

Original Paper

Communicative Functions of Textbook Images: A Multimodal Analysis of *New Standard English* (Selective Compulsory Book 2)

Shiman Zhang¹

¹ Harbin Institute of Technology, Harbin, Heilongjiang, China

Received: February 12, 2026
doi:10.22158/eltls.v8n3p43

Accepted: April 22, 2026

Online Published: May 7, 2026

URL: <http://dx.doi.org/10.22158/eltls.v8n3p43>

Abstract

With the increasingly prominent position of multimodal ability in high school English education, the function carried by textbook images is far more than page decoration. They are no longer just a visual accompaniment of written texts, but participate in meaning construction and classroom communication in a more direct way. This study takes New Standard English (Selective Compulsory Volume II) as the object of analysis, explores the communication function of textbook images under the theoretical framework of visual grammar, and adopts multimodal discourse analysis methods to examine how different visual resources promote the understanding and interpretation of textbook discourse.

The analysis found that the images of the textbook generally play a positive role in discourse understanding, cultural awareness cultivation and value guidance in English learning, but the focus of each function is not exactly the same. Some images mainly help learners enter specific discourse situations and grasp the text background more smoothly; others highlight the cultural content with visual details such as clothing, character behaviour, natural appearance and surrounding environment. There is also a class of images that are more closely related to emotional response and attitude guidance. In contrast, some images choose to use symbols, visual metaphors or coherent series of images to clearly explain abstract concepts. When abstract concepts are transformed into more intuitive image forms, learners tend to get a more transparent understanding experience. Frankly speaking, the generation of meaning in these images does not rely on an isolated visual factor, but on the coordinated operation of elements such as mood, salignance, narrative reproduction and image organisation, which jointly affect the generation and interpretation of meaning.

At the same time, the analysis shows that there are still some shortcomings in the images of the textbook. The interaction between some images and the audience is relatively limited, while others fail

to form a sufficiently close semantic association with the supporting text. In some cases, the unclear reproduction structure may even interfere with the learner's grasp of the meaning of the image. Based on the above findings, this study puts forward the following suggestions to teachers: in classroom teaching, students should not only be allowed to passively browse the illustrations of textbooks, but should be guided to think about the teaching value and interpretation functions of these images, and students should be encouraged to actively explore visual information. At the same time, textbook writers also need to work harder on layout design to make the connection between images and text and interaction design smoother and more natural. Only in this way can visual resources truly help discourse understanding, cross-cultural cognition and value guidance in English learning.

Keywords

Visual Grammar, High School English Textbook, Textbook Image, Communication Function, Multimodal Discourse Analysis

1. Introduction

Under the background of the continuous promotion of basic education reform in China, the role of visual resources in English education has changed significantly. Today's language learning no longer focusses only on the understanding of vocabulary, grammar or written texts. Students' ability to understand multimodal discourses and process information on different symbolic resources is receiving more and more attention. General High School English Curriculum Standards (2017 Edition, Revised in 2025) clearly points out that in addition to listening, speaking, reading and writing, language ability also includes "seeing", that is, the ability to interpret meaning through images, charts, symbols, animations, videos and other visual forms. Against this background, textbook images have gradually become an important part of discourse construction and classroom meaning generation.

The research on textbook images in foreign academics started relatively early, and the existing results are mainly concentrated in two major areas: the role of images in understanding and the construction mechanism of visual meaning. Peeck (Peeck, 1993) pointed out that illustrations affect learners' processing of explanatory text, and their function is not only at the decorative level. After that, Kress and van Leeuwen (Kress, G., & van Leeuwen, 2006) proposed the theory of visual grammar, advocating that images have a semantic structure similar to that of language. This theory makes it possible to systematically examine the visual meaning. At the same time, Mayer (2009) cuts in from the perspective of multimedia learning, pointing out that when visual information and verbal information are presented synchronously, learners can usually understand the content more effectively. This discovery actually reveals the fact that in the learning process, the application of images is much more than decoration. In the earlier period, most foreign studies regarded illustrations as auxiliary teaching materials, but since then, the research direction has gradually changed, and more studies have begun to pay attention to how images themselves participate in discourse construction and multimodal meaning generation.

