Vocabulary learning through videos: captions, advance-organizer strategy, and their combination

(Mark) Feng Teng

Department of Education Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong

ABSTRACT
This study intended to examine L2 young learners’ vocabulary learning (i.e. form, meaning, and use). The research design involved a 2 (advance-organizer strategy: present vs. absent) × 4 (caption type: glossed full captions, glossed keyword captions, full captions, keyword captions) between-subjects design. A total of 240 Chinese ESL primary school students were recruited and randomly and equally assigned to eight conditions. Findings revealed that captioned videos produced significant effects on learning regarding the three dimensions of vocabulary knowledge. Glossed full captions were found to be the most effective caption type. Administration of the advance-organizer strategy also resulted in better gains in learning form, meaning, and use than the absence of this strategy. The combination of the advance-organizer strategy and glossed full captions led to the best learning performance for each dimension of vocabulary knowledge. Relevant pedagogical implications, including use of the advance-organizer strategy and captions, are provided.

KEYWORDS
Advance-organizer strategy; captions; glossed full captions; vocabulary learning; vocabulary knowledge

1. Introduction
Vocabulary is essential to language learning. It is a particularly important factor for students learning English as a second language (ESL). However, vocabulary acquisition can be arduous for ESL learners (Webb & Nation, 2017). Instructed classroom time for students is often limited to learning the 8,000- to 9,000-word families required for text comprehension (Nation, 2006). Learners require more time to learn words independently, incidentally, and continuously out of class. In Hong Kong, the English language curriculum in primary school has long attached great importance to the learning and teaching of vocabulary-building skills. However, learning vocabulary is not easy. Considering the limited English classroom hours allocated for primary school students in Hong...
Kong, particularly in public schools where Chinese is the medium of instruction, it is not surprising that learners may lack the capacity to handle this vocabulary learning burden (Lo & Murphy, 2010). ESL learners in Hong Kong, as well as in similar contexts, must seek outside help to learn vocabulary.

Fortunately, today’s language learners have ever-increasing access to culturally rich and enjoyable online materials. Different approaches have been used to assist and motivate learners in learning vocabulary from these materials. Technological advances have afforded learners various tools that may allow better access to such materials as well. Multimedia language learning tools, including videos, apps, online vocabulary games, and digital texts, are becoming more prevalent with the increased availability of computer technology. The pervasiveness and ubiquity of technology is well accepted. Thus, integrating technology into vocabulary instruction is valuable; this integration may be suited to the urgent need to improve students’ vocabulary, which can enhance their literacy in school and beyond. Widely available digital tools may be one resource teachers can harness to provide scaffolds for students to learn words more effectively.

Among multimedia tools, captioned videos, which are accompanied by synchronous on-screen L2 texts to reinforce video comprehension (Danan, 2004), have received significant attention in recent years (Teng, 2020). Captioned videos have helped learners at various skill levels to visualize what they have heard, thereby enhancing vocabulary acquisition (Peters & Webb, 2018; Teng, 2019a, 2019b, 2019c). Winke, Gass, and Sydorenko (2010) documented that captioned videos may help learners perform spoken-word processing at greater depth, become engaged in video content, and reinforce knowledge structures. Thus, the use of captioned videos may be considered an added benefit rather than an imposed burden, through which learners can better connect auditory input with visual input to acquire new words (Vanderplank, 2016).

However, non-native speakers may have limited cognitive ability to speculate about, predict, and comprehend information when immersed in a flow of foreign utterances in today’s digital world due to a lack of sufficient language proficiency and cultural background knowledge (Taylor, 2005). On-screen text—in the absence of attention-enhancing techniques—may not help learners process the bimodal channel of audio input and visual pictures (Montero Perez, Peters, & Desmet, 2015). Challenges in learning from multimedia materials may become more evident when attempting to engage ESL learners in vocabulary learning. In addition, linking the meaning of new words to their orthographic representation while engaging in reading or listening tasks can be difficult (Chen & Teng, 2017; Teng, 2018a). Visual salience techniques or
attention-enhancing techniques may be necessary to help learners adapt to English natives’ natural interactions, speech-delivery speed, or the syntactic complexity of multimedia materials (Vanderplank, 2016).

The current quasi-experimental study, which infused praxis-based pedagogy into empirical investigations, addresses two attention-enhancing techniques: (a) visual salience in the captioning line (e.g. glossed full captions, full captions, glossed keyword captions, and keyword captions); and (b) visual salience in advance-organizer activities, where students were either administered these activities or not. Despite the potential, studies are scarce regarding the interplay of attention-enhancing techniques in captioned videos and advance-organizer activities in vocabulary learning. This innovative study is expected to establish a blueprint for how ESL young learners’ vocabulary learning can be augmented by incorporating both attention-enhancing techniques (i.e. captioned videos and advance organizers) into vocabulary learning tasks.

2. Literature review

2.1. Theoretical framework for captions

L2 captions, a type of on-screen textual information presented in learners’ target language, can lead to a triple association between image, sound, and text (Danan, 2004). According to Multimedia Learning (Mayer, 2001), effective learning can occur when multimedia instruction helps learners reduce their extraneous cognitive load and increase their germane cognitive load during learning. Based on this theoretical conceptualization, the combination of image, sound, and text in captioned videos may prompt learners to perceive, understand, subsume, and merge new information within their mental system (Plass & Jones, 2005). Ineffective learning may occur when learners’ attention is distracted. Learning vocabulary from English videos alone may not help learners construct links to retrieve meanings (Taylor, 2005). Learners may need visual aids, such as attention-enhancing captions, to construct stronger meaning representations for retrieval.

2.2. Theoretical framework for advance organizers

‘Advance organizers’ have been operationalized as relevant materials or activities introduced prior to the learning material itself to establish a meaningful learning condition (Corkill, 1992). Such a strategy links unfamiliar knowledge to what learners already know and functions as a subsuming bridge that can connect new learning materials and mental ideas (Hall & Strangman, 2008). In this study, the advance organizer is
defined as the provision of background information to improve students’ comprehension of foreign-language materials under the assumption that comprehension can be gained from an introductory sequence of exercises. Through such a previewing stage prior to video exposure, EFL learners may be encouraged to acquire some words.

