English Language Teaching and Linguistics Studies
ISSN 2640-9836 (Print) ISSN 2640-9844 (Online)
Vol. 8, No. 5, 2026
www.scholink.org/ojs/index.php/eltls

Original Paper

A Comparative Study of Multimodal Discourse in Micro-Learning Videos via Intersemiotic Complementarity Theory

Binyan Ma[1] & Hanxu Liu[1]

[1] School of Foreign Languages and Cultures, Jilin University, China

Received: September 28, 2026 Accepted: October 19, 2026 Online Published: November 2, 2026

doi:10.22158/eltls.v8n5p135 URL: http://dx.doi.org/10.22158/eltls.v8n5p135

Abstract

With the increasing popularity of online educational videos, how to teach and disseminate these videos has become a challenge for many creators. But few of the current studies discussed the intersemiotic complementarity of micro-learning videos and how to promote their effectiveness through this angle. Based on the intersemiotic complementarity theory, this research explores how the collaborative interaction between micro-learning videos' image and text affects the effectiveness of teaching and dissemination.

This study compares and analyzes the image-text relationship of the two videos, Crash Course (CC) and Home Revise (HR) from ideational, interpersonal and compositional intersemiotic complementarity (IC). The study finds that: (1) For ideational IC, both videos use synonymy, collocation and hyponymy, and repetition is often combined with other relationship types to form a composite type. (2) For interpersonal IC, CC uses a lot of demand shots, and the text mood is diverse, so as to build a sense of interaction. (3) For compositional IC, the layout of CC is more flexible. In contrast, the layout of HR is relatively fixed and monotonous.

The results of this study show that good ideational IC can reduce the cognitive burden and promote the teaching effectiveness; highly interactive interpersonal IC can improve audience participation; good compositional IC can balance teaching and fun.

This study expands the application of intersemiotic complementarity in video analysis, and also provides inspirations for the design and evaluation of micro-learning videos.

Keywords

Intersemiotic Complementarity, Multimodal Discourse Analysis, Micro-Learning Videos, Teaching Effectiveness, Communication Effectiveness

1. Introduction

According to data released by the International Telecommunication Union (ITU) (2024), the number of global internet users had reached 5.5 billion by the end of 2024. In China, the number of internet users has reached 1.123 billion, with an internet penetration rate of 79.7% (China Internet Network Information Center, 2025). The use of social media has become an inseparable part of people's daily lives. Meanwhile learning online has become a trend. Tutorial videos on online video platforms that provide the necessary knowledge, offering an affordable and easily accessible learning path for many people. Micro-learning videos have become an important learning resource. However, many micro-learning videos lack effective communication and teaching effectiveness, and these aspects need to be improved.

Numerous studies have explored the communication strategies of knowledge-based educational videos. Unfortunately, these existing studies are still insufficient. In terms of the video categories, previous studies have focused on short videos themed on science popularization, health and medical knowledge (Li & Fu, 2024; Yuan & Qiu, 2026). While micro-learning videos, as an important online learning resource received few attentions on communication effectiveness. In terms of research methods, most of the previous studies on videos' communication strategies are based on video duration, shooting techniques, background music, presenting style and titles (Liu *et al*, 2025; Yu & Wu, 2024), which pays no attention on the text and image interaction of videos. In terms of theoretical framework, intersemiotic complementarity is often used in textbook research (Xiong *et al.*, 2024), while not many studies expand this theory into video studies.

Attempting to address this gap, this study examines two micro-learning videos different in teaching and communication effectiveness, and compares how they differ in text and image interaction. To achieve this purpose, this study applies intersemiotic complementarity theory, to compare the two micro-learning videos from the perspectives of ideational, interpersonal and compositional complementarity. This research aims to explain how micro-learning videos with better teaching and communication effectiveness design their text and image, and provide guidance for micro-learning video makers.

On the theoretical side, this study tries to encourage the application of intersemiotic complementarity theory to video research. On the practical side, by applying multimodal discourse analysis to visual presentation patterns, this study seeks to enhance the teaching effectiveness and communication effectiveness, thereby providing a reference for the production of micro-learning videos.

2. Literature Review

2.1 Studies on Communication and Teaching Effectiveness of Knowledge-Based Educational Videos

Currently, there is a substantial body of domestic research on the communication strategies of knowledge-based educational videos. In terms of thematic focus, science popularization videos constitute one of the core research areas. Many scholars have conducted studies on the operational

practices of science popularization journals on platforms such as Douyin and WeChat Video Channels (Liu *et al*., 2026; Yuan & Qiu, 2026; Zhan & Sun, 2025). Zhan & Sun (2025) use quantitative methods such as variance analysis and regression analysis, they investigate the factors influencing the communication effectiveness of science popularization videos, propose corresponding optimization strategies, and analyze the existing dilemmas in this field of communication. Health and medical knowledge videos are another research focus in this field. Some studies aim at health science popularization and explore optimization strategies through elements such as communication genres, interactive models, topic selection and text editing (Liu *et al*., 2024; Wang *et al*., 2025). Another study analyzes the innovative path of short videos of health science popularization books from the perspective of social support (Li & Fu, 2024). In addition, some studies have also discussed the communication strategy of national security (Sun *et al.*, 2025) and the popularization of emergency knowledge videos (Shen *et al.*, 2024).

Micro-learning videos are an important carrier of online knowledge dissemination. However, despite the wide range of these research topics, they are still relatively lacking in exploration. Micro-learning videos are different from short videos and long videos in many ways. In terms of duration, short videos usually last 15 to 60 seconds, medium videos last 1 to 30 minutes, and long videos last more than 30 minutes (Wang *et al*., 2024). In this framework, micro-learning video belongs to the category of medium video. In contrast, the short course video discussed earlier belongs to the short video format, while the course video of more than half an hour is classified as long videos. Although short course videos have a strong potential for online dissemination, they often convey fragmented knowledge that is easily forgotten. Although the content of long course videos is more systematic, it is often difficult to maintain the audience's attention due to the length, which often leads to low course completion rate and limited dissemination effect. Micro-learning videos provide an ideal compromise solution: compared with short videos, the knowledge it conveys is more structured and systematic; compared with long videos, its content is more refined and more conducive to wide dissemination. Therefore, micro-learning video is an important learning resource in the digital environment. Although a large number of micro-learning videos have emerged on major online platforms, the lack of research on relevant communication strategies leads to the lack of theoretical guidance in their design and production. This gap not only causes poor communication effect, but also limits its educational value.

In terms of content, most of the studies that explore the factors affecting the effectiveness of video communication focus on factors such as video duration, shooting techniques, presentation form (horizontal screen/vertical screen; animation/real person), content characteristics, title tags, commentary style and account attributes (Liu *et al*., 2025; Zhan & Sun, 2025). However, there are relatively few studies carried out from the perspective of video graphic presentation mode and design. Although some scholars have implemented multimodal content analysis for high-quality popular science videos, most of these studies only focus on single modality, such as visual modality (Xu & Liu, 2023), or regard different modalities as independent entities for separate discussion (Chen & Zhou,

2023), thus ignoring the interaction between different modalities in the same video and collaboration, that is, intersemiotic complementarity. This includes, for example, presenting images to further explain the content of the text, or annotating images with text to explain, so as to improve the efficiency of the audience's understanding. There is no existing research specifically for the complementarity analysis of micro-learning video.