In the domestic context, the research on images of English textbooks has also shown an obvious growth trend. The existing research mainly involves teaching application, image use strategy, cultural characterisation and high-order thinking ability cultivation. However, despite touching these levels, some inherent shortcomings have not been completely improved. Studies have noticed that some textbook images still rely too much on decorative presentation, lack close connection with text content, or are not effective in supporting understanding (Cheng & Cong, 2020). At the same time, many studies are still at the level of descriptive discussion, and limited attention is paid to the internal process of visual meaning generation (Yu, 2025). From a multimodal perspective, researchers have proposed an analytical framework to explain the interaction of different symbol resources in meaning generation (Zhang, 2008), which provides useful theoretical support for further analysis of images in English textbooks.

On this basis, this study applies the visual grammar theory to the image analysis of *New Standard English (Selective Compulsory Volume II)*, and adopts the multimodal discourse analysis method to reveal how the textbook images realise their communication function, and how visual resources participate in the construction of discourse meaning. At the same time, this article also hopes to add a new perspective to the image research of existing textbooks, and bring some practical inspiration to classroom teaching and textbook writing.

2. Theoretical Framework of Visual Grammar

2.1 Visual Grammar Theory

The visual grammar theory proposed by Kress and van Leeuwen points out that images are like language and have the ability to construct structured meaning. According to this theory, the generation of visual meaning can be explained from the three dimensions of reproduction, interaction and composition. Since this study focusses on the function of textbook images in English learning, the following analysis will take the above three dimensions as the main theoretical support.

2.1.1 Representational Meaning

The meaning of reproduction is related to the path of images presenting people, objects, behaviours and environmental elements, focusing on how human experience is visualised and characterised, and how the viewer can identify the situation with the help of images. In high school English textbooks, images are not a simple background of text. Sometimes they are also used to present thematic situations, classroom activities or cultural fields more intuitively, so as to lower the cognitive threshold of the content of the text and help students enter the understanding process more smoothly.

2.1.2 Interactional Meaning

The interactive meaning focusses on the construction of the relationship between the image and the viewer, which usually involves factors such as gaze direction, social distance, perspective selection and mood marking. With the help of visual choices such as shooting angle, lens distance, colour use and image quality clarity, the textbook image affects the learner's sense of participation and emotional

involvement. In this way, visual resources can play a differentiated communication function in classroom discourse.

2.1.3 Compositional Meaning

The meaning of composition focusses on the internal organisation of visual information, among which the information value, salient and framing framework are the key elements of this dimension. The arrangement of picture elements in terms of position layout, size contrast, colour intensity, and the connection or separation between visual components may guide the audience's attention flow and affect the interpretation of the overall meaning of the image.

2.2 *Realization of Communicative Functions of Textbook Images within the Visual Grammar Framework*

The communication function of textbook images cannot be fully explained by a single meaning dimension, but depends on the coordinated operation of the triple meaning of reproduction, interaction and composition. Different image types often focus on different dimensions, so the role of each dimension in different images is also different. Take the reproduction meaning as an example, the core is to present the participants, behaviours and environment in the picture. The meaning of interaction should focus on how the connection between the viewer and the image is established, and how the sense of participation and resonance is generated. As for the meaning of composition, the focus is on how the visual information is arranged and highlighted, such as why some elements are more prominent visually.

Different communication functions usually rely on different visual characteristics. When images are used to construct a communication context, participant behaviour, environmental background and viewing position are often more critical to shaping the context; in images presenting cultural information, visual elements such as cultural symbols, social behaviour patterns and spatial background are more likely to attract attention; images involving emotional attitudes rely more on pictures. The atmosphere, visual focus and emotional tension affect the learners' psychological feelings and responses. Cognitive-oriented images often rely on visual metaphors, symbolic characterisation and information layout to make the relatively abstract connotation easier to understand. Although the functions carried by various images are different, the final realisation of their functions still depends on the interaction of the three dimensions of meaning, rather than any isolated visual factor.

