ReCALL (2023), 35: 1, 105–121
doi:10.1017/S0958344022000088

RESEARCH ARTICLE

Efficacy of different presentation modes for L2 video comprehension: Full versus partial display of verbal and nonverbal input

Chen Chi
National Taiwan Normal University, Taiwan (onlyc215@gmail.com)

Hao-Jan Howard Chen
National Taiwan Normal University, Taiwan (hjchenntnu@gmail.com)

Wen-Ta Tseng
National Taiwan University of Science and Technology, Taiwan (wenta.tseng@ntust.edu.tw)

Yeu-Ting Liu*
National Taiwan Normal University, Taiwan (yeutingliu@gapps.ntnu.edu.tw)

Abstract

Video materials require learners to manage concurrent verbal and pictorial processing. To facilitate second language (L2) learners' video comprehension, the amount of presented information should thus be compatible with human beings' finite cognitive capacity. In light of this, the current study explored whether a reduction in multimodal comprehension scaffolding would lead to better L2 comprehension gain when viewing captioned videos and, if so, which type of reduction (verbal vs. nonverbal) is more beneficial. A total of 62 L2 learners of English were randomly assigned to one of the following viewing conditions: (1) full captions + animation, (2) full captions + static key frames, (3) partial captions + animation, and (4) partial captions + static key frames. They then completed a comprehension test and cognitive load questionnaire. The results showed that while viewing the video with reduced nonverbal visual information (static key frames), the participants had well-rounded performance in all aspects of comprehension. However, their local comprehension (extraction of details) was particularly enhanced after viewing a key-framed video with full captions. Notably, this gain in local comprehension was not as manifest after viewing animated video content with full captions. The qualitative data also revealed that although animation may provide a perceptually stimulating viewing experience, its transient feature most likely taxed the participants' attention, thus impacting their comprehension outcomes. These findings underscore the benefit of a reduction in nonverbal input and the interplay between verbal and nonverbal input. The findings are discussed in relation to the use of verbal and nonverbal input for different pedagogical purposes.

Keywords: multimedia; video captions; animated video; multimodality; English as a foreign language

Cite this article: Chi, C., Chen, H.-J. H., Tseng, W.-T. & Liu, Y.-T. (2023). Efficacy of different presentation modes for L2 video comprehension: Full versus partial display of verbal and nonverbal input. *ReCALL* 35(1): 105–121. https://doi.org/10.1017/S0958344022000088

*All correspondence regarding this publication should be addressed to Yeu-Ting Liu (Email: yeutingliu@gapps.ntnu.edu.tw).

© The Author(s), 2022. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited..

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

1. Introduction

For second language (L2) video comprehension, the pedagogical potency of captions – verbatim transcript in the same language as the spoken narration – has been established by a multitude of studies (e.g. Montero Perez, Van Den Noortgate & Desmet, 2013; Winke, Gass & Sydorenko, 2010, 2013). Despite the supportive evidence, studies based on Mayer's (2002, 2005) cognitive theory of multimedia learning (CTML) have expressed reservations about the use of full-captioned videos (e.g. Mayer, Lee & Peebles, 2014). CTML, which was originally proposed for learning math and science, stipulates that multimedia instruction should be designed in light of human beings' limited cognitive capacity; otherwise, incoming information that exceeds learners' processing limit will lead to cognitive overload and will thus inhibit learning.

To design multimedia material that does not overburden one's cognitive system, instructors should consider three kinds of cognitive demand (Mayer, 2005). First, intrinsic cognitive load is determined by learners' perceived difficulty or complexity of the learning material, and is thus not directly malleable to the control of the instructor. Extraneous cognitive load, on the other hand, is determined by the design and presentation of the material and is therefore more amenable to the instructor's control. High extraneous load results when the instruction requires learners to simultaneously process a large amount of input presented in different modalities. It interferes with schema acquisition and automation because learners need to devote their cognitive resources to unnecessary processing (Kam, Liu & Tseng, 2020). Lastly, germane cognitive load stems from the mental effort required to make sense of the learning material, and thus contributes to schema acquisition and learning. Given that the three kinds of cognitive load are additive, scholars generally agree with the need to minimize extraneous cognitive load (e.g. Kam *et al.*, 2020; Mayer & Moreno, 2010), but how this goal could be realized in L2 multimedia learning environments has not been thoroughly examined.