2.2 Studies on the Applications of Intersemiotic Complementarity Theory

Currently, most of the application of intersemiotic complementarity theory is in textbook research (Xiong *et al*., 2024), focusing on static image-text materials. There are also studies that apply intersemiotic complementarity theory to the research of dynamic images, such as city promotional videos (Jing, 2025), tourism advertisements (Liang & Jiang, 2015). However, there is a research gap in the combination of pedagogy and video research, and there is almost no intersemiotic research on educational videos.

Nowadays, international research on video communication strategies mainly focuses on the content published by political party social media accounts. Scholars regard social media as a political tool to disseminate specific political ideologies through the Internet. Some researchers have studied the dissemination strategies of such videos to figure out the impact of social media on democracy (Di Nubila et al., 2023). In addition, there is a study targeting short videos released by commercial brands, but their investigation of communication strategies is limited to aspects such as titles, music, and subtitle presentation (Yu & Wu, 2024), without addressing the multimodal construction of the videos themselves. In contrast, micro-learning videos, which serve the functions of knowledge transmission and teaching, have not yet entered the research purview of communication strategies, leaving a research gap. Since the pandemic, studies on knowledge-based educational videos have emerged, and some researchers have gradually come to recognize that the teaching effectiveness of educational videos is not always satisfactory. However, their study mostly focuses on factors such as online class fatigue, learning engagement, and course satisfaction (Cho & Im, 2024), concentrating on the outcomes of teaching effectiveness rather than investigating the image-text presentation patterns of the videos themselves. They also do not address the multimodal construction of the course, and thus fail to provide specific guidance for optimizing the communication and teaching effectiveness of micro-learning videos.

There are not many relevant studies on the application of the intersemiotic complementarity theory abroad. Existing studies which mainly focus on pedagogy, apply intersemiotic complementarity theory to study middle school teachers' behaviors in the classroom (Nakakuwa & Jawahar, 2020), or examine the image-text complementarity in English textbooks (Chang & Tsai, 2022). However, relevant research on micro-learning videos still scarce.

Based on the above conclusions, this paper takes micro-learning videos as the research object and employ the multimodal discourse analysis framework of intersemiotic complementarity theory to analyze the complementarity between the image modality and text modality in micro-learning videos.

Intersemiotic complementarity theory is grounded in systemic functional linguistics; therefore, this study proceeds from three aspects: the ideational features, interpersonal features, and compositional features of multimodal discourse, exploring the image-text interaction patterns in micro-learning videos, with the aim of addressing how to improve the communication and teaching effectiveness of micro-learning videos through image-text design.

3. Theoretical Framework

3.1 Multimodal Discourse

Multimodal discourse refers to the communication phenomenon carried out through a variety of senses such as hearing, vision, touch, etc., using diversified means and semiotic resources such as language, image, sound and action (Zhang, 2009), and these multiple semiotic resources construct meaning simultaneously. Unlike the traditional discourse analysis, which focuses primarily on language, multimodal discourse analysis concentrates on how different modes contribute meaning in the overall message.

Systemic functional linguistics provides an applicable theoretical framework for multimodal discourse analysis. According to systemic functional linguistics, language embodies three meta-functions in communication—ideational function, interpersonal function and textual function (Halliday, 1994). From a multimodal perspective, semiotic systems other than language also realize the same three meta-functions (Zhang, 2009). For instance, the theory of visual grammar proposed by Kress and van Leeuwen (Kress & van Leeuwen, 1996) extends the concept of meta-functions to images and can be applied to the analysis of the visual modality in multimodal discourse. For instance, an image can represent objects and events, establish relationships through gaze, and structure information through composition. Therefore, the analysis of multimodal discourse can also be conducted from three aspects: ideational, interpersonal, and textual.

In the context of micro-learning videos, multiple modes operate simultaneously. There is spoken narration providing explanations, images and animations illustrating abstract concepts, and background sound enhancing involvement. These modes do not work in isolation; rather, they interact and cohere to support the viewers' comprehension and engagement. Thus, multimodal discourse analysis is suitable to be applied into micro-learning video studies.

3.2 Intersemiotic Complementarity

Intersemiotic complementarity refers to the coordination and complementary relationship between text and visual modal in multimodal discourse. Barthes (1977) concluded the relation between text and image as that all images are polysemous and they imply a "floating chain" of signifieds and text help to fix this "floating chain". Thus, text and image, as resources for the expression of meaning, jointly construct meaning and present coherent multimodal discourse to readers (Yang, 2012). The complementary analysis of multimodal discourse is also based on Halliday's systemic functional linguistics—focusing on ideational function, interpersonal function and textual function. Based on this,

Royce (2007) proposed ideational intersemiotic complementarity (ideational IC), interpersonal intersemiotic complementarity (interpersonal IC), and compositional intersemiotic complementarity (compositional IC).

Ideational IC requires first observing the visual message elements in the image, and then focusing on analyzing the relationships between these visual message elements and the textual information, including relations such as hyponymy, synonymy, antonymy, meronymy, repetition, and collocation. Specifically, hyponymy, synonymy, antonymy, and meronymy are understood here in the same way as they are in semantics. Repetition will occur, when the words or phrases said in the narration appear on the screen at the same time. Collocation means that text and image elements appear frequently at the same time in typical contexts, but they do not belong to any of the above five relationships.

Interpersonal IC focuses on connecting readers/audiences with texts through moods (offer, demand, statement, question). For the analysis of image, Kress and van Leeuwen's (1996) framework of interactive meaning was adopted, which examines images from three dimensions: gaze, social distance, and angle. In intersemiotic complementarity studies, these dimensions are called contact, social distance and attitude respectively (Zhang, 2013). The contact indicates whether the image participant directly stares at the audience (demand) or not (offer). The social distance conveyed through the shooting distance (close-up, medium and long shots) reflects the social relationship between the image and the audience. The attitude expressed through the shooting angle (front/oblique, high/low) conveys the perspective of the image to the audience.

The compositional IC involves the layout between different symbol systems and the way the combination is coherent, including information value, salience and reading paths within the framework (Royce, 2007). To be specific, information value refers to how different information is laid out. Different positions on the screen represent different levels of importance of the information. Salience concerns how information attracts viewers' attention through elements such as fonts, color, size, etc. Reading path is about the order in which viewers watch, such as from left to right or from center to edge, etc.

4. Methodology

4.1 Research Questions

The main objective of this paper is to analyze the image-text modality complementarity patterns of micro-learning videos through a comparative study, and to summarize how image-text design can enhance the communication effectiveness and teaching effectiveness of micro-learning videos, thereby providing references for the design and production of other micro-learning videos. Therefore, this study will investigate the following research questions:

(1) What are the similarities and differences between the *Crash Course* and the *Home Revise* in terms of ideational IC, interpersonal IC and compositional IC?