The technique of advance organizers is supported by subsumption learning theory (Ausubel, 2006). According to this theory, learners can acquire information via visual means and text materials (Corkill, 1992). The presentation of linguistic and pictorial information can help learners establish a more robust mental representation of content prior to watching a video. The adoption of advance organizers is based on Ausubel (2006), who found that learning is enhanced when superordinate, representational, and combinatorial processes occur during information comprehension. Despite criticisms of not specifying the teacher’s role (Barnes & Clawson, 1975), research has demonstrated the effectiveness of advance organizers in promoting learners’ comprehension of screen-based materials (Hung & Chao, 2007). An advantage is that the substantive content of a given organizer can be selected based on its suitability for explaining, integrating, and interrelating the preceding materials. This type of design can enhance the organizational strength of cognitive structures and lead to better learning outcomes; for instance, this technique prepares cognitive structures (e.g. schemas and conceptual patterns) for a subsequent learning condition.

2.3. An overarching framework for combining captions and advance organizers

The effectiveness of combining captioned videos and advance organizers can be explained through dual coding theory (Paivio, 2008), in which information simultaneously encoded verbally and nonverbally is likely to be stored in memory rather than in either verbal information processing or nonverbal information processing alone. According to the cognitive-affective model (Vanderplank, 2018), whether a learner-viewer would put the essential effort into watching a captioned video to gain in language learning is related to the perceived self-efficacy that a learner-viewer is likely to bring to watching. Chung (2002) evaluated different modality combinations on listening performance by measuring the independent and combined effects of advance organizers and captioned videos. Results indicated that the combination of advance-organizer activities and captioned videos was most effective for listening comprehension. We may assume that learners’ self-efficacy in learning from captioned videos can be strengthened by integrating captions and advance organizers.
Based on Teng (2019c), we may see the value of adding previewing activities (e.g. advance-organizer activities) to captioned videos when aiming to enhance learners’ ability to comprehend and learn unknown words. In this respect, combining advance organizers and captioned videos could function as a springboard for L2 learners’ vocabulary development through simultaneous provision of audiovisual and practice information. The process may not appear as overwhelming bimodal input; rather, it may serve as a support to reduce cognitive load when supplemented by previewing advance-organizer activities (Chung, 2002). This approach can thus help learners synthesize multiple representations of the same information. Paivio’s (1986) early dual coding theory posited that memory consists of one verbal and one visual system to be activated. Drawing from this theory, dual-modal information presentation that demonstrates an internal relationship between captioned videos and graphic organizers may be effective in helping learners reorganize and integrate background and newly acquired information.

2.4. Empirical studies on using captions for vocabulary learning

Videos can be a teaching aid for young learners’ vocabulary learning. Neuman, Wong, Flynn, and Kaefer (2018) conducted two studies to evaluate the benefits of online streamed videos on vocabulary learning. The first study involved a content analysis of 100 educational media language programs. The coding of 2,000 scenes revealed that the videos included ostensive cues (assisting young learners in gaining definitional information) and attention-directing cues (orientating learners to a target word). The second study adopted eye-tracking technology to identify the predictive effects of videos on learners’ ability to recognize program-specific vocabulary. Attention-directing cues, rather than ostensive cues, played more important roles in the young learners’ vocabulary learning outcomes. These results pointed out the value of salience in digital media—including videos—in enhancing young learners’ vocabulary learning opportunities.

The effects of captioned videos on vocabulary learning have received extensive attention as of late. For example, Montero Perez, Peters, Clarebout, and Desmet (2014) examined Flemish undergraduate students’ vocabulary gains within four groups: full captioning with highlighted keywords, full captioning, keyword captioning, and a control group. Differences among the groups lay in the amount of text presented on-screen (full captions, keyword captions, or no captions) and the visual salience of keywords (highlighted keywords group). Learners who viewed captions while watching videos outperformed the control group on form and meaning recognition tests. Visual salience (highlighted keyword
groups in full captions) did not help learners achieve higher vocabulary mean scores in form and meaning than those obtained by the full captioning group.

Montero Perez, Peters, and Desmet (2018) examined two enhancement techniques for vocabulary learning from videos: L2 captioned videos (glossed keyword captioning, full captioning, keyword captioning, and no captioning) and a test announcement (informing or not informing). This $2 \times 4$ quasi-experimental design involved 227 Dutch-speaking university students. The design targeted receptive form and meaning knowledge, including form recognition, clip association, and meaning recall. Results revealed that within the condition wherein learners were informed of the test, glossed keyword captioning resulted in the best outcomes for the three dimensions of vocabulary knowledge. Meaning recall constituted learners’ least acquired form of knowledge. A questionnaire regarding learners’ look-up behaviors while watching videos also showed that behavior in looking up a word was connected to the learning of that word.

Teng (2019a) examined the effect of two independent variables on 257 Chinese primary school students’ vocabulary learning. The variables included captioning videos (i.e. full captioning, keyword captioning, and no captions) and target word exposure frequency (one and three encounters). This $2 \times 3$ quasi-experimental design explored the recognition of word form/meaning and recall of word meaning. Results showed that the full captioning group earned the highest scores, followed by the keyword-captioning group and the no-captioning group. Three encounters with target words led to more successful learning than one encounter. The most realistic condition for learning the three dimensions of vocabulary knowledge was therefore a combination of full captioning and three encounters. Teng’s study was innovative in exploring ESL primary students’ vocabulary learning through captioned videos and provided insights into ESL contexts where learners were heavily dependent on print literacy activities for vocabulary learning.

After reviewing relevant literature, some issues may need to be considered. First, findings related to using captioned videos appear inconsistent. Full captioning does not necessarily yield promising results in vocabulary learning in all cases (Montero Perez et al., 2018). However, in some studies (e.g. Teng, 2019a), full captioning yielded significantly better results for vocabulary learning than keyword captioning. As Teng (2019b) explained, different results from viewing captioned videos may be due to learners’ English proficiency, which influenced their noticing and attention. In addition, videos might include complex and incomprehensible language input, for which learners may need additional enhancement
techniques to focus on key aspects of video input comprehension (Teng, 2020). Hence, captioned videos have the potential to enhance vocabulary learning, but certain underexplored enhancement techniques may need to be tested when stimulating learners’ ‘engagement’ in language input comprehension (Schmitt, 2008, p. 338). The technique should enhance learners’ noticing and attention for acquisition, which represents an initial and crucial step in vocabulary acquisition (Hulstijn, 2001) and second language acquisition (Gass, 1999).

Second, vocabulary learning from watching videos has been deemed challenging. In previous studies (Montero Perez et al., 2014; 2018), only partial learning gains were detected. As explained in Schmitt (2010), various dimensions of vocabulary knowledge (i.e. word form and meaning) are relatively amenable to learning. Other knowledge dimensions, such as word use, are more contextualized and much more difficult to learn within a limited time. Montero Perez et al. (2015) suggested using attention-enhancing features (i.e. the process of making language input salient to the learners) to facilitate vocabulary learning, particularly word use.