Hitherto, in the realm of L2 captioning research, extraneous cognitive load when viewing and understanding multimodal video materials could be reduced at the verbal level by utilizing partial captions – on-screen transcripts of only selected words from the oral discourse (Guillory, 1998; Montero Perez, Peters & Desmet, 2014; Teng, 2019). In contrast to full captions, partial captions can highlight the targeted learning information and direct L2 learners' attention to specific content (Guillory, 1998). Although it is unresolved whether partial captioning can unequivocally lead to superior video comprehension compared to full captioning, some studies, albeit limited in number, have demonstrated the added advantages of partial captions on cognitive load reduction and attention guiding (Rooney, 2014). L2 learners' cognitive load was found to be higher when viewing videos with full captioning than when watching videos with partial captioning (Mohsen & Mahdi, 2021). Furthermore, Mirzaei, Meshgi, Akita and Kawahara (2017) found that presenting only selected words in the captioning line may help avoid L2 learners' over-reliance on caption reading and better prepare them for real-world listening.

Besides the verbal approach, extraneous cognitive load in viewing and understanding multimodal video materials may also be reduced at a nonverbal level. Studies based on native speakers have found that fast-changing images (e.g. animations) would impose greater extraneous load than slow-changing images (Hegarty, 2004; Höffler & Leutner, 2007). To this end, presenting a series of static key frames or images extracted from the animation may help reduce the extraneous processing of transitory pictorial input (Paas, Van Gerven & Wouters, 2007). Similar to partial captions, which present key ideas of an oral discourse, static key frames are key screenshots that present essential pictorial information for understanding the gist of the video. The missing details in the deliberately chosen frames encourage learners to fill in the gaps using their prior knowledge (Hegarty, 1992). From the perspective of CTML, presenting partial captions (the verbal approach) and static key frames (the nonverbal approach) in L2 videos are viewed in this study as two plausible ways to minimize extraneous cognitive load and to promote germane cognitive load.

Nevertheless, in previous CTML studies (e.g. Kam *et al.*, 2020; Lee, Liu & Tseng, 2021), the reduction in extraneous load was mostly discussed at the verbal rather than at the nonverbal level.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

More research is needed to shed light on the optimal strategies for reducing extraneous cognitive load and their potential benefits in L2 multimedia learning scenarios. To address the gap in the research literature, the following research questions were examined:

1. Does the reduction in verbal information through the manipulation of caption presentation modes (i.e. full captions vs. partial captions) affect L2 learners' video comprehension?
2. Does the reduction in nonverbal information through the manipulation of pictorial presentation modes (i.e. animation vs. static key frames) affect L2 learners' video comprehension?
3. How do L2 learners perceive their cognitive load when viewing videos with different verbal supports (i.e. full captions vs. partial captions) and different pictorial content (i.e. animation vs. static key frames)?

2. Literature review

2.1 Theoretical accounts of processing multimodal input in captioned videos

Mayer's (2002, 2005) CTML specifies the cognitive processes of multimedia learning, as shown in Figure 1. Starting from the left side, words and pictures enter through the eyes and ears and are briefly held in sensory memory. With only parts of the information selected into working memory, the learner then draws on long-term memory to make sense of what has been heard or seen by piecing together the attended input and then organizing it into a coherent verbal/pictorial representational model; the resulting model may or may not enter long-term memory, depending on whether it can be integrated with prior knowledge. According to Mayer (2014), the efficiency of the above processes in working memory can be facilitated through the design of multimedia materials, where two assumptions must be considered. One is the *dual-channel* assumption, which indicates humans' use of interconnected channels to process input presented in different modalities. The ideal presentation mode is the distribution of the information across both channels. The second assumption is *limited capacity*, which stresses that each channel has only finite cognitive resources. It explains why viewers have to selectively attend to different parts of multimodal video content at a time, rather than processing the exact copy of the input in working memory (see Hsieh, 2020).

Based on the above-mentioned two assumptions, when a viewer watches a narrated animation, the narration enters the auditory/verbal channel, while the animation enters the visual/pictorial channel. The concurrent activation of the two channels is beneficial for deeper processing. However, when full captions are also presented, they enter the verbal channel as well as the visual channel, introducing more than one source of information to the same channel at the same time.