3. Multimodal Analysis of the Communicative Functions of Textbook Images within the Visual Grammar Framework

3.1 Construction of Communicative Contexts



Figure 3.1 Image from Unit 6 “Developing Ideas”



Figure 3.2 Image from Unit 5 “Understanding Ideas”



Figure 3.3 Image from Unit 1 “Memorable Moments”

In high school English textbooks, an important function of contextual images is to lead learners into the discourse situation and grasp the background of text development. Most of the images in this textbook play the function of “providing”, that is, conveying situational information through the presentation of characters, environment and action processes, helping students to initially establish a preliminary understanding of the background of the discourse. At the same time, some images are also biased towards the “solking” function. In addition to displaying the picture, try to use interactive resources such as gaze direction and social distance to bring the relationship between the viewer and the image closer, thus stimulating students’ thinking motivation and willingness to participate in the classroom. Figure 3.1 (p.91) and Figure 3.2 (p.68) can be used as typical examples. From the point of view of the meaning of reproduction, both images are narrative reproduction, which are mainly organised around the action process. Participants, together with vectors composed of body movement and direction, jointly construct activity scenes in a specific environment. Figure 3.1 shows the scene of astronauts walking on the barren surface of Mars and surrounded by a strange environment; Figure 3.2 shows a nature conservation worker carrying two dogs through the wild environment of Macquari Island. Although the two themes are different, both transform the text description into visible scenes through a

clear action process, making the abstract text easier to perceive. More importantly, both images provide visual construction for the discourse background with the combination of the participants' actions and environment.

In terms of interactive meaning, the two images do not establish eye contact between the participants and the audience. The audience focusses on being placed in the position of the bystander, which is the typical feature of "providing" images. At the same time, both of them use the long-view frame to place the characters in a relatively grand environment. Nevertheless, the interaction between the two is not exactly the same. The image of Mars appeared at the beginning of the article, creating a low-modal effect with the pink sky, which is different from the natural scene of the earth, which weakens the sense of reality and strengthens the distant and strange meaning of alien exploration. In contrast, the image of Macquary Island shows a stronger sense of reality. Details such as conservation workers and dog companions are highly consistent with the text description, conveying a high-modal effect, enhancing the credibility of environmental protection actions, and forming a more direct support for the second half of the text.

In terms of the meaning of composition, the environmental elements occupy a large visual space in both images, so that the situational information is more prominent. Therefore, the characters no longer have a visual dominant position, and the overall scene becomes the core focus. Take Figure 3.1 as an example, the eyes often fall on the exploration behaviour of astronauts first, and then slowly move up to the question in the upper right corner - "Is there life on Mars?", thus forming a visual advancement between exploration and questioning. Figure 3.2 is more balanced. The characters and natural scenery are evenly distributed in the picture. No side has an overwhelming advantage, and the whole presents a peaceful and coordinated visual effect. This kind of composition further strengthens the meaning of human and nature's companionship and protection, and at the same time makes the text description on the side more authentic and convincing in the eyes of the audience.

Figure 3.3 (p. 7) is different from the orientation of the first two images. Although the image also belongs to the context category, its main function is not simply to provide background information, but to enhance learners' sense of participation while building a campus communication field. Several students in the picture walk in the relaxed campus atmosphere, echoing the theme of "personal growth" and "different stages of life". Compared with the first two images, the picture has a closer relationship with the viewer. The forward-looking participants are facing the audience, and the close-up lens and the head-up perspective reduce the social distance. This kind of visual condition makes it easier for students to link the picture with their own campus experience, instead of staying in an alienated onlooker. In addition, the students' natural facial expressions and stretched body posture also create a friendly and harmonious atmosphere. Visual details such as bright corridors and fresh tones further strengthen the texture of the campus scene, making the image more effective in promoting emotional participation and classroom dialogue about "life's taste moments" and "life attitude".

3.2 Presentation of Cultural Information

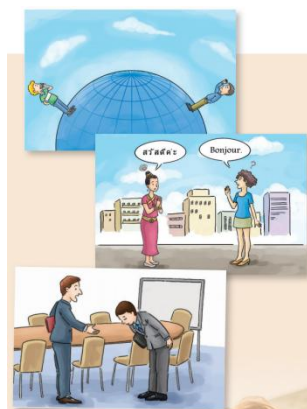


Figure 3.4 Image from Unit 4 “Starting Out”



Figure 3.5 Image from Unit 5 “Developing Ideas”

Cultural information is another dimension that the images of textbooks frequently touch. In this textbook, a large number of images are used to visualise cultural differences and cultural contexts. Some images focus on reproducing people’s greetings and communication behaviours in different cultural scenes, while others pay more attention to regional scenery and cultural space. Compared with simple textual interpretation, this kind of visual characterisation makes cultural content easier for learners to perceive and grasp. The following are two examples of images to illustrate how the function is implemented in the textbook.