2.5. Empirical studies on using advance-organizer activities for vocabulary learning

Advance organizers—visual and graphic displays that present connections between facts, concepts, and ideas within a learning activity—prepare learners to sort essential information, outline key concepts, and sequence main ideas (Hall & Strangman, 2008). As proposed by Vandergrift (2007), advance organizers can help learners build connections between prior linguistic and content knowledge and working memory, thus helping them outline audio and visual information in a multimedia technology setting. The assumption is that learners may be better able to interpret incoming information and achieve meaningful learning when information is organized in a visual format. Thus, advance organizers appear useful for organizing ideas, understanding complex structures, conveying relationships between concepts, integrating new and previously learned knowledge, and evaluating understanding or diagnosing misunderstanding.

A review of the literature on advance organizers unveils numerous benefits for language acquisition, including vocabulary learning. For example, in an early study (Chung & Huang, 1998), learners understood video content better when advance organizers were implemented prior to viewing videos. In line with Herron (1994), the advance-organizer strategy aided video comprehension as learners were offered opportunities to preview structures, vocabulary, and cultural information before video-
watching sessions. Li (2014) also delineated the effectiveness of combining advance-organizer activities and video viewing to promote learners’ content comprehension. Ponce, Mayer, Figueroa, and López (2018) conducted two experimental studies on the advance-organizer strategy: learners were presented with a text on their computer screen and required to highlight the unknown words, and the instructor used system-generated, summarized information about learners’ selected unknown words to guide learners in understanding the text. In Experiment 1, college students who were exposed to this strategy scored higher on a vocabulary test. In Experiment 2, high school students exposed to the strategy scored higher on a standardized reading comprehension test. In a recent study (Teng, 2019c), adding advance organizer activities prior to video viewing enhanced primary school students’ L2 collocation learning.

The reviewed literature involves several assumptions. First, the advance-organizer strategy offers learners an overview of a knowledge domain before actual learning sessions. This characteristic demonstrates the value of using an advance-organizer strategy, as the growing availability of multimedia technologies can result in more informative advance organizers. Presumably, learners can be oriented to multiple retrieval routes to language input and possibly achieve better vocabulary learning outcomes (Teng, 2020). Second, caution is warranted regarding the use of advance organizers. As argued by Khoii and Sharififar (2013), ESL learners—who are commonly exposed to a dominant educational system oriented toward memorization and examinations—may be more likely to perform rote memorization rather than to use advance organizers. Third, the effects of an advance-organizer strategy on ESL learners’ vocabulary learning can be constrained by (a) the extent of learners’ attention to and engagement with vocabulary learning and (b) their preferred personal strategies for vocabulary learning (Teng, 2019c). As proposed by Griffin, Malone, and Kameenui (1995), primary school learners required instructional support with graphic organizers to comprehend novel textual material. Finally, words are learned to varying degrees, a principle known as incrementality (Schmitt, 2010). Learners may need to understand a word in multiple contexts, including by connecting its relationship to other words (Nagy & Scott, 2000; Teng, 2019d). This requirement underscores the essence of presenting new words in meaningful contexts. Different types of advance-organizer activities, such as storytelling, skimming, and graphic organizers, are common approaches that may help learners learn, attend to, and understand new information more easily. In the present study, the advance-organizer activity aims to present unfamiliar vocabulary words in sentences that depict events in scenario-based schemata. This technique was expected to orient young learners’ attention toward vocabulary learning.
2.6. The present study

The present study attempts to expand previous literature by testing L2 captioned videos and advance organizers and examining the combination of captions and the advance-organizer strategy. The assumption is that the use of captions and the advance-organizer strategy can promote video-based vocabulary learning. Such a multidimensional approach may be useful for helping learners better process learning materials. To the author’s limited knowledge, no studies have been conducted to measure vocabulary learning on the basis of using complex and nuanced captions or the advance-organizer strategy. Three research questions guide this study:

1. What is the effect of the advance-organizer strategy on vocabulary learning?
2. What are the effects of caption types on vocabulary learning?
3. What is the effect of combining L2 caption types and the advance-organizer strategy on vocabulary learning?

3. Method

3.1. Participants

Participants included 240 Chinese-speaking students from six public primary schools in Hong Kong. The medium of instruction for the six schools was Chinese, and all participants were taking ESL courses. Students were recruited using convenience sampling. The author’s friend, who was a primary school teacher for sixth-grade students, helped contact five other sixth-grade teachers. Roughly 100 students were enrolled in each school, and all were invited to participate in the study. Ultimately, 320 students took part. This research focused on an analysis of data collected from 240 students, as they were of intermediate level on a vocabulary test. 80 students’ data were excluded because 44 students were of advanced level, and 36 students were of low proficiency level. The insufficient number of participants at the advanced and low proficiency level were not included. The participants were between 11 and 12 years old. The 240 participants were gathered and then equally and randomly assigned to eight groups (see the Research design section). Based on the ANOVA results of a vocabulary test (see the Instruments section), the eight groups did not show significant differences in their vocabulary level, $F(7, 232) = .832$, $p = .562$. Signed consent forms were obtained from participants and their parents for learners’ participation. The participants were instructed that this study
involved watching English language videos and completing exercises; however, the true aim of testing vocabulary learning from viewing videos was not revealed. Participants received a meal coupon as thanks for their involvement.

Six primary school teachers who were not teaching the participants volunteered to take part in the study. They assisted with related activities, including selecting videos, keywords, and target words; inserting captions in videos; proctoring the learning and testing processes; and scoring learners’ tests.

3.2. Research design

The present study—focusing on two independent variables (advance-organizer strategy and caption types)—adopted a $2 \times 4$ between-subjects design. The first independent variable was the presence or absence of using an advance-organizer strategy (AOS) before viewing videos. The second independent variable included four captioning types (i.e. on-screen texts) in the videos: full captioning (FC), keyword captioning (KC), glossed keyword captioning (GKC), and glossed full captioning (GFC). The combination of variables resulted in eight experimental groups: FC + AOS, KC + AOS, GKC + AOS, GFC + AOS, FC, KC, GKC, and GFC. The 240 participants were gathered and then assigned to groups equally and randomly.