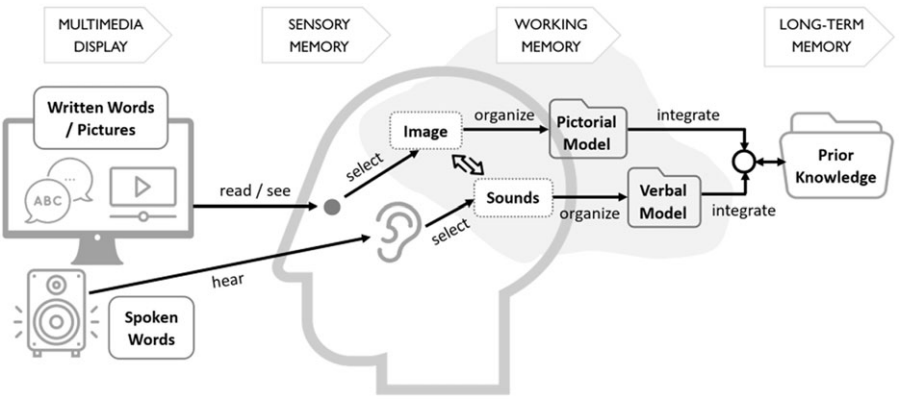


Figure 1. Cognitive processing of multimedia learning

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

In this situation, learners need to reconcile the incoming spoken and written text, and constantly scan the animation to search for images corresponding to the text (Mayer & Moreno, 2010). For L2 learners with lower working memory capacity, reading full captions in the L2 may overburden their working memory capacity and result in less information intake (Kam *et al.*, 2020; Winke *et al.*, 2010; see also Lee *et al.*, 2021).

Despite the fact that full-captioned videos may create redundancy and impose additional cognitive demands on L2 learners, the CTML does not elucidate how to reduce the extraneous load in such cases. Therefore, the current study examined whether a reduction in input presentation would have positive effects on L2 learners' comprehension of the video content and, if so, which level of input reduction (i.e. verbal or pictorial) would be more effective.

2.2 Reducing extraneous cognitive load in L2 captioned videos

2.2.1 Partial captions as the verbal approach

In the extant captioning research, partial captions are usually operationalized in the form of keyword captions. By excluding on-screen text that is less crucial for the video content, keyword or partial captions may prevent visual or verbal channel overload, which is likely to occur in a full-captioning viewing environment. The pioneering research on the use of partial captions by L2 learners was conducted by Guillory (1998). The study revealed that both full and partial captions led to superior comprehension compared with no captions. In particular, partial captions were considered as more effective for video comprehension than full captions due to no significant differences being found. Yet with less text in the visual channel, learners viewing with partial captions may be less prone to cognitive overload.

Supporting evidence regarding partial captions was also provided by Yang, Chang, Lin and Shih (2010), who compared the effects of full captions, captions of only nouns, and captions of only verbs on video comprehension. The result showed that captions comprising only nouns led to comparable comprehension to full captions. However, with captions comprising only verbs, the learners' video comprehension was significantly poorer than that of the other groups. Although Yang *et al.* did not explain what caused the differences between the two types of partial captions, their findings implied the importance of partial selection criteria. A more recent study by Mirzaei *et al.* (2017) investigated the effects of partial and synchronized captions (PSC), which consisted of words less frequently seen in corpora and/or uttered at a high speech rate. Consistent with Guillory (1998) and Yang *et al.* (2010), the result showed that the PSC group performed as well as the full-captioning group in listening comprehension (see also Mohsen & Mahdi, 2021). Moreover, PSC seemed to help reduce the students' reliance on caption reading, thus better preparing them for real-world listening.

Although partial captions were postulated to be as effective as full captions in the aforementioned studies, others have shown contradictory results. For instance, Montero Perez *et al.* (2014) investigated the effects of full, partial, and no captions on global and detailed comprehension. Results revealed no significant differences between the three types of captions in detailed comprehension, but in global comprehension, the full-captioning group significantly outperformed the others. Full captions were thus suggested to be more assistive for L2 learners. However, the authors acknowledged that the lack of inferencing comprehension questions prevented the study from gaining a more complete view of the efficacy of different captions. Similar to Montero Perez *et al.* (2014), Teng (2019) found full captions more beneficial than partial captions for L2 learners' global comprehension, and no significant differences were found in detailed comprehension. The researcher speculated that because of its unusual or incomplete presentation, the partial captions might have drawn so much attention from the viewers that they were not able to holistically interpret the video content.