Figure 3.4 (p. 49) shows greeting etiquette and social interaction norms from different cultural backgrounds in the form of illustrations. In terms of the meaning of reproduction, this group of images mainly adopts the narrative reproduction path. The participants’ gestures, body dynamics and communication behaviour together constitute a distinct action process. Different greeting expressions and body movements concretize the differences in cross-cultural communication, making the relatively abstract etiquette norms clear and tangible. Images do not rely on words to explain cultural differences, but convert them into visible communication scenes.

At the level of compositional meaning, the group of images adopts a weak framing separation, and several scenes are presented side by side, forming a relationship between juxtaposition and contrast. This arrangement guides the audience to compare cross-cultural communication behaviour. The image sequence has also produced changes in the visual focus: some scenes show a broader communication environment with a longer-range view, while others enhance the prominence of the participants and

draw attention to clothing, gestures and interactive details. In this way, cultural differences become more visually recognisable, and learners can more intuitively grasp the distinction between different etiquette practises.

Figure 3.5 (p. 74) focusses on the presentation of regional cultural information with the help of realistic photography. The image mainly constructs the natural scenery of Shennongjia, including mountains, snow, vegetation and surrounding environmental features. In terms of the meaning of reproduction, the image is more inclined to conceptual reproduction, because it does not highlight the dynamic behaviour process, but presents the landscape as a relatively stable and symbolic cultural space. The natural environment becomes the carrier of regional culture here, and visual elements such as mountains and forests act as symbolic attributes, indirectly conveying the connotation of local culture.

In the sense of composition, the natural scenery occupies the vast majority of the visual space and achieves high prominence in the picture. The traces of the characters have been greatly weakened, and the audience's eyes can be focussed on the environment itself. At the same time, the adoption of real-life photography techniques enhances the meaning of interaction and gives the image a stronger sense of mood and reality. Compared with illustrations, photographic reproduction makes the landscape more realistic and credible. In this way, learners may form a clearer understanding of the cultural connotation associated with the natural landscape of China.

3.3 Evocation of Emotional Attitudes

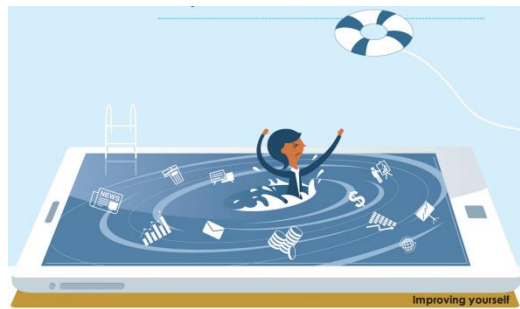


Figure 3.6 Image from Unit 2 “Understanding Ideas”



Figure 3.7 Image from Unit 4 “Starting Out”

From the perspective of emotional guidance, the images of this textbook can be roughly divided into two categories. One type mainly creates negative emotional experiences, using visual characteristics such as stress, anxiety or bondage to generate warning effects; the other type focusses more on positive emotional identity, trying to convey value concepts such as openness, friendliness, cooperation and connectivity through a warm visual atmosphere.

Figure 3.6 (p. 19) is a typical representative of the first category. The image uses metaphorical visual design to guide learners to reflect on the problem of social media dependence. From the observation of the reproduction meaning, the image constructs a narrative scene in which an individual is trapped in social media. The participant's raised arms and struggling body posture formed a clearly visible vector, visually showing the process of being involved and trying to get out. With this action reproduction, the abstract concept of "addiction" is transformed into a concrete and emotionally perceptible visual experience. As a result, learners are more prone to emotional reactions such as nervousness, anxiety and critical reflection.

The meaning of interaction is mainly strengthened through the visual atmosphere. The picture is dominated by cold tones, creating a stronger emotional distance and psychological suppression. This visual selection enhances the warning effect and makes the overall atmosphere particularly solemn and depressing. In terms of compositional meaning, the mobile phone screen occupies the vast majority of visual space and becomes the absolute dominant element. In comparison, the participants on the screen are small, while the vortex structure inside the device is very prominent. This kind of arrangement implies that the individual is surrounded and controlled by the medium. At the same time, the lifebuoy placed in the upper right corner of the picture is a symbolic attribute. In addition to balancing the composition, it also hints at the possibility of rescue or escape, thus linking the image more closely with the topic of "social media detoxification".