(2) How does image-text intersemiotic complementarity influence the communication effectiveness

and teaching effectiveness of micro-learning videos?

4.2 Research Object Selection

In order to ensure the effectiveness of the comparison, this study sets four criteria for the selection of two micro-learning video samples. First of all, the video must revolve around the same topic to control the content differences. In this study, "Introduction to Biology" is chosen as the topic. Secondly, their duration should be comparable, and the ideal range is between 13 and 16 minutes. Third, the release date of the video should be within about a year to reduce the impact of platform algorithms and technological evolution. Fourth, the two videos need to exhibit pronounced differences in communication and teaching effects, with one enjoying notably wider recognition and more positive reception than the other.

Based on these four criteria, this study selected two micro-learning videos on "Introduction to Biology" for comparison. The two videos are highly comparable: they share the same topic, their durations fall within the ideal range, they were posted within about a year of each other, and they show pronounced differences in communication and teaching effectiveness. Such comparability allows the videos' divergent performance to be examined in relation to their image-text design rather than external factors. The two video files used for coding were supplied to the researcher as local copies by an overseas contact.

The video titled "Introduction to Biology: Crash Course Biology" made by *Crash Course* (hereafter abbreviated as *CC*) has a duration of 13 minutes and has attracted considerably wider attention and more favorable reception than the other selected video, and is therefore treated as the high-performing video. There is a lot of information packed in the video, plus animated visuals popping up here and there. The presentations are lively and pretty engaging. It is a good example of how online educational video can be designed successfully. These micro-learning videos mix the instructor's spoken explanations with real case demonstrations. In the case demo part, pictures and text often work together. That makes it really suitable for analyzing intersemiotic complementarity. The other video, titled "Introduction to Biology: What is Biology" made by *Home Revise* (hereafter abbreviated as *HR*), also satisfies the above criteria. It has a duration of 16 minutes and has received limited attention and less favorable reception than *CC*, and is therefore treated as the low-performing video. Despite its earlier release, this limited reception conforms to the characteristics of poor communication and teaching effectiveness.

4.3 Research Procedures

This study uses the screenshot function of the computer to complete the screenshot operation according to the change of the shots. Here, shots conversion is the main basis for screenshots; if a single shot lasts for more than 10 seconds, the picture will be captured every 10 seconds. After organization, a total of 117 shots is captured from the *CC*, and 149 shots are captured from the *HR*.

Next, according to these screenshots, extract the text content that appears at the same time as each picture. After that, sort out the screenshots and the corresponding text content into the Excel worksheet

for easy viewing. Six separate worksheets are made. These worksheets serve to analyze ideational IC, interpersonal IC, and compositional IC respectively.

At the ideational level, this study mainly explores the semantic connection between image modality and text modality from the semantic perspective. According to the ideational IC classification method proposed by Royce (2007), the method covers many aspects such as repetition, collocation, meronymy, hyponymy, synonymy and antonymy. The image-text content is coded and specific relationship types are annotated. Finally, using Excel software, the frequency and proportion of various ideational IC types in the two videos are counted respectively, so as to compare and analyze the similarities and differences between the two videos in this regard.

At the interpersonal level, this study draws on the interactional meaning explained by Kress and van Leeuwen in the theory of visual grammar (Kress & van Leeuwen, 1996), paying special attention to contact, social distance and attitude. Contact refers to the expression of different speech functions through the gaze or gestures of the people in the image, including offer (no contact), demand (contact). The notion of social distance draws on the photographic concept of "shot distance". Three distinct categories of shots can be identified: close-up, medium shot, and long shot. In essence, these shot types function to convey the social connection between the image and the viewer. When it comes to camera attitude, this factor is closely related to the camera angle, frontal angle and oblique angle, along with high angle and low angle (Zhang, 2013). From the perspective of the mood system, the analysis of the text was carried out; this system includes declarative sentences, imperative sentences, and interrogative sentences, each of which fulfills the different functions. Finally, an analysis is conducted to determine whether the interpersonal features of the two modalities are consistent and formed complementarity.

At the compositional level, the interaction and coordination between the image and text modalities are examined from two perspectives: layout and salience. Since reading path is closely related to salience, it is seen as a part of salience analysis. Thus, reading path is not explained separately here. In this way, this study compares the two videos in two respects: the overall layout features of shots where images and text appear at the same shot, and the ways in which viewers' attention are directed through salience design. The arrangement of images and text is analyzed to determine which parts of information are placed in more salient positions and which parts are placed in secondary positions.

For illustrative purposes, line-drawn schematic illustrations presented in this paper were produced with generative-AI assistance. These illustrations are reconstructed based on coded video screenshots observations extracted from source educational videos, rather than direct reproductions or adaptations of original video screenshots.

5. Results

5.1 Multimodal Comparative Analysis of Micro-Learning Videos with Intersemiotic Complementarity Theory

After data collection and analysis, it is found that the ideational IC in the two videos can be categorized

into two types: the single type and the composite type. Single type refers to that there is only one ideational IC type in a shot. Composite type indicates that there are more than one ideational IC types in a shot, such as repetition + synonymy.

The *CC* exhibited nine types of ideational IC, among which synonymy accounted for the highest proportion, followed by unrelated, collocation, hyponymy, collocation + repetition, meronymy, repetition, hyponymy + repetition and synonymy + repetition. No instances of antonymy are found.

The *HR* also exhibits nine types of ideational IC, among which hyponymy accounted for the highest proportion, followed by collocation, synonymy, repetition, collocation + repetition, hyponymy + repetition, and synonymy + repetition. No instances of meronymy or antonymy are found. In addition, two unrelated shots and 16 shots categorized as "repetition with irrelevant background" are identified.

In both videos, the relationship of repetition frequently appeared in combination with other relationship types, primarily manifested as subtitle elements within the images.

The results are presented in Table 1 and Table 2 below.

Table 1. Frequency and Proportion of Ideational IC Types in *CC*

Ideational IC Types	Frequency	Proportion
Synonymy	43	36.7%
Collocation	15	12.8%
Hyponymy	12	10.2%
Meronymy	5	4.2%
Repetition	4	3.4%
Unrelated	27	23.0%
Collocation + Repetition	7	5.9%
Hyponymy + Repetition	3	2.5%
Synonymy + Repetition	1	0.8%

Note. The 27 "unrelated" shots in *CC* are scenes where the host speaking directly to the camera without any visual element corresponding to the text, thus not constituting a design flaw but rather a distinct presentation pattern.

Table 2. Frequency and Proportion of Ideational IC Types in *HR*

Ideational IC Types	Frequency	Proportion
Hyponymy	31	20.8%
Collocation	22	14.7%
Synonymy	21	14.0%
Repetition	19	12.7%
Unrelated	2	1.3%

Repetition with Irrelevant Background	16	10.7%
Collocation + Repetition	14	9.3%
Hyponymy + Repetition	14	9.3%
Synonymy + Repetition	10	6.7%

Note. The "Repetition with Irrelevant Background" category in *HR* refers to shots displaying section titles where the text on screen repeats the narration, but the background image has no semantic relation to the text.