One factor that may have caused the mixed results in previous studies is the type of selected videos. For example, the videos of situational conversation (in Guillory, 1998) and TED Talks

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

(in Mirzaei *et al.*, 2017) pertain to a more "unidimensional talking-head genre" (Yeldham, 2018: 375), in which the images are less dynamic and it is less crucial to understand the content. Such visual distinction may have affected how viewers allocated their attention, further determining their comprehension outcomes.

2.2.2 Static key frames as the nonverbal approach

Empirical evidence based on young native speakers has shown that narrated animations were more beneficial for story comprehension than audio picture books (Takacs, Swart & Bus, 2015). With dynamic images, animations attract visual attention and provide a closer match between verbal and nonverbal information, which makes the plot more explicit and easier to comprehend (Takacs & Bus, 2016).

Nevertheless, these benefits may only be realized in animations accompanied by spoken text, where the viewer's visual and auditory channels deal with one source of input at a time (Mayer & Moreno, 2010). When other sources of visual information are presented concurrently (e.g. captions), viewers may experience split attention between the text and images. Particularly for novices lacking sufficient background knowledge, the presence of multiple dynamic visuals may cause constant searches for pertinent information, which consume their cognitive resources required for deeper processing (Ayres & Paas, 2007). To minimize the extraneous processing in animated lessons, the nonverbal approach – presenting viewers with only the key segments of a video – may have the potency to reduce cognitive load resulting from constant, transient pictorial displays, and may serve as an attention-getting device to (re)direct learners' focus to key frames that are crucial for meaning interpretation.

In this vein, Moreno (2007) examined the effects of the "segmentation" and "signalling" methods on video-instructed lessons. Through segmentation, the dynamic video content was divided into smaller chunks, so that the viewers only had to process portions of the information at a time. Through signalling, the viewers could see a written list on the screen, which was expected to guide their selection of the most relevant pictorial displays. Results showed that the segmented materials could lead to better performance on retention and transfer tests. However, the signalled video/animation did not benefit the viewers' test performance, and it even imposed higher cognitive load than the original material. The major cause, as pointed out by Moreno, was that the verbal signals may have introduced another source of visual information, which split the viewers' attention between the dynamic video content and the signalling text. Moreno's finding suggests that reduction (of transient pictorial plays) was more helpful than addition (of artificially imposed input enhancement or attention-getting devices) in retaining the processed information in a multimodal environment.

It is feasible that presenting the most representative frames of a portion of a video – static key frames – may merge the benefits of segmenting and attention guiding. Paas *et al.* (2007), for instance, employed static key frames as a follow-up session to an animation-based lesson. After viewing the animation, some of the participants studied all the key frames simultaneously, while the others studied the frames sequentially. No significant between-group differences in comprehension were found; however, the participants exposed to sequential key frames perceived less mental effort, which in turn indicated higher instructional efficiency. In contrast, participants viewing the key frames simultaneously all indicated perceiving a higher cognitive load. Paas *et al.* thus postulated that when the learners must simultaneously pay attention to multiple sources of key information – a processing environment similar to the viewing of animation – extraneous cognitive load would increase. However, when the learners only had to focus on one piece of key information at a time, it required less mental effort and decreased extraneous cognitive load. Given that the research examining L2 learners' processing of dynamic images is still in its infancy, more studies are warranted to explore how the amount of pictorial input affects L2 learners' cognitive load and video comprehension.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

3. Method

3.1 Participants

A total of 62 students majoring in applied English at a public vocational high school in northern Taiwan participated in this study. They were all 12th graders recruited from two intact classes and were all motivated to improve their English proficiency through viewing various kinds of video content online. The two classes had comparable English competence according to their average scores on the achievement tests in the previous academic years. Additionally, based on their performance in a standardized English proficiency test, the participants' English proficiency was considered to be high-intermediate, according to the Common European Framework of Reference for Languages.

