Press.

Prihatiningsih, F., Petrus, I., & Silvhiany, S. (2021). Cultural representation in EFL textbooks for the seventh graders: A multimodal analysis. *Lingua Cultura*, *15*(1), 121–133.

Qi, M. (2024). Analyzing image meanings in Chinese EFL textbooks: A multimodal perspective. *Theory and Practice in Language Studies*, *14*(8), 2498–2509.

Smith, C. (2022). *Understanding Multimodal Discourses in English Language Teaching Textbooks: Implications for Students and Practitioners* (1st ed.). Bloomsbury Publishing Plc. https://doi.org/10.5040/9781350256989.

Smith, C. A. (2024). Mapping reflexivity in situ: A multimodal exploration of negotiated textbook discourses in Korean university EFL classrooms. *Language Teaching Research*, *28*(4), 1368–1390.

Souryana, Y. (2012). Culture Issues, Ideology and Otherness in EFL Textbooks: A Social Semiotic Multimodal Approach [Doctoral dissertation, Université Mouloud Mammeri].

Torres, G. (2015). Reading'world link: A visual social semiotic analysis of an EFL textbook. *International Journal of English Language Education*, *3*(1), 239–253.

Weninger, C. (2021). Multimodality in critical language textbook analysis. *Language, Culture and Curriculum*, *34*(2), 133–146.

Xiong, T., Feng, D., & Hu, G. (Eds.). (2022). *Cultural knowledge and values in English language teaching materials: (multimodal) representations and stakeholders*. Springer. https://doi.org/10.1007/978-981-19-1935-0.

Xu, Y., & Feng, D. (2022). Values in Business English Textbooks: A Multimodal Analysis Approach. In T. Xiong, G. Hu, & D. Feng (Eds.), *Cultural Knowledge and Values in English Language Teaching Materials* (pp. 21–40). Springer. https://doi.org/10.1007/978-981-19-1935-0_2

Zhang, D. L. (2009a). Multimodal Discourse Theory and Its Application to Foreign Language Teaching with Modern Media Technology. *Foreign Language Education*, *30*(4), 15–20.

Zhang, D. L. (2009b). On A Synthetic Theoretical Framework for Multimodal Discourse Analysis. *Foreign Languages in China*, *6*(1), 24–30.

Zhang, D. L. (2012). Exploration of multimodal learning competence cultivation model. *Foreign Languages Research*, (2), 9–14.

Zhang, D. L., & Zhang, S. J. (2010). An investigation into the compilation principles of multimodal foreign language textbooks. *Foreign Language World*, (5), 26–33.

Zhu, Y., & Yan, S. (2011). *Rethinking systemic functional linguistics* (p. 120). Fudan University Press.