This research shows that although there are significant differences between the two videos, they also have some similarities. Compared with *HR*, *CC* are more diverse in terms of ideational IC types, and are unevenly distributed, biased towards specific types—most notably, the frequency of synonyms is significantly higher than that of collocations and hyponymy. In contrast, *HR* is more balanced in the distribution of ideational IC types, and the frequency difference between each type is relatively small. Both videos use repetition techniques, but the frequency of repetition in *CC* is much lower than that in *HR*. In addition, *CC*'s "unrelated" shots are direct-to-camera host segments rather than mismatched image-text pairs, while the video of *HR* contains a small number of such irrelevant shots.

The commonality between the two videos lies in the fact that, excluding unrelated shots, the three most frequently used ideational IC types in both videos are synonymy, hyponymy, and collocation. In addition, repetition is often appearing with synonymy, hyponymy, or collocation within the same shot.

In terms of visual form, *CC* contains 59 demand (contact) shots and 58 offer (no contact) shots, among which there are significantly more medium shots and frontal angles than other types. In contrast, the number of offer (no contact) shots in *HR* is much more than that of demand (contact) shots, and medium shots, front-angle shots and oblique-angle shots are common. There is a significant difference in the use of shots in these two videos: *CC* is more inclined to balance the use of demand and offer shots, while *HR* prefers to use offer shots. In terms of social distance, there is almost no difference between these two videos. They mainly use medium shots and rarely use long shots. In terms of attitude, the two videos also exhibit differences: most shots in *CC* are presented from a frontal angle, whereas *HR*, in addition to frontal angles, also features a substantial number of shots from oblique angles.

The results are shown in Table 3 below.

Table 3. Visual Interpersonal Features in the Two Videos

Video	Contact	No Contact	Close-Up	Medium Shot	Long Shot	Frontal Angle	Oblique Angle	High Angle	Low Angle
CC	59	58	24	87	6	90	18	5	4
HR	14	135	38	104	7	66	78	5	0

After analyzing the text corresponding to the camera, it is found that the text mood in both videos is mainly declarative. However, *CC* shows greater diversity in mood types. In addition to declarative sentences, they also integrate interrogative sentences and imperative sentences. In contrast, *HR* only uses a statement throughout the whole process and only plays the single function of providing information.

The results of comparing text modality are presented in Table 4 below.

Table 4. Comparison of Mood Types in the Two Videos

Video	Declarative	Interrogative	Imperative
CC	67	6	3
HR	102	0	0

When analyzing data at the compositional level, this study proceeded from two perspectives: layout and salience. Since the reading path of a video frame is typically associated with salience—moving from areas of high salience to low salience—reading path was incorporated into the salience analysis.

The *CC*'s creators do not add bottom subtitles. Therefore, from the 117 captured shots, this study picked out 38 shots that contained both textual and visual content simultaneously. In these shots, text is mostly positioned at the periphery of the frame, playing a supporting role in conveying meaning alongside the images, with relatively low salience. In only three shots does text have higher salience than images, with the text positioned at the center of the frame. In general, the video adopts a variety of text colors, different font sizes and flexible arrangement.

The creator of the *HR* added white subtitles with a black background at the bottom of the video for narration, and added blue subtitles with a green background at the top to mark the chapter title. Some shots also have subtitles in the upper left corner, just below the chapter title. These text elements are pretty noticeable and take up a good amount of space. They catch the viewer's eye pretty quickly. The layout follows common reading habits—top to bottom, left to right, moving from the title into the details of the content. In the video, in addition to the scenes of other text and images, there are also two layouts: the first is that the text is located in the center of the image, the image is the background, and the text is more prominent than the image; second, text positioned below images to introduce the image content, where text salience is lower than images. In the set of screenshots examined, the second type of image-text co-occurrence occurs considerably more often than the first. Across the entire *HR* video, text styling remains uniform: white characters appear on a black background, with very little change in font size.

5.2 Influence of Intersemiotic Complementarity on Teaching and Communication Effectiveness

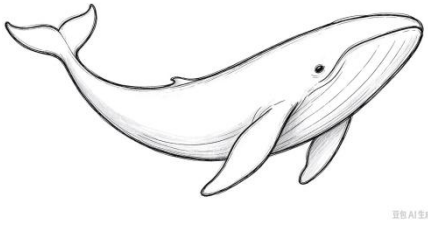
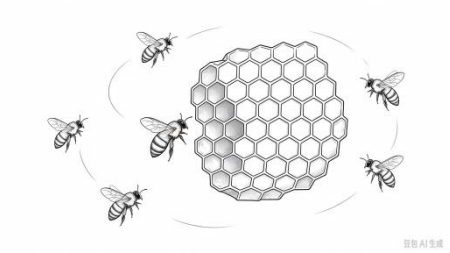
Ideational intersemiotic complementarity has an impact on how well viewers can understand the content. When it is used properly, ideational IC can obviously reduce the cognitive burden on viewers,

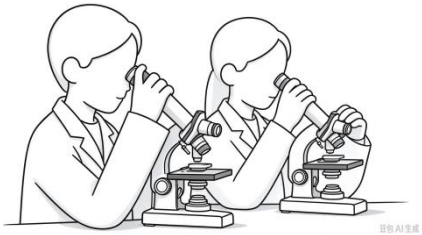
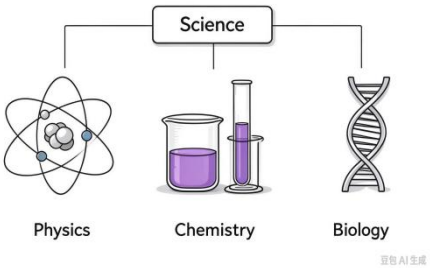
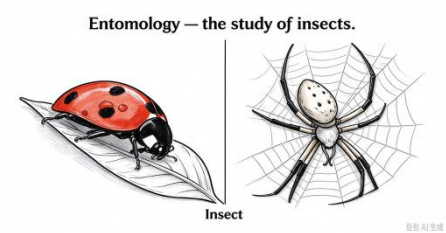
and in turn improve teaching effectiveness. The results of this study show that both videos have acceptable levels of ideational IC, though each focuses on different aspects. In the *CC*, synonymy makes up the biggest part, accounting for 36.7%—that's more than any other type. This method helps turn abstract ideas into something more concrete.

To give an example (see Table 5-1), when the narration says "ferns", the corresponding image directly shows fern plants, so viewers can grasp the concept without having to imagine it with much effort. Because synonymy is used extensively, viewers can obtain information straight from the visuals, which bypasses the step of turning textual descriptions into mental images and thus cuts down cognitive load.

The *HR* makes more use of hyponymy. It takes up 20.8% of all the strategies used. This kind of hyponymy is pretty helpful for building knowledge structures. Here is an example (see Table 5-2). The narration says "organisms survive and reproduce." In the visuals, they pick whales and bees as specific cases. Whales show survival—they are an ancient species. Bees show reproduction, because they are really good at breeding. Using hyponymy this way helps viewers understand how abstract ideas turn into concrete forms. Theoretical knowledge gets grounded in things can be actually seen.