3.2 Materials

3.2.1 Video selection

The viewing material, dubbed by a native English speaker, was selected based on the following three criteria. First, the topic of the video was familiar to the participants to prevent comprehension barriers caused by lack of background knowledge (Othman & Vanathas, 2017). Second, to measure the participants' inferential comprehension, the video used in this study did not include materials that completely presented concrete and factual information, as such materials may not be the most feasible genre for measuring inferential comprehension (Montero Perez *et al.*, 2014). Finally, it was ensured that the video matched the participants' language proficiency to ensure that they could manage the intrinsic cognitive load imposed by the material (Martin & Evans, 2018). Accordingly, an animated Ted-Ed video explaining the clustering phenomenon of competing stores was adopted. All participants viewed the video with headphones.

3.2.2 Caption modes

Participants viewed the video with either full or partial captions. With full captioning, there were at most two lines of text presented on screen at the same time; with partial captioning, only the selected words or phrases were shown in the captioning line (see Figure 2). To select the words and phrases in the partial captions, the current study utilized MonkeyLearn, online software that can automatically extract high-frequency words and co-occurring lexical strings from a given text.[1] Additionally, two experienced English teachers were invited to watch the video, read the automatically extracted phrases, and delete semantically redundant words. The final set of keywords and phrases for partial captioning accounted for approximately 45% of the full transcript – close to the 50% caption ratio, which is considered most effective for L2 learners' listening comprehension (Rooney, 2014).

3.2.3 Pictorial modes

The participants watched the video with either animation or static key frames. All videos – whether animation or key-framed – were presented to the participants in high-definition format. To extract the key frames, the present study utilized VLC software, which can automatically

[1]According to Meyer (2011), the cognitive load of multimodal content can be reduced through "explicit material scaffolding" (e.g. showing only novel or unfamiliar terms or vocabulary) or "embedded material scaffolding" (e.g. showing shortened lexical strings consisting of high-frequency keywords that should be known to the viewers). The former scaffolding can be employed when the purpose of multimodal video viewing mainly concerns (explicit) vocabulary learning; the latter scaffold can be utilized when the purpose is pertinent to multimodal comprehension and to activating students' prior knowledge. Both types of scaffolding aim to reduce the cognitive load through reducing the amount of verbal information, but the latter (i.e. focusing on high-frequency words) is more relevant to the context of this study.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

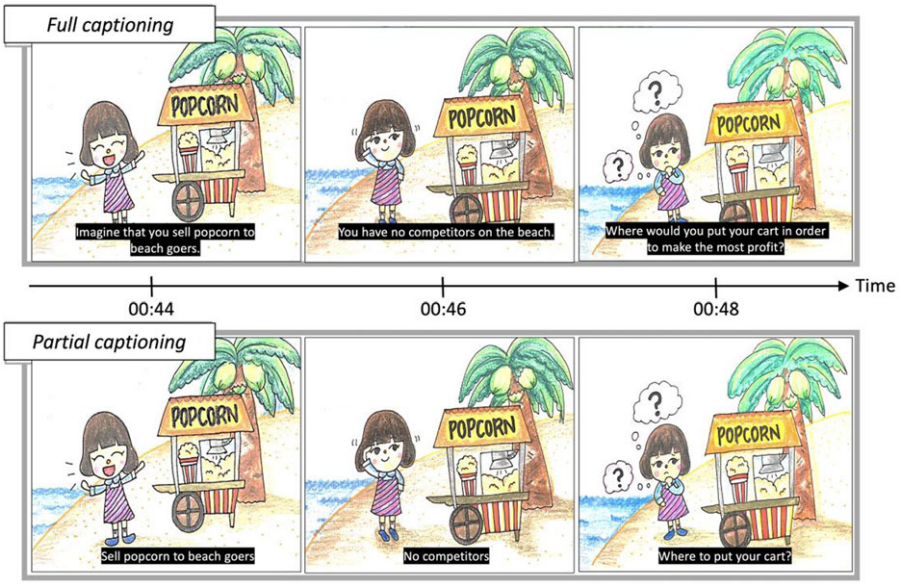


Figure 2. Illustrative examples of the full captions and partial captions utilized in this study

capture a frame at a particular interval. The recording ratio was set to capture one frame every 2 seconds, resulting in 120 frames extracted as candidate key frames. Among them, blurry, incomplete, similar, and repeated images were manually deleted. By doing so, we minimized the amount of redundant pictorial input. There were 50 images reserved as the static key frames. Finally, to create the key-framed videos, the full/partial captions were imported via VLC. The duration of each key frame was reset to ensure that all video materials were identical in length.