3.3. Materials

3.3.1. Videos

Four short storytelling videos on YouTube were used for all groups. Brief videos were selected for these students because learners of this age may not be able to cope with information in four long videos due to their cognitive load. In addition, storytelling videos were intended to pique these young learners’ curiosity. The average length of each video was 4 minutes, 11 seconds. Six primary school teachers suggested that the topics and image functionality of videos were appropriate for primary school sixth-grade students. The Flesch Reading Ease scores of the videos’ text readability ranged from 65.7 to 71.7, suggesting similar difficulty levels across videos.

3.3.2. Videos with captions

Videos included either full captions, glossed full captions, keyword captions, or glossed keyword captions. Full captions consisted of verbatim transcription of the narrator in the videos and were presented
concurrently with spoken audio. Video viewers who received full captions observed synchronized on-screen texts related to what was being said in the videos. Keywords were important words essential to learners’ meaning identification and construction and were selected by the six primary school teachers by reading the video transcripts and participating in a discussion session. Each keyword consisted of either one word (e.g. ‘extravagant’) or a maximum of four consecutive words (e.g. ‘abide by the rules’). Similar to full captions, keyword captions appeared in the center of the captioning line. The duration of keyword appearances on the video was 1.6 seconds but was subject to keyword length and type. The final set of keywords accounted for 19.44% of all words in the transcripts (i.e. 413 words out of 2,124 words).

Glossed full captions and keyword captions were defined as captions including access to meaning. One or two difficult words were salient in each captioned line, and students were allowed to use the mouse to click each word. These salient words helped learners better comprehend the sentence. This action paused the video and simultaneously displayed the L1 meaning of the clicked word above the target words. When the learners clicked on the mouse again, the video resumed, and the gloss disappeared. Provision of glosses was in line with Teng’s (2019d) suggestion that glosses (i.e. short definitions or explanations of the meaning of words) have the potential to decrease students’ incorrect guesses, connect background knowledge and text, and improve learners’ comprehension of new words. Glossing in the present study differed from in-text translation; in this case, a gloss refers to a brief explanation of a difficult or obscure word or expression in one’s target language. For this study, presenting explanations in learners’ L1, rather than L2, may facilitate learners’ comprehension, attention, and recall of video content; learners can have visual exposure to language input when the learning experience is in the language the students are learning. The use of glosses in captioned videos was assumed to provide the following benefits to young learners: (1) minimize video interruptions and ensure the flow of reading and listening; and (2) help learners avoid interference when comprehending new words. Figure 1 presents a snapshot of glossed full captions; glossed keyword captions are illustrated in Figure 2.

Videos included narration by a native speaker who was invited to read the transcripts aloud. The speed for reading aloud and presenting captions was 90 words per minute, a reasonable rate for young learners watching English videos (Tyler et al., 2009). The new audio file was inserted through the program Wondershare Filmora to replace the original audio track. Video captions were inserted using the program MAGpie 2.5.
3.3.3. Target words
The target words consisted of 20 words unknown to the learners. Six teachers independently identified difficult words from transcripts and then discussed their chosen words. As most words in the videos were simple, the teachers included several difficult words with similar meanings to replace some of the easier words that appeared in videos. For
example, ‘extravagant’ replaced ‘luxury.’ Fifty simple words were replaced with more challenging words. According to the six teachers, the 50 target words did not appear in learners’ textbooks. Based on the RANGE program (Nation & Heatley, 2002), all target words extended beyond the 3,000-word families. Based on the results from a vocabulary test (see the Instrument section), the participants did not reach strong mastery of the 3,000- and 5,000-word families. It is assumed that the participants may not have had prior knowledge of the target words. These results were further confirmed by a pre-test, which was adopted from Wesche and Paribakht (1996) vocabulary knowledge scale (Table 1). This test was used to measure learners’ pre-study knowledge of target words (Teng, 2018b). In addition, a set of 50 high-frequency words were mixed with the 50 difficult words. High-frequency words were assumed to reduce the difficulty of this test, as a test containing all low-frequency words may diminish test-takers’ confidence.

Test results showed that students could identify the meaning of 10 words, all of which were then excluded from the list. The remaining 40 words were inserted into videos. Among them, 20 words, including 8 verbs, 8 adjectives, and 4 nouns, were target words (see Table 2). Although the participants reported not knowing what the other 20 words meant, they mentioned having seen them before. These 20 words were thus excluded to avoid the possibility that participants might in fact know the target words. Each target word was shown only once in one of the videos.

### 3.3.4. Advance-organizer strategy

The advance-organizer strategy, a cognitive instructional strategy, was developed to strengthen learners’ linguistic knowledge gained from pictorial information and promote the learning and retention of new information.

---

**Table 1.** An example of the vocabulary knowledge scale for testing target words.

<table>
<thead>
<tr>
<th>Extravagant</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. I don’t remember having seen this word before.</td>
</tr>
<tr>
<td>II. I have seen this word before, but I don’t know what it means.</td>
</tr>
<tr>
<td>III. I have seen this word before, and I think it means ______________________ (synonym or translation).</td>
</tr>
<tr>
<td>IV. I know this word. It means ______________________ (synonym or translation).</td>
</tr>
<tr>
<td>V. I can use this word in a sentence (write a sentence): _______________________________. (If you do this section, please also do Section IV.)</td>
</tr>
</tbody>
</table>

**Table 2.** Overview of target words per video and per type.

<table>
<thead>
<tr>
<th>Type</th>
<th>Video 1</th>
<th>Video 2</th>
<th>Video 3</th>
<th>Video 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs</td>
<td>Disturb; Nibble</td>
<td>Regulate; vibrate</td>
<td>Glow; Shiver</td>
<td>Nourish; Hover</td>
</tr>
<tr>
<td>Adjectives</td>
<td>Desperate; Despicable</td>
<td>Extravagant; Ridiculous</td>
<td>Stubborn; Startled</td>
<td>Elegant; Gloomy</td>
</tr>
<tr>
<td>Nouns</td>
<td>Bridegroom</td>
<td>Illumination</td>
<td>Twilight</td>
<td>Misery</td>
</tr>
</tbody>
</table>

---
One advantage of the advance organizer is that it can provide learners an overview of information and help them organize and interpret new, incoming information prior to learning sessions (Mayer, 2003). Information included in an advance organizer is not an overview but instead functions as a platform to help learners grasp verbal and/or visual information before viewing a video. The present study, following Li (2014) and Teng (2019c), used pictures and text information from videos, which was effective in providing clues to help learners acquire new and unfamiliar material.