Figure 3.7 (p. 49) adopts a completely different emotional construction path. The image does not create tension or warning, but focusses on strengthening the positive emotional identity related to cross-cultural communication. In terms of reproduction, several young people from different cultural backgrounds circled the globe and gathered their eyes on it. The direction of their bodies and the direction of their eyes are pointed to the centre, forming a clear vector, creating a visual scene with joint participation and consistent focus. With this narrative reproduction, the abstract concept of "breaking boundaries" becomes more concrete and perceptible.

The meaning of composition also plays an important role in the image. The globe is in the centre, has the highest prominence in the visual structure, and naturally becomes the focus of the audience. At the same time, the circular distribution of the surrounding participants strengthens the sense of connectivity and collective belonging. The meaning of interaction can be further enhanced through warm tones and natural light and shadow, and the two together create a more friendly and harmonious atmosphere. Compared with the previous image, this picture has a stronger sense of friendliness and closeness.

Under this influence, learners may be able to generate positive emotional responses to cross-cultural communication and cooperation more naturally.

3.4 Facilitation of Cognitive Interpretation



Figure 3.8 Image from Unit 2 “Starting Out”



Figure 3.9 Image from Unit 3 “Presenting Ideas”

In order to support learners’ cognitive understanding, textbooks often use vivid visual characterisation to make abstract concepts, attitudes and values concrete and easier to perceive. As a result, the process of learning difficult concepts tends to be more intuitive, which helps to reduce the burden of understanding. This study found two main types of cognitive-oriented images: one is to explain the abstract meaning with the help of symbols that learners are already familiar with, and the other is to

show the process of change and development through the image sequence connected by the weak frame.

Figure 3.8 (p. 17) is a typical example of the first category. The image takes the clock, which is extremely common in life, as the core visual symbol of time, and through anthropomorphic design, a villain pushes the pointer forward, making the abstract concept of “making full use of time” visually easy to understand. At the level of reproducing meaning, the body posture of the villain and the direction of movement of the pointer together form a vector, constructing a process of suggesting force, pushing and active control. As a result, the image visually strengthens the connotation of “time should be actively managed rather than passively experienced”.

At the same time, the image is not intended to directly reproduce reality, but is a visual metaphor. The symbolic relationship between the villain and the clock enables the image to convey the meaning beyond the characterisation of the literal narrative. The meaning of interaction is enhanced through close-up scooping, shortening the interval between the viewer and the visual object, and making the time theme more close. In terms of the meaning of composition, the numbers and scales on the clock face occupy a prominent position in the vision, thus gaining high salience. This arrangement further strengthens the audience’s association of time pressure and efficiency. Through the superposition of narrative reproduction and symbolic meaning, the image effectively assists learners to grasp a relatively abstract concept.

Figure 3.9 (p. 48) characterises another type of cognitive-oriented image. The three pictures are connected to each other with a weak frame, forming a continuous visual sequence, presenting “the change of our learning style”. In terms of the meaning of reproduction, all three scenes depict the action process. As the subject of behaviour, the teacher’s gestures and body orientation constitute a vector pointing to the blackboard or teaching materials. These vectors create dynamic classroom situations and highlight the process of teaching activities.

Compared with Figure 3.8, the image sequence depends more on the meaning of the composition. The three pictures are arranged side by side, implying the advancement of the time dimension. When the line of sight moves along the sequence, the changes in the classroom environment become clearer. The transition from the traditional blackboard classroom to a more modern and technologically empowered learning space reflects the continuous evolution of educational resources and learning conditions. The change of significance between sequences also further emphasises this evolutionary process. As a result, the abstract concept of “change of learning mode” is converted into concrete visual experience, so that learners can more directly perceive the pulse of educational development through image sequences.

3.5 Evaluation of Functional Realization and Pedagogical Implications

Based on the above analysis, this section comprehensively evaluates the effectiveness of textbook images in different communication functions.