3.3 Design

The study used a 2 x 2 factorial design in which a verbal factor (i.e. full captions vs. partial captions) was paired with a nonverbal factor (i.e. animation vs. static key frames). For grouping, a randomized block design was conducted to equally assign the students of Class A and Class B to one of the four viewing conditions. First, the average scores of the students' mid-term and final English examinations in the previous two semesters – including Class A and Class B – were collected. Second, students were numbered according to the ascending order of their average scores (from the lowest to the highest). Based on the numerical order, students were sequentially assigned to the four viewing conditions. Figure 3 visually schematizes the grouping process.

3.4 Instruments

3.4.1 Video comprehension test

This test consisted of 12 multiple-choice questions, among which four items assessed global understanding, another four gauged local understanding (details or understanding of a particular sentence), and the remaining four required inference making (see Appendix in the supplementary material for example items). Each question was followed by four options, with only one correct answer and three distractors. One point was given for each correct answer, so the participants received a maximum of 12 points.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

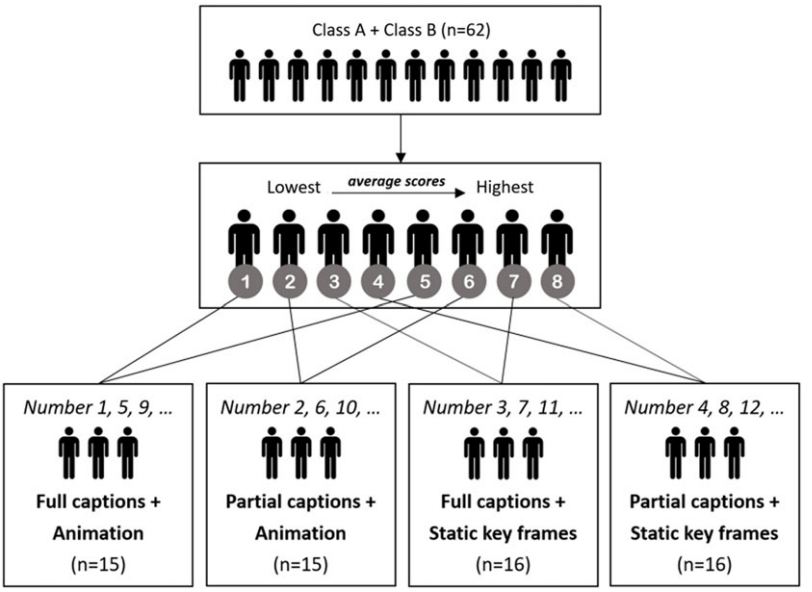


Figure 3. The participants' grouping process

Additionally, to ensure that the participants in the partial information groups were as capable of answering the questions as those viewing with full information, all questions were derived from the information available in the partial captions and key frames. This particular focus on partial information was suggested by Montero Perez *et al.* (2014), who targeted the passages with keywords in the question items, so that participants' comprehension could be solely attributed to the mode of video presentation rather than the amount of information they received.

3.4.2 Cognitive load questionnaire

After the video comprehension test, all participants completed a 9-item questionnaire (see Table 6) to evaluate their cognitive load during the video viewing session. For measuring intrinsic and extraneous cognitive load, six items were adapted from Leppink, Paas, Van der Vleuten, Van Gog and Van Merriënboer's (2013) subjective rating scale, which has been proved to be a valid instrument for measuring the two types of cognitive load (Sweller, van Merriënboer & Paas, 2019). Furthermore, to probe the participants' germane cognitive load, three items developed and validated by Klepsch, Schmitz and Seufert (2017) were adapted; the original questions were slightly rephrased to conform to the context of this study. The participants were required to rate each statement on a 10-point Likert scale (0 = *strongly disagree*; 10 = *strongly agree*).