The advance organizer contained 10 episodic pictures from each video for a total of 40 episodic pictures. Four pictures were organized on one webpage, and learners were asked to link each picture with its corresponding two- to three-line English caption (Figure 3). The lines were displayed at the top of the screen, and pictures were displayed at the bottom. Participants could click one option (A, B, C, or D) noted on the right side of each picture, and then the selection would appear in the box. If the selected option did not match the picture after learners clicked ‘Submit the Answer,’ an icon of a sad face appeared on the computer screen. Incorrect answers were highlighted, and learners would be prompted to try once more. The learners sometimes needed to try several times before successfully completing the page. An icon of a happy face appeared on the screen if all chosen answers were correct, after which the learners would be led to the next page. Learners could also listen to a recorded message of each line through their headsets by clicking on the speaker icon next to each sentence. This advance-organizer activity ended after learners provided correct answers on the last webpage. The average time for completing this activity was approximately 50 minutes.

The main rationale for developing the advance organizer was to help learners associate pictorial information with linguistic information to activate cognitive systems for meaningful learning. This technique differs from a traditional drill exercise. As noted by Li (2014), using authentic pictures from videos can provide accurate clues about video content and can encourage learners to reflect on the learning process. This approach also provides a brief synopsis about an upcoming event. The usefulness of this technique for the learning process was assessed using a post-test of 20 words, wherein the advance-organizer technique could offer learners opportunities to activate knowledge of those words.

3.4. Instruments

3.4.1. Vocabulary knowledge test
Learners’ vocabulary level correlates to their vocabulary learning outcomes and provides a rough estimate of English proficiency (Teng, 2020). The
Vocabulary Levels Test (VLT), developed by Schmitt, Schmitt, and Clapham (2001), was used in this study; the test measured learners’ written receptive vocabulary knowledge at the 2,000- to 10,000-word level and was suitable for measuring learners’ vocabulary size. However, as target participants were young Chinese-speaking learners, VLT was used to measure their knowledge at the 2,000-, 3,000-, and 5,000-word frequency levels of English word families. The 240 participants’ average scores were 25.89, 13.17, and 2.11 (out of 30) for the 2,000-, 3,000-, and 5,000-word levels, respectively. The participants appeared to master the 2,000-word level to some degree. The Cronbach’s alpha for this test was .81, indicating sound reliability.

Figure 3. A screenshot of the advance-organizer strategy.
3.4.2. Vocabulary post-test
Vocabulary learning outcomes were measured using a computerized vocabulary test, which included three parts. The purpose of the vocabulary post-test in this study was to measure vocabulary knowledge; according to Nation (2001), the test should consider word form, meaning, and use.

The first part of the test measured word-form recognition. Three distractors and one target word were intermixed to reduce wild guessing. Learners were required to select a word that they believed had appeared in the videos. After finishing the first part, learners clicked ‘next’ and the computer automatically continued to the second part. The second part measured whether learners selected the correct response from four translations in Chinese. Learners clicked onto the next page after finishing this one. The test automatically brought learners to the third part after finishing the second. The third part measured learners’ ability to use the target word in a sentence. Learners selected the correct sentence from four presented options. Cronbach’s alpha for the test sections ranged from .79 to .82, indicating acceptable reliability. Learners were not allowed to return to previous test items after clicking ‘next.’ Learners who answered incorrectly in the first part could not move onto the second part; the same principle applied to the second and third parts. The test period lasted 30 minutes based on the teachers’ recommendation; after 30 minutes, the test session ended automatically. Figure 4 presents a snapshot of the test.

3.5. Scoring
All tests were scored binomially. One point was assigned for a correct answer with no points for an incorrect answer. Two of the six teachers were invited to rate the tests, and if disagreements occurred, a third teacher would be called upon. No scoring-related discrepancies occurred between the first two raters; thus, the third rater was not asked to assist.

3.6. Procedures
The learning sessions lasted for 120 minutes for learners participating in advance-organizer activities and 60 minutes for learners without advance-organizer activities. However, the test results should be generalized with caution, as time-on-task can influence task performance. Ackerman and Kanfer (2009) found that subjective fatigue increased with increasing time-on-task. In addition, learners’ performance was enhanced in conditions involving a longer test. Considering that participants in this study
Figure 4. A snapshot of the computerized test parts (form, meaning, use).
were between 11 and 12 years old, the possible effect of test length on their performance is worth acknowledging. However, the test results would become unreliable if students had not been allotted a specific amount of time for completion. The time limitation was set based on the time required for learners to complete similar exercises. Learning sessions were administered in computer laboratories on a Saturday; all students were required to gather together prior to being randomly assigned into different groups.

Learners were administered the pre-test one month prior to the study to minimize the carryover effect from pre-test to post-test. The instructions for the learning sessions were as follows. First, all students received instructions related to the need to understand video content and how to work individually on a computer and with a headset. Learners receiving the advance-organizer strategy received an additional instruction session in which they completed sample exercises. Second, following the online advance-organizer activity, learners individually clicked a link that allowed them to watch each video only once. After watching the four videos, learners were directed to a page to complete the vocabulary post-test.

3.7. Data analysis

The present study involved a $2 \times 4$ research design. MANOVA was employed to measure overall captioning type and advance-organizer strategy effects on each type of vocabulary knowledge; interactions between the two main variables were also assessed. Bonferroni post-hoc comparisons were performed to analyze data pairs. The MANOVA was followed by discriminant function analysis (Field, 2013). While MANOVA explored whether groups differed along a linear combination of outcome variables, discriminant analysis deconstructed the linear combination in greater detail.

4. Findings

Table 3 presents the descriptive data on word form, meaning, and use, respectively. Figures 5–7 display these results graphically for the three aspects of vocabulary knowledge.

The descriptive statistics in Table 3 indicate a consistent decline in mean scores from less-demanding (e.g. word form) to more-demanding tests (e.g. word use) across all learning conditions. GFC seemed to support optimal vocabulary learning performance among the four caption types. Adding AOS to each captioning condition yielded higher scores in
either word form, meaning, or use compared to the absence of such a strategy.

In terms of the three figures (Figures 5–7), the central line on each box-whisker plot denotes the median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data and the points not considered outliers.