In general, the image that evokes emotional response reflects the most effective integration of three meta-functions. This kind of image usually integrates narrative reproduction with symbolic process, and constructs specific emotional scenes at the level of reproduction meaning. At the same time, with the help of the mood in the interactive meaning and the prominence in the composition meaning, the emotional atmosphere and value orientation are also enhanced together. Therefore, they effectively enhance learners' sense of participation and emotional resonance.

In contrast, although the images presenting cultural information have successfully constructed cultural content at the level of reproduction and composition, they are relatively limited at the level of interactive meaning. Most of these images play the function of "provision", mostly using long-term and low-interaction perspectives, mainly placing learners in the position of bystanders. Therefore, it is necessary to enhance the interaction between images and learners by introducing direct gaze, reducing social distance or adjusting the viewing angle, so as to enhance the communication effectiveness of cultural information.

In terms of images that construct the communication context, although some "resunit" images can effectively stimulate students' participation, the connection between a small number of images and text content is not strict enough. This shows that simply analysing the image structure within the framework of visual grammar is not enough to fully explain its function, and the inter-function relationship between graphics and text is also worth considering. In the future, the design of textbooks should focus on strengthening the connection between image and text semantics to avoid the disconnect between context construction and discourse meaning.

Images that help cognitive interpretation constitute a special highlight of this textbook. These images transform abstract concepts into concrete and tangible visual forms through documentary photography and conceptual illustrations, and achieve a balance between conceptual depth and cognitive reach, which is in line with the physical and mental development characteristics of learners at this stage. However, some images still have limitations at the level of reproduction meaning, such as the reproduction process is not clear enough or the symbolic relationship is unstable, which may cause the risk of misunderstanding. To this end, it is necessary to strengthen the correspondence between the image and the concept it wants to convey to improve the accuracy and effectiveness of cognitive support.

4. Conclusion

This study shows that the overall design of the textbook image of *New Standard English* (Selective Compulsory Volume II) is reasonable and can realise diversified communication functions in English learning. Through the interaction of reproduction, interaction and composition, these visual resources not only help to understand the text, but also participate in the processes of cultural cognition, emotional mobilisation and thinking development. At the level of context construction, many images help learners enter specific discourse situations more clearly and understand background information

with the presentation of behaviour, scenes and viewing relationships. At the level of cultural presentation, the visual reproduction of behaviour, dress style and regional landscape makes cultural information more intuitive and recognisable, which supports the cultivation of students' cross-cultural awareness to a certain extent.

The study also found that emotional guidance is a key function of textbook images. Some images form a warning effect through the visual characterisation of anxiety, stress or imbalance, while others strengthen positive emotional identity with the help of a warm and collaborative atmosphere. In addition, cognitive-oriented images rely on visual metaphors, symbolic elements and orderly image sequences to help learners understand abstract concepts in a more intuitive way. By transforming complex ideas into visual and concrete forms, these images reduce the difficulty of understanding and make the meaning construction more intuitive.

Overall, the research results show that textbook images play a significant role in the construction of multimodal discourse, which can support the development of the core literacy of English subjects. Therefore, visual resources are not an edissable embellishment in textbooks, but a meaningful component that participates in communication, interpretation and value guidance.

Analysis can also lead to some revelations. At the level of classroom teaching, teachers should pay more attention to the interpretation and communication value of textbook images, rather than just regarding them as auxiliary decoration. Under the framework of visual grammar, students can be guided to analyse images from different perspectives such as characterisation, interaction and visual organisation to enhance their multimodal literacy and interpretation ability. At the level of textbook writing, the future design needs to take into account the closer semantic relationship between images and text, and further optimise the interactive features such as orientation, social distance and visual atmosphere to improve learners' sense of participation and integration. In this way, textbook images can play a more effective role in supporting understanding, cultural transmission and value guidance.

References

- Cheng, X. T., & Cong, L. (2020). Design and use of visual resources in English textbooks. *Curriculum, Teaching Material and Method*, 40(8), 78-85.
- Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). London: Routledge.
- Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press.
- Ministry of Education. (2025). English Curriculum Standards for Senior High Schools (2017 Edition, Revised 2025). Beijing: People's Education Press.
- Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction*, 3(3), 227-238.
- Yu, S. M. (2025). Functional construction of textbook images from a visual narrative perspective. *Global Education*, 54(5), 31-43.

Zhang, D. L. (2009). A framework for multimodal discourse analysis. *Foreign Languages in China*, 6(1), 24-30.