3.5 Procedure

Prior to the study, all participants were provided with experiences watching full-/partial-captioned videos as well as animated/key-framed videos in their regular class hours. This was to avoid the confounding variables arising from unfamiliarity with certain presentation modes. The experiment began with a 5-minute oral instruction, which informed the participants of their assigned viewing conditions and the overall procedure. Next, the participants watched the system-paced video twice, following the repeated viewing practice in Winke *et al.* (2013) and Teng (2019). Immediately after the two-time viewing, which took approximately 10 minutes, the participants were given 20 minutes to complete the video comprehension test and the questionnaire. All

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

participants were invited to comment on their responses after they had completed the questionnaire (informal post-study interview).

4. Results

4.1 Descriptive statistics

As shown in Table 1, participants viewing full captions + static key frames attained the highest mean score ($M = 8.75$), followed by those viewing partial captions + static key frames ($M = 8.12$), and then partial captions + animation ($M = 7.53$), whereas full captions + animation ($M = 7.47$) led to the lowest score. Besides an overview of the participants' performance in each viewing condition, the results showed a higher mean score of video comprehension for full captions ($M = 8.13$) than for partial captions ($M = 7.84$) in caption mode, and a higher mean score for static key frames ($M = 8.44$) than for animation ($M = 7.50$) in pictorial mode.

Table 2 shows each group's performance on different types of comprehension questions (global, local, and inferential comprehension). The participants viewing the animation video consistently scored lower than their counterparts viewing the key-framed video (probably due to higher cognitive load. Notably, seeing key-framed video content with full captions ("FC+S") appears to lead to the highest scores on both global ($M = 2.93$) and local ($M = 3.06$) comprehension, while for inferential comprehension, seeing the key-framed video with partial captioning ("PC+S") yielded better performance ($M = 2.88$) – a result empirically supporting the assumption of this study that partial display of multimodal input, whether verbal or nonverbal, may promote inference making.

4.2 Two-way ANOVA analysis

The normality of the participants' overall comprehension scores was assessed through the Shapiro–Wilk test, which demonstrated a normal distribution ($W = .962$, $p = .054$) with a significance value greater than 0.05. The Levene's test, an instrument used to assess the variances between two or more groups, also indicated homogeneous performance of the participants in all four groups. As shown in Table 3, the variances were insignificant in all types of comprehension scores. Accordingly, the statistical data fulfilled the prerequisites of normality and homogeneity for further ANOVA analysis.

Table 1. Descriptive statistics of overall comprehension scores

Caption mode	Pictorial mode	*n*	*M*	*SD*
Full captions	Animation	15	7.47	1.81
	Static key frames	16	8.75	1.57
	Total	31	8.13	1.78
Partial captions	Animation	15	7.53	1.19
	Static key frames	16	8.12	2.16
	Total	31	7.84	1.75
Total	Animation	30	7.50	1.50
	Static key frames	32	8.44	1.88
Total		62	7.98	1.76

Note. The maximum score was 12.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

Table 2. Descriptive statistics of different types of comprehension scores

		Global comprehension		Local comprehension		Inferential comprehension	
Viewing condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FC+A	15	2.67	1.05	2.53	0.64	2.33	1.18
FC+S	16	2.93	0.68	3.06	0.68	2.75	1.00
PC+A	15	2.20	0.86	2.73	0.80	2.60	0.99
PC+S	16	2.81	0.98	2.44	0.89	2.88	0.89
Total	62	2.66	0.92	2.69	0.78	2.65	1.01

Note. FC = full captions; A = animation; S = static key frames; PC = partial captions. The maximum score for each type of comprehension question was 4.

Table 3. Levene's test for equality of variances on video comprehension scores

Type of comprehension	*F*	*df1*	*df2*	*p*
Overall	1.916	3	58	.137
Global	1.364	3	58	.263
Local	1.320	3	58	.277
Inferential	.494	3	58	.688

Results of the ANOVA are displayed in Table 4. While caption mode did not cause a significant main effect on any type of comprehension, pictorial mode significantly affected overall comprehension, $F(1, 58) = 4.85$, $p < .05$, with a medium effect size ($\eta_p^2 = .073$). This suggested that the presentation of pictorial mode in the video was particularly effective in terms of promoting the participants' well-rounded performance in the major aspects of comprehension, including global, local, and inferential comprehension, and this was true irrespective of whether the video came with partial or full captions. In addition, the interaction effect between caption and pictorial mode also reached significance in local comprehension