Table 5. Examples of Ideational Intersemiotic Complementarity

No.	Text	Image	Ideational IC
1	Beetles, ferns, algae, yourself, your neighbor.		Synonymy
2	These adaptations are traits that help organisms survive and reproduce.	 	Hyponymy

3	We study biology for a whole bunch of different reasons.		Collocation
4	There are three main branches of science, namely, Physics Chemistry and Biology.		Repetition + Collocation
5	Entomology — the study of insects.		Hyponymy + Repetition

Collocation shows up a lot in both videos. In *CC* it makes up 12.8% of the strategies, and in *HR* it is 14.7%. Take Table 5-3 as an example. The phrase “study biology” comes with a picture of kids looking through a microscope. That kind of pairing turns abstract research activities into scenes people can recognize more easily. Viewers can use what they already know while learning new things.

Repetition is more common in the *HR*. It often appears with other relationship types—like repetition + collocation, or repetition + hyponymy. Table 5-4 gives an example. The words “physics”, “biology”, and “chemistry” are each repeated. And each repeated word gets paired with its own matching image. Later, when the video introduces a new topic, entomology—the study of insects (see Table 5-5), the word “insects” is repeated again, and a picture showing a beetle and a spider is used to help explain the term. Repeating important words from the text and linking them to relevant images can stress key points and leave a stronger impression on viewers. That said, the 10.7% of shots labeled as “repetition with irrelevant background” in the *HR* are worth noting—in these cases, the title text repeats what the narration says, but the background image has no semantic connection to the content. Such a design does not make full use of what image-text complementarity has to offer.

Each ideational IC strategy plays a different part in helping teaching work better. Synonymy turns text into images that match. That way, learners pick up new concepts more directly. Hyponymy uses

concrete examples to stand for abstract categories. It helps viewers build knowledge structures that are well organized. Collocation brings up familiar, related scenes to activate prior knowledge. This makes unfamiliar information easier to get. Repetition helps lock things into memory by bringing back key points more than once. But for it to really work, the repeated images need to be meaningful. No single strategy is absolutely better or worse than the others; instead, video designers should pick the dominant strategies that fit the nature of the knowledge content they are dealing with.

When it comes to interpersonal IC, both videos manage to achieve consistency and complementarity across the image and text interpersonal systems. That is, the *CC* shows more variety in its interpersonal IC types and comes across as more interactive. Looking at visual modality, the *CC* has a nearly equal number of demand shots and offer shots, with counts of 59 and 58. By contrast, the *HR* contains far more offer shots than demand shots, with counts of 135 and 14 respectively. The host in the *CC* frequently looks straight into the camera (see Figure 1), which creates eye contact with viewers. Such eye contact affords viewers a heightened sense of involvement; they experience the feeling of being addressed directly, rather than being relegated to the periphery as passive onlookers. In the *HR*, by contrast, no host ever appears on screen. Consequently, viewers are consistently positioned as observers, which creates a greater degree of psychological distance. With respect to social distance and attitude, both videos predominantly employ medium shots along with frontal and oblique camera angles. As illustrated in Figure 2, this combination of shooting distance and angle proves suitable for the objective presentation of educational content. When paired with declarative textual material, both videos thus realize interpersonal IC on this particular dimension.

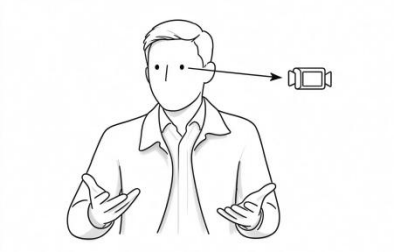
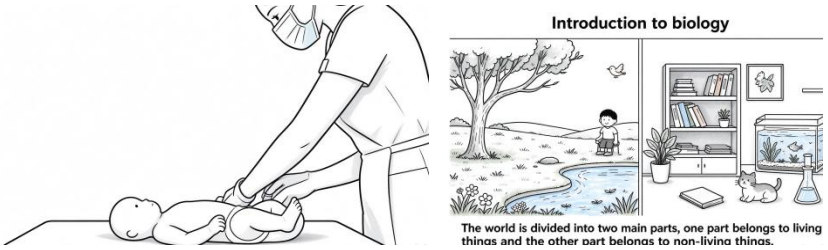


Figure 1. Demand Shot (with Contact) in *CC*

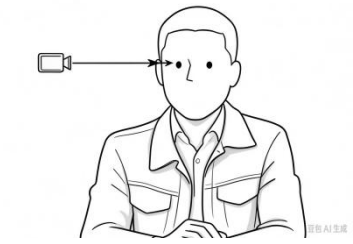


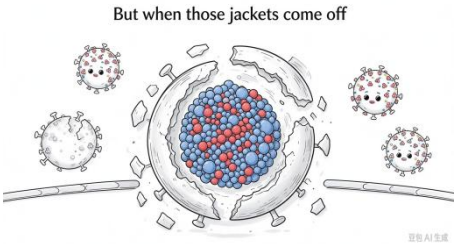
(a)*CC* (b)*HR*

Figure 2. Medium Shot and Oblique Angle Shot in *CC* and *HR*

From the perspective of textual modality, both videos primarily use declarative sentences to convey objective knowledge. However, the *CC* occasionally employs interrogative and imperative sentences, such as (see Table 6) “If we didn’t, how else would we know that there are living microbes up in those clouds that affect the weather?” or “Let’s head to the Thought Bubble…” or “Watch out.”. Through this shift in mood, the video directly engages viewers and encourages them to think actively instead of just taking in information without reflection. This finding indicates that using interrogative sentences in educational materials can draw viewers’ attention and increase their level of engagement. Meanwhile, questions posed by the host encouraged viewers to comment or discuss under the video, which in turn helps improve both teaching and communication outcomes.

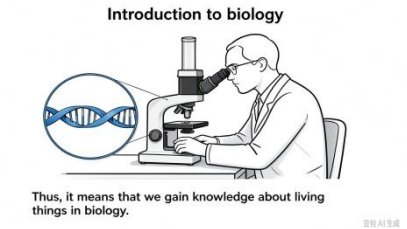
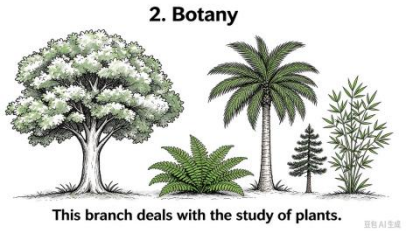
Table 6. Examples on Different Mood in *CC*

No.	Text	Image
1	Debating what does and doesn’t count as life might seem kind of up in the clouds—but hey, if we didn’t, how else would we know that there are living microbes up in those clouds that affect the weather? (interrogative)	
2	Let’s head to the Thought Bubble… (imperative)	

3	But when those jackets come off, watch out. (imperative) Viruses seem wildly alive.	

In contrast, the *HR* uses declarative sentences exclusively. The tone remains uniform throughout, with the sole purpose of conveying factual information. The linguistic style resembles that of written rather than spoken discourse, and as a result, interactivity remains weak (see Table 7).

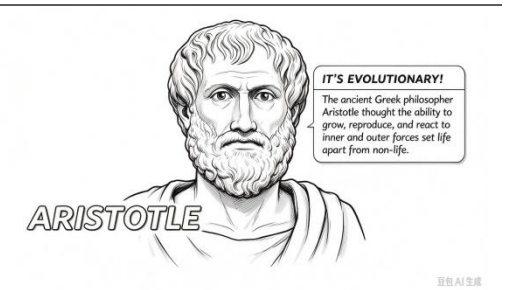
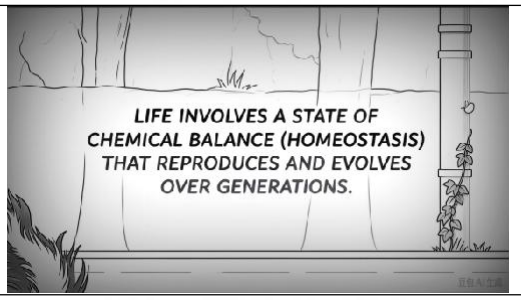
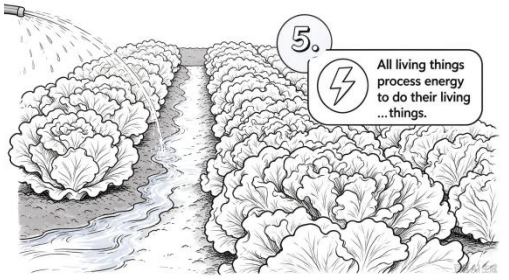
Table 7. Examples on Mood in *HR*

No.	Text	Image
1	Thus, it means that we gain knowledge about living things in biology. (declarative)	
2	This branch deals with the study of Plants. (declarative)	

In short, the *CC* mimics a real classroom setting through eye contact and changes in mood. Viewers therefore feel more involved and engaged. This way of interacting with the audience proves more effective at grabbing attention, boosting viewership and satisfaction, and ultimately delivering better communication results. As for the *HR*, its interpersonal style stays neutral and factual. But on online video platforms, people can click away anytime, so holding onto their attention for long periods is more of a challenge.

Regarding compositional IC, the two videos' layout choices differ a lot. And those different choices each shape teaching results and communication success in their own ways.

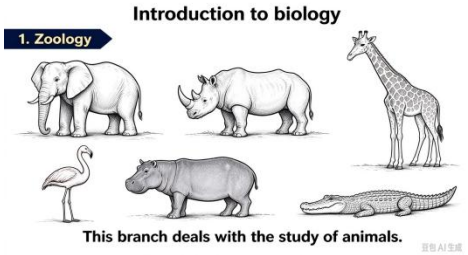
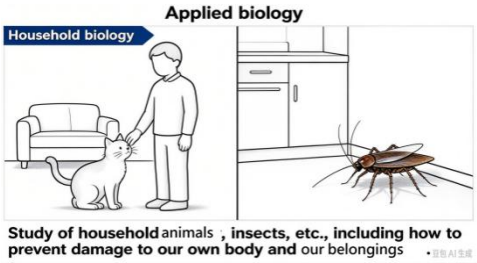
Table 8. Examples on Compositional Intersemiotic Complementarity in *CC*

No.	Text	Image
1	Like, the ancient Greek philosopher Aristotle thought the ability to grow, reproduce and react to inner and outer forces set life apart from non-life.	
2	Modern-day biologists tend to agree that life involves a state of chemical balance that reproduces and evolves over generations.	
3	So, this episode of Crash Course is brought to you in part by a lettuce plant, an olive tree, and, oddly enough, the milk of a sheep.	

In terms of layout design, *CC* is characterized by dynamic and flexible layout. The text on the screen is in a variety of colors and font sizes, and its position will be adjusted as needed, but it is usually still close to the edge of the screen. In this design, visual elements carry the main information, while words only play an auxiliary role. Because the layout of each picture will be adjusted according to the display content, it can maintain the audience's attention for a long time and effectively avoid visual fatigue. One example (see Table 8-1) comes from the section on Aristotle: his portrait sits at the center of the screen, his name appears in a larger font, and his theory of evolution is written in a smaller font and placed in the corner. That is because the aim of this shot is to cite Aristotle's division of life and non-life, instead of his evolutionary theory. While in another shot (see table 8-2), text is in the center of the image, and the road sign and telegraph pole are on the two sides with dark colors. In this example, the text is more important, the background is accessory. In several shots, identical text is placed in the corner of the screen, signifying they belong to the same topic and thus strengthening their cohesion (see table 8-3). However, viewers need to adapt to different screen layouts, which to some extent increases the cognitive cost of information localization.

The *HR* adopts a fixed and unified layout system: black-background white subtitles at the bottom transcribe the narration, green-background blue subtitles at the top mark chapter titles, and occasional subsection titles appear in the upper left corner (see Table 9-1). When text and images appear together in a shot, the text usually sits underneath the images (see Table 9-2) and stands out less than the images do. Through this layout, the audience can quickly adapt to the clear reading mode. They will know where to look to read the title, where the picture is located, and where the subtitles are placed. This will reduce the cognitive load required by the audience to find information. However, there are also potential disadvantages to maintaining the same layout in all shots. As time goes by, the audience may develop visual fatigue. After watching for a long time, their attention is likely to gradually decline.

Table 9. Examples on Compositional Intersemiotic Complementarity in *HR*

No.	Text	Image
1	This branch deals with the study of animals.	
2	Study of household animals, insects etc., including how to prevent damage to our own body and to our belongings.	

In *CC*, through techniques such as color contrast and change of font size, important information is highlighted and the focus of readers' attention is clarified. On the other hand, the layout of *HR* remains unified, and all information is presented in the same form. Although static compositions allow the audience to quickly grasp the overall picture, they also have the risk of being monotonous. On the contrary, although dynamic compositions may bring a greater cognitive burden to the audience, they can maintain visual appeal. *CC* achieves an effective balance between practicality and attractiveness. Although the composition of each image is different, the subject always stands out visually. In contrast, the design style of *HR* is clear and predictable. However, its visual monotony may make it difficult for the audience to keep their attention when watching for a long time.

So, when designing educational videos, one has to weigh "usefulness" against "interest". The goal is to find a middle ground—where layout predictability meets visual variety—that fits the particular content.

6. Conclusion

6.1 Findings

This study has the following findings. Firstly, for the ideational IC, most of the shots in both videos well match images with text. However, the quality of each video's design is not the same, and that affects how well these complementarity strategies work in practice. According to the research, ideational IC types fall into two categories: single and composite. Take the single types first. Synonymy, hyponymy, and collocation show up a lot in both educational videos. They ease the mental workload for viewers and make teaching more effective. With hyponymy, general categories get illustrated through specific examples, and that helps viewers put together their own knowledge structures. Collocation puts abstract content into a context, calling on what viewers already know. Both videos use composite ideational IC types, such as repetition + collocation. When key information comes up, repeating the text while also using image-text complementarity can increase the frequency of main points, and add extra semantic support through visuals. It is worth noting that, in the *HR*, 10.7% of the shots fall under the label "repetition with irrelevant background." Such images do not help with meaning construction; instead, they may pull viewers' attention away and add unnecessary mental load.