### Table 3. Descriptive statistics of various vocabulary knowledge in each group.

<table>
<thead>
<tr>
<th>Groups</th>
<th>n</th>
<th>Form M</th>
<th>Form SD</th>
<th>Meaning M</th>
<th>Meaning SD</th>
<th>Use M</th>
<th>Use SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFC + AOS</td>
<td>30</td>
<td>19.033</td>
<td>.89</td>
<td>16.867</td>
<td>.97</td>
<td>14.633</td>
<td>.81</td>
</tr>
<tr>
<td>FC + AOS</td>
<td>30</td>
<td>15.167</td>
<td>.83</td>
<td>13.167</td>
<td>.91</td>
<td>10.967</td>
<td>.88</td>
</tr>
<tr>
<td>GKC + AOS</td>
<td>30</td>
<td>16.967</td>
<td>.85</td>
<td>15.033</td>
<td>.71</td>
<td>12.967</td>
<td>.81</td>
</tr>
<tr>
<td>KC + AOS</td>
<td>30</td>
<td>13.033</td>
<td>.71</td>
<td>10.967</td>
<td>.80</td>
<td>8.933</td>
<td>.74</td>
</tr>
<tr>
<td>GFC</td>
<td>30</td>
<td>11.033</td>
<td>.76</td>
<td>8.933</td>
<td>.82</td>
<td>6.933</td>
<td>.74</td>
</tr>
<tr>
<td>FC</td>
<td>30</td>
<td>6.900</td>
<td>.75</td>
<td>5.033</td>
<td>.85</td>
<td>2.933</td>
<td>.73</td>
</tr>
<tr>
<td>GKC</td>
<td>30</td>
<td>8.933</td>
<td>.69</td>
<td>6.967</td>
<td>.76</td>
<td>4.933</td>
<td>.78</td>
</tr>
<tr>
<td>KC</td>
<td>30</td>
<td>5.100</td>
<td>.80</td>
<td>2.867</td>
<td>.89</td>
<td>1.167</td>
<td>.91</td>
</tr>
</tbody>
</table>

*Note.* The maximum score for each test part is 20 points.

![Figure 5. Word form scores.](image1)

![Figure 6. Word meaning scores.](image2)
Outliers are plotted individually using the ‘+’ symbol. Although GFC in Figure 4 shows little variability, evidence suggests that GFC + AOS (glossed full captioning video and advance-organizer strategy) scores were the most spread and yielded the highest mean scores in learning word form, meaning, and use.

Table 4 presents the result of Box’s test, which measures the assumption of equality of covariance matrices. Based on the *p* value (*p* = .512, larger than .05), this statistic was not significant. Hence, the covariance matrices were roughly equal as assumed.

As the covariance assumption was not violated, the next step was to use Wilks’ lambda in multivariate tests to illustrate the effects of independent variables (i.e. caption types and advance-organizer strategy; see Table 5). As shown in Table 5, Wilks’ lambda reached the criterion for significance (*p* < .05). The caption type, use of the advance-organizer strategy, and interaction between captions and the advance-organizer strategy each appeared to exert significant effects on vocabulary learning.

The next step was to identify between-group differences. Univariate test results were determined to identify the nature of the effect. Before presenting these results, a summary result of Levene’s test for equality of variances for each dependent variable was obtained as shown in Table 6. Results of the Levene’s test were non-significant for the three dependent variables (*p* > .05). These findings strengthened the
case for assuming that the multivariate test statistics were robust and that it was reasonable to run univariate tests.

Univariate test results are presented in Table 7. The $p$ values in Table 7 indicate a significant difference between caption types in terms of word form ($p < .001$, $\eta_p^2 = .89$), meaning ($p < .001$, $\eta_p^2 = .876$), and use ($p < .001$, $\eta_p^2 = .88$). There were also significant differences between using or not using AOS in terms of word form ($p < .001$, $\eta_p^2 = .946$), meaning ($p < .001$, $\eta_p^2 = .959$), and use ($p < .001$, $\eta_p^2 = .961$). Similarly, significant interaction effects of caption types and AOS interaction were found on word form ($p < .05$, $\eta_p^2 = .606$), meaning ($p < .001$, $\eta_p^2 = .602$), and use ($p < .001$, $\eta_p^2 = .609$). The post-hoc pairwise comparison in Table 8 indicates significant differences among caption types. Results showed that, irrespective of word knowledge dimensions, GFC was the most effective caption type. The post-hoc pairwise comparison in Table 9 demonstrated that, again irrespective of word knowledge dimensions, using AOS was more effective than not using AOS. The post-hoc pairwise comparison in Table 10 demonstrated that, in terms of learning word form, meaning, and use, GFC + AOS was more effective than the other seven conditions ($p < .001$).

The MANOVA was followed by discriminant analysis, which revealed two discriminant functions: caption types and AOS. The first function (caption types) explained 63.2 percent of the variances (canonical $R^2 =$
whereas the second function (+/- AOS) explained 36.8 percent of the variances (canonical $R^2 = .12$). The combination of these discriminant functions significantly differentiated the treatment groups, $L = .71, \chi^2(4) = 9.67, p = .03$. When removing the first function, the second function also significantly differentiated the treatment groups, $L = .82, \chi^2(1) = 8.81, p = .04$. Correlations between the outcomes and 

---

### Table 8. Post-hoc pairwise comparisons among the caption types.

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form GFC FC</td>
<td>4.000*</td>
<td>0.144</td>
<td>0.000</td>
<td>3.616</td>
<td>4.384</td>
</tr>
<tr>
<td>GKC</td>
<td>2.083*</td>
<td>0.144</td>
<td>0.000</td>
<td>1.699</td>
<td>2.468</td>
</tr>
<tr>
<td>KC</td>
<td>5.967*</td>
<td>0.144</td>
<td>0.000</td>
<td>5.582</td>
<td>6.351</td>
</tr>
<tr>
<td>FC</td>
<td>-4.000*</td>
<td>0.144</td>
<td>0.000</td>
<td>-3.842</td>
<td>-3.616</td>
</tr>
<tr>
<td>GKC</td>
<td>-1.917*</td>
<td>0.144</td>
<td>0.000</td>
<td>-2.301</td>
<td>-1.532</td>
</tr>
<tr>
<td>KC</td>
<td>1.967*</td>
<td>0.144</td>
<td>0.000</td>
<td>1.582</td>
<td>2.351</td>
</tr>
<tr>
<td>GKC</td>
<td>-2.083*</td>
<td>0.144</td>
<td>0.000</td>
<td>-2.468</td>
<td>-1.699</td>
</tr>
<tr>
<td>FC</td>
<td>1.917*</td>
<td>0.144</td>
<td>0.000</td>
<td>1.532</td>
<td>2.301</td>
</tr>
<tr>
<td>KC</td>
<td>3.883*</td>
<td>0.144</td>
<td>0.000</td>
<td>3.499</td>
<td>4.268</td>
</tr>
<tr>
<td>GKC</td>
<td>-5.967*</td>
<td>0.144</td>
<td>0.000</td>
<td>-6.351</td>
<td>-5.582</td>
</tr>
<tr>
<td>FC</td>
<td>-1.967*</td>
<td>0.144</td>
<td>0.000</td>
<td>-2.312</td>
<td>-1.582</td>
</tr>
<tr>
<td>KC</td>
<td>-3.883*</td>
<td>0.144</td>
<td>0.000</td>
<td>-4.268</td>
<td>-3.499</td>
</tr>
</tbody>
</table>

### Table 9. Post-hoc pairwise comparisons between using or not using advance organizers.

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without AOS</td>
<td>8.058</td>
<td>0.102</td>
<td>0.000</td>
<td>-8.260</td>
<td>-7.857</td>
</tr>
<tr>
<td>With AOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without AOS</td>
<td>8.058</td>
<td>0.109</td>
<td>0.000</td>
<td>-8.274</td>
<td>-7.843</td>
</tr>
<tr>
<td>With AOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without AOS</td>
<td>-7.883*</td>
<td>0.104</td>
<td>0.000</td>
<td>-8.088</td>
<td>-7.678</td>
</tr>
<tr>
<td>With AOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
discriminant functions revealed that learning related to word form, meaning, and use loaded higher onto the first function ($r = .81$) than the second function ($r = .65$). The discriminant function plot (Figure 8)
shows that the first function discriminated the groups (horizontal dis-
tance between centroids), as did the second function (vertical distance).
However, the effect of the second function was not as dramatic as the
first. Evidently, GFC + AOS was the more effective condition.

5. Discussion
This study examined the effect of the interplay between captioning types
and advance organizers on vocabulary learning among primary school
L2 learners. The research included eight treatment groups across two
advance-organizer strategy × and L2 captioning type conditions. Each
group included 30 participants with similar vocabulary proficiency who
were introduced to 20 novel words. Learners were uninformed about the
test ahead of time. The test focused on three aspects of knowledge: form,
meaning, and use. Overall, findings indicated the effectiveness of cap-
tioning type and the advance-organizer strategy on word learning,
including for word form, meaning, and use. Among the eight groups,
applying the advance-organizer strategy before viewing glossed full cap-
tioned videos yielded significantly better results in word form, meaning,
and use than the other seven combinations.

First, similar to previous studies (Montero Perez et al., 2014; Teng,
2019a, 2019b), captions—a creative means of transforming script into
animated text—may help L2 learners gain more access to online
streamed texts and comprehend unknown words. In Montero Perez et al.
(2018), glossed keyword captioning resulted in a better outcome than full
captioning, keyword captioning, and no captioning for three dimensions
of vocabulary knowledge, namely form recognition, clip association, and
meaning recall. In particular, the present study shows clearer evidence
for using glossed full captioned videos rather than glossed keyword cap-
tioned videos in terms of learning word form, meaning, and use.

Word use while using captioned videos has not been measured in ear-
lier studies; this research reveals a pronounced effect of using glossed full
captioned videos on word use. The benefit of glossed full captions high-
lights the value of making words salient in captioned videos. As sug-
gested by Neuman et al. (2018), attention-directing cues afford learners
more opportunities to learn vocabulary. Other means for enhancing the
visual salience of keywords, such as full captioning with highlighted key-
words (Montero Perez et al., 2014) and full captioning with repeated
encounters with target words (Teng, 2019a), also improved the quality of
learners’ attention to new words. A possible explanation was that
enhancement techniques geared toward word salience in captioned vid-
eos helped learners encode the form–meaning link for new words.
In this study, although on-screen texts were expected to encourage L2 young learners to learn new words more efficiently, glossed keyword captions, full captions, and keyword captions appeared to have less pronounced effects than glossed full captions. Under the full and keyword caption conditions, L2 young learners still found it challenging to infer new word meaning from the real-time nature of video viewing. Even under the glossed keyword condition, the learners may have only discerned partial word meanings. Given the limited time allowed for inferring word meaning from videos and L2 young learners’ limited language proficiency (Teng, 2019b), the process of inferring word meaning was slow, challenging, and time-consuming (Tonzar, Lotto, & Job, 2009). Learners needed to pay attention to the videos and adapt to the video input (Vanderplank, 1990), an approach that was intended to guide L2 young learners in deciphering visual clues while comprehending the meaning of videos. L2 young learners with limited language proficiency still appeared dependent on contextual clues to help them connect new words with previously known words. As suggested by Godfroid, Boers, and Housen (2013), the overlap between known words and novel words could help learners take advantage of cues, thus enhancing opportunities for word learning. In the present study, glossed full-caption videos helped L2 learners notice contextual cues and absorb information from video input. One possible explanation was that the glossed full-caption condition may have helped L2 young learners manage their working memory—a capacity for temporarily holding information for immediate conscious perceptual and linguistic processing (Baddeley, 2012)—more effectively and efficiently. The learners might have been better able to coordinate language input from their strengthened visuo-spatial sketchpads (mainly formed through visual recall exercises) and phonological loops (mainly formed by storing verbal information together with a rehearsal mechanism) while watching videos. Such multiple cueing systems refined Teng’s (2019a) dual-modal presentation technique (Figure 9). This presentation technique, which draws upon the integration of a video’s verbal modal (aural) and pictorial modal (visual), may help learners manage working memory when coupled with activation of L2 learners’ prior vocabulary knowledge from visual enhancement techniques. Learners’ working memory systems functioned as central executives in coordinating input from storage systems of the visuo-spatial sketchpad and phonological loop. The central executive function finally led to a corroborated form–meaning link for new words.

Second, although the effect of the advance-organizer strategy was not as significant as that of captions, these findings underscore the importance of using advance organizers in L2 vocabulary learning. Similar to Ponce et al. (2018), using the advance-organizer strategy elicited higher scores in
vocabulary learning. Complementing studies that highlighted advance organizers in video content comprehension (Chung, 2002; Chung & Huang, 1998; Li, 2014; Teng, 2019c), the present study suggests the potential of using the advance-organizer strategy to help L2 young learners learn new words. As argued by Vandergrift (2007), advance organizers can be applied as a strategy in L2 learning because they may activate learners’ prior knowledge and reduce learners’ cognitive processing load. Activation of prior knowledge and reduction of cognitive processing load may lead to better management of working memory, thus strengthening the comprehension of unknown words. By contrast, learners in each caption condition who did not receive the advance-organizer strategy failed to build a form–meaning link or understand the use of novel words. A potential explanation is that learners lacking information access in the previewing stage could only rely on their linguistic knowledge to cope with caption-reading input. Apparently, these L2 young learners had limited linguistic knowledge to cope with a large amount of information from video input.