Secondly, the interpersonal IC and the compositional IC are important in the effectiveness of video communication, and *CC* is obviously superior to *HR* at both levels. Thanks to the host's direct look at the camera and the combination of interrogative sentences, imperative sentences and declarative sentences, *CC* reproduces the "face-to-face" educational environment and effectively shortens the psychological distance. On the other hand, *HR* lacks this engagement. Viewers are just bystanders, and it is difficult for them to feel that they are one of them. In terms of layout, *CC* adopts a dynamic and flexible design. By using different colors, font sizes and typesetting, it not only maintains aesthetic attractiveness, but also prevents attention from drifting away. Although the static layout in *HR* reduces the cognitive load of information retrieval, its uniform appearance will cause visual fatigue for viewers, thus reducing their continuous attention. The design differences directly explain the different reception of the two videos, and show that interpersonal interaction and visual appeal are crucial on an online video platform where viewers can exit the video at any time.

Thirdly, the three dimensions of intersemiotic complementarity are not isolated, but jointly determine the effect of the video. The ideational dimension realizes effective knowledge transfer, the interpersonal dimension ensures the continuation of the audience's interest, and the compositional dimension optimizes the presentation of visual information. Creators of micro-learning videos must take these

three dimensions into account.

6.2 Implications

Drawing on the findings above, this study offers several suggestions for how micro-learning videos might be designed and produced.

At the ideational level. Designers need to pick the main complementarity strategies based on what kind of knowledge they are dealing with. When the goal is to explain new ideas or abstract processes, synonymy is a good first choice: present the content through matching visuals, and viewers can develop an intuitive grasp of it. For teaching things like taxonomies or hierarchical relationships, hyponymy works well—use concrete examples to stand for broader categories, and viewers can build systematic knowledge frameworks. Collocation fits situations where abstract concepts need a context; showing typical scenes helps call up what viewers already know. As for repetition, it should always be combined with other relationship types. The repeated content and the images need to work together meaningfully, so avoid mindless repetition or designs where the background has no connection to the text.

At the interpersonal level. On online video platforms, viewers can leave a video any time they want. So, boosting interactivity becomes key to better communication outcomes. One thing designers can try is putting a host on screen. Direct eye contact with the camera creates a kind of virtual interaction with the audience. For the text itself, try adding a few questions or commands here and there, not just statements. That way, viewers stay alert and think things through—they are not just passively taking everything in. Medium shots and frontal camera angles also do a good job when designers want to present knowledge objectively. They make the experience more immersive but still keep things authoritative.

In terms of composition, the structure of each shot should maintain a certain balance: not only be predictable enough to reduce cognitive burden, but also have enough diversity to maintain interest. When the video adopts a consistent structure, the audience can quickly master the specific viewing style. This reduces the energy required for the audience to identify these elements. This fixed structure is very suitable for content that requires a specific cognitive framework. On the contrary, the flexible design can maintain visual appeal and help offset the visual fatigue over time. Therefore, it is more suitable for medium and long videos and content that requires the audience's continuous attention. When highlighting specific elements, designers must consciously guide the audience's attention. Techniques such as color contrast, different font sizes and dynamic visual overlay can be used to highlight important information, so as to avoid too flat information hierarchy. In addition, when designing the screen layout, the viewing habits and cultural background of the target audience must be taken into account.

6.3 Limitations and Suggestions for Future Studies

Every study has its limits, and this one is no different. To start, this study only looked at two introductory videos from biology. The future studies can collect videos covering different topics, durations and production styles, and test their intersemiotic complementarity characteristics.

Secondly, this study focuses more on the multimodal characteristics of the video and it pays less attention to the audience's reaction. In future research, intersemiotic analysis can be combined with empirical data of user behavior, such as the number of views, the number of likes and the content of comments. Given that there may be differences in the reaction of audiences from different cultural backgrounds to the complementarity of images and text, cross-cultural comparisons are also meaningful.

Acknowledgements

The authors have no acknowledgements to declare.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of AI use

AI tool is used to produce the line-drawn schematic illustrations in this manuscript, based on the author's observations of coded video screenshots. All data collection, data processing, data analysis, and conclusions were completed by the author. The author has reviewed and takes full responsibility for the entire content of the manuscript, including the AI-assisted elements.

References

Barthes, R. (1977). *Image-music-text* (S. Heath, Trans.). Fontana.

Chang, P., & Tsai, H. J. (2022). Text-image complementarity and genre in English as foreign language textbooks. *Semiotica*, *2022*(244), 53-80. https://doi.org/10.1515/sem-2020-0103

Chen, W. C., & Zhou, Y. L. (2023). An empirical study on factors influencing dissemination effect of short videos in popular science journals in China: Focusing on 50 Chinese outstanding popular science journals in 2020. *Chinese Journal of Scientific and Technical Periodicals*, *34*(12), 1616-1622. https://doi.org/10.11946/cjstp.202305070333

China Internet Network Information Center (2025, July 21). *56th Statistical Report on China's Internet Development*. https://www3.cnnic.cn/n4/2025/0721/c88-11328.html

Cho, S. Y., & Im, T. (2024). The structural relationships among online class related factors, zoom fatigue, learning engagement, perceived achievement, and class satisfaction in university online classes. *Education and Information Technologies*, *29*(15), 19987-20003. https://doi.org/10.1007/s10639-024-12658-3

Di Nubila, K., Ballesteros-Herencia, C. A., Etura, D., & Martin-Jimenez, V. (2023). Technopopulism and Politainment in Brazil: Bolsonaro Government's Weekly YouTube Broadcasts. *Media and Communication*, *11*(2), 137-147. https://doi.org/10.17645/mac.v11i2.6470

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

International Telecommunication Union (2024, November 10) *Facts and Figures 2024-Report index*. https://www.itu.int/itu-d/reports/statistics/facts-figures-2024/index/

Jing, Y. (2025). Multimodal self-construction of international city image: An ideational intersemiotic analysis of Shenzhen Next. *Foreign Language and Literature Studies*, *42*(2), 43-55.

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

Li, R. H., & Fu, X. J. (2024). She hui zhi chi shi jiao xia jian kang ke pu tu shu duan shi pin de chuang xin lu jing—Yi dou yin ping tai wei li [Innovative approaches to health education books and short videos from a social support perspective—A case study of the Douyin platform] (in Chinese). *Science-Technology & Publication*, *43*(6), 108-115. https://doi.org/10.16510/j.cnki.kjycb.20240613.003

Liang, B., & Jiang, P. (2015). Multimodal tourism discourse analysis and the spreading of Chinese culture. *Foreign Language Research*, *30*(2), 155-158. https://doi.org/10.16263/j.cnki.23-1071/h.2015.02.030

Liu, J. R., Wu, K. R., & Shang, J. H. (2026). Ke pu qi kan yu ke pu qi kan duan shi pin: Gong sheng ji li yu gong sheng lu jing [Science popularization journals and science popularization short videos: Mechanisms and pathways of symbiosis] (in Chinese). *Publishing Research*, *42*(1), 53-59. https://doi.org/10.19393/j.cnki.cn11-1537/g2.2026.01.011

Liu, J. R., Xue, Z. H., & Wu, K. R. (2025). Factors influencing the dissemination effect of WeChat short videos in youth science popularization journals based on HSM model and platform optimization strategies. *Chinese Journal of Scientific and Technical Periodicals*, *36*(2), 207-217. https://doi.org/10.11946/cjstp.202409281080

Liu, Y., Zhao, N., Wang, L., Yu, S. S., Deng, L. X., Zhang, X., Wu, Y. B., Shi, L. L., Fang, Q., & Li, N. (2024). Exploration of science popularization creation strategies based on academic journal content: A case study of Medical Journal of Peking Union Medical College Hospital. *Chinese Journal of Scientific and Technical Periodicals*, *35*(10), 1416-1424. https://doi.org/10.11946/cjstp.202404110357

Nakakuwa, T., & Jawahar, K. (2020). Intersemiotic complementarity in Namibian physical science teachers' classroom practice. *African Journal of Research in Mathematics, Science and Technology Education*, *24*(3), 387-399. https://www.tandfonline.com/doi/full/10.1080/18117295.2020.1837509

Royce, T. D. (2007). *New directions in the analysis of multimodal discourse*. Routledge.

Shen, H. Z., Zhu, J., Huang, S. J., & Yuan, Q. J. (2024). Research on the dissemination effectiveness of emergency knowledge short videos: An analysis based on different types of publishers. *Information Studies: Theory & Application*, *47*(11), 101-110. https://doi.org/10.16353/j.cnki.1000-7490.2024.11.011

Sun, Y. Z., Tang, S. H., Zhao, Y. X., & Li, X. S. (2025). How do short videos empower the "viral

dissemination" of national security knowledge?—Research based on the hybrid method of NCA-fsQCA. New Media and Society, *15*(5), 61-80. https://kns.cnki.net/kcms2/article/abstract?v=_oeCfPpuuXFRGA-fO5khG0O3dMrFP11KdUzaqFACVuz0ImnH4bQEB0vnRsl1A_V0GiTu0UZO5CRtEE1GsSb7GF63F3aVDr8qpvnRYxsdA_5OE2Ibke83OsbI8NsxZkFuiZcOGd5zqQMcFuSMuwMv1bTUFrDhVH1E5_7ywLj-9Vg=

Wang, L. T., Duan, L. M., & Tu, W. J. (2024). *Jian ying zhuan ye ban: Duan shi pin chuang zuo an li jiao cheng (quan cai mu ke ban) [Jianying pro: Short video creation case study tutorial (full-color MOOC edition)]*. Posts & Telecom Press.

Wang, X. D, Zhao, Y. X., Song, X. K., & Zhao, M. Y. (2025). Genre types and interactions of online health information outreach practices for older adults—Insights from a Bilibili-based netnography. *Journal of Modern Information*, *45*(8), 31-42. https://doi.org/10.3969/j.issn.1008-0821.2025.08.003

Xiong, Z. W., Zhao, Z. H., & Chen, X. (2024). Fu ji hu bu li lun xia ying yu jiao cai tu xiang zi yuan de fen xi yu shi yong [Analysis and use of visual resources in English textbooks based on the theory of inter-semiotic complementarity] (in Chinese). *Foreign Language Teaching in Schools*, *47*(9), 44-49. https://kns.cnki.net/kcms2/article/abstract?v=_oeCfPpuuXEkyI4tS5wtNzGegiAAXcDgKZ9msPeQz5_iaop21anDWUjOfwcDzE_AT7PFbSDVA3Y7GReU3MMjFh_ETsVyhq6CPw7DH9M6VYwxNPRUBUDcD8Ki5sNqG6DLStSXCI6KDpxrxV5S4qxqXe8q9qfiYPswg71zBytUCi4=

Xu, J., & Liu, Y. (2023). A study on the influencing factors of information communication effect of popular science videos: A case study of the institute of physics CAS official account on Bilibili. *Library Journal*, *42*(11), 108-116. https://doi.org/10.13663/j.cnki.lj.2023.11.012

Yang, S. (2012). Fu ji hu bu li lun—Duo mo tai hua yu fen xi li lun kuang jia [The theory of intersemiotic complementarity: A theoretical framework for multimodal discourse analysis] (in Chinese). *Journal of Yangtze University (Social Sciences)*, *35*(7), 50-52. https://kns.cnki.net/kcms2/article/abstract?v=_oeCfPpuuXGvyPTT2q25vOvvTCXw2TX5i21mQalqjB1B0BVYkKnnVRt3oc6ge6h_qQwpfksODh89kx3_TomKdoRI_2T2NKUhDG_xbNvLvMIKQgxGZp__hhuRO5pmfCDdM8HvUI-Cs38yHZIAiNa8drR4jXzJ3avsV9AxGjGSqcE=

Yu, S., & Wu, Z. (2024). Research on the influence mechanism of short video communication effect of furniture brand: Based on ELM model and regression analysis. *BioResources*, *19*(2), 3191-3207. https://doi.org/10.15376/biores.19.2.3191-3207

Yuan, S., & Qiu, L.Y. (2026). Dou yin ke pu lei shi pin bo chuan bo ce lue tan jiu—Yi "zhi zao de yuan li: zhong guo zhi zao" zhang hao wei li [An exploration of dissemination strategies for educational videos on douyin: A case study of the "principles of manufacturing: made in china" account] (in Chinese). *News World*, *38*(1), 20-22. https://doi.org/10.19497/j.cnki.1005-5932.2026.01.030

Zhan, L. J., & Sun, X. H. (2025). Influencing factors and optimization strategies of short video communication effect of popular science journals: Empirical analysis based on the content of

WeChat channels of “China’s Excellent Popular Science Journals.” *Chinese Journal of Scientific and Technical Periodicals*, *36*(7), 897-906. https://doi.org/10.11946/cjstp.202503100234

Zhang, D. L. (2009). On a synthetic theoretical framework for multimodal discourse analysis. *Foreign languages in China*, *6*(1), 24-30. https://doi.org/10.13564/j.cnki.issn.1672-9382.2009.01.004

Zhang, Y. L. (2013). Intersemiotic complementarity in print public service advertising discourse—A case study of world tuberculosis day poster. *Journal of Zhongzhou University*, *30*(5), 80-83. https://kns.cnki.net/kcms2/article/abstract?v=_oeCfPpuuXGEsZlQBXDJdJlnVgrhSvH_NAH4vjposDNc_RqdxyyJT_edoFkdSc1nUJNb21g1TF22YUiHOFatfhRmYN8-erBZ86Ikqrb6l4pDwjA1_3m-dDYcvDrIqfNL0YkCUehu273YEbT24daxTyj4UBmtmCdvPTNwnKwGmHk=