The potential of the advance-organizer strategy has been further supported by the subsumption learning theory (Ausubel, 2006). According to this theory, creating instructional material that helps learners comprehend content before learning sessions could make learning more meaningful. In the current study, the presentation of linguistic and pictorial information prior to video viewing may have helped L2 young learners organize a more
robust mental understanding of content. Another possible explanation is that the advance-organizer strategy may function as a cognitive tool that helped young learners link textual and pictorial content to activate vocabulary knowledge in their existing cognitive structure. As proposed by Tang (1992), the strategy of advance organizers prepared learners to structure incoming learning sessions and subsume new information into their existing cognitive structures. Herron (1994) argued that using this strategy prior to video viewing was consistent with Vygotsky’s theory of one’s ‘actual and potential development level,’ namely the zone of proximal development. Essentially, the advance-organizer strategy functions as a tool to bridge the gap between one’s ‘actual development level’ (i.e. what learners are actually able to do) and ‘potential development level’ (i.e. what learners have the potential to do). In this study, advanced organizers highlighted target items, which increased the frequency of exposure to those items. This pattern also accounts for the value of AOS in vocabulary learning. As argued by Teng (2019a, 2019d), word exposure frequency is a determinant of effective vocabulary learning. Advance organizers can guide learners’ internalization process to reach higher development levels and serve as a bridge to link new material and existing knowledge, leading to meaningful learning rather than ‘parrot-like’ memorization techniques (Teng, 2019c). As such, development of the advance-organizer strategy may help L2 young learners decrease cognitive demands, motivate comprehension seeking, develop multidimensional and in-depth understanding of new words and their meanings, and facilitate—perhaps through pre-learning some imagery and text information—their ability to process contextual clues related to L2 target words.

Finally, the present study suggests the potential of combining glossed full captioning and the advance-organizer strategy. This integrated technique greatly enhanced participants’ learning of word form, meaning, and use. For example, in the post-test, learners achieved a mean score of 19.03, 16.86, and 14.63 out of 20 possible points for word form, meaning, and use, respectively (Table 3). The MANOVA was followed up with discriminant analysis, which revealed the advantage of combining the advance-organizer strategy and glossed full captions. Although one may argue that higher exposure to learning contents generally results in greater learning, it is important to consider this study’s objective: to measure incidental vocabulary learning among L2 young learners. Chinese L2 students were not informed of the test and were speakers with limited language processing ability. Findings imply that these students managed to acquire a certain degree of vocabulary knowledge, especially in terms of recognizing form–meaning links. Compared to Montero Perez et al. (2018), wherein Dutch university learners achieved scores of 11.43 and 12.73 (out of 18) on word form and meaning after
watching glossed captioned videos, findings in the present study are encouraging. The advance-organizer strategy may have contributed to the glossed captioning condition. Information organized in the advance-organizer activity may have also helped L2 learners develop a relatively holistic picture of the target video materials (Li, 2014). Additionally, this approach may have enabled L2 young learners to discern the details of the video plot and pay particular attention to unfamiliar words rather than listening to and reading the text for the main idea (Teng, 2019c). The current study thus suggests the potential benefit of using advance-organizer activities prior to watching glossed captioned videos as a means of enhancing L2 young learners’ vocabulary learning.

6. Limitations

Limitations in the present study call for caution in applying these findings. First, participants consisted of L2 primary school Chinese students within a single context. Future studies involving learners at other education levels are needed to generalize the results to diverse L2 learners, particularly those in a reading-texts-dependent ESL setting. Second, a control group of learners not using the advance-organizer strategy or captions was not included for comparison. Third, future studies should include different advance-organizer activities. Fourth, this research only involved storytelling videos; subsequent work should investigate the effects of different video types, including television reports, programs, and documentary videos. Fifth, glosses provided for salient words were in Chinese. Future studies may include more varied glosses. Sixth, variables such as participants’ reading speed or English proficiency may have influenced the results but were not controlled; such variables could be controlled as covariates in future research. Seventh, there is also a need for a delayed test to measure retention. Finally, students who received advance-organizer activities spent twice the time with content as their peers. Even though there was a time limit in each activity, many students finished earlier. One limitation is the lack of control for the precise time-on-task. As argued by Huang, Eslami, and Willson (2012), the larger amount of time in an activity may lead to better learning outcome. Future studies could take the precise time for each activity as a control variable.

7. Implications and concluding remarks

Despite the limitations, this study provides the following implications. First, vocabulary learning is challenging for L2 young learners. Results support the use of captioned videos, particularly glossed full captions, in
L2 vocabulary learning. This may represent an innovation in ESL learning, as vocabulary teaching was enhanced through exposure to lexical chunks in video content. Second, an organizer-aided activity prior to viewing captioned videos may activate L2 young learners’ vocabulary knowledge and help them process prior mental knowledge and upcoming information in videos. This activity enables learners to better process visual images and spoken text. Students who were presented with previewing questions during advance-organizer activities may have focused on finding and confirming their answers during subsequent video viewing. Moreover, the applied question format provided retrieval cues that can activate information when viewing videos. As reported by Chung (2002), open-ended items require students to draw conclusions, understand themes, and synthesize video sections into an overarching concept. The current study provides evidence that an advance-organizer activity of using question items that require learners to seek answers may increase vocabulary learning gains simply through captioned videos. Third, the combination of the advance-organizer strategy and glossed full captions may help L2 young learners discern and internalize information in videos for more effective vocabulary learning. These findings offer insight into the development of curriculum and English materials. Finally, vocabulary learning, through combining glossed full captioning and the advanced-organizer strategy, led to only partial gains. Vocabulary learning for L2 young learners is inherently difficult. Future research could thus extend the present study to gain greater insight into learners’ acquisition of the different components of vocabulary knowledge. In summary, this empirical study is innovative in combining the use of advance organizers with glossed full captions to guide primary school L2 learners to develop vocabulary knowledge from English videos. This technique represents an alternative approach to vocabulary instruction in an L2 context where learners depend heavily on printed texts.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

(Mark) Feng Teng is a language teacher educator with extensive teaching and research experience in China. His professional interests include metacognition and L2 writing, and L2 vocabulary acquisition. He has published widely in international flagship journals, including Language Teaching Research, Applied Linguistics, TESOL Quarterly, Applied Linguistics Review, Computers & Education, and others. He has edited several special issues for international journals, including Asian EFL and Journal of Writing Research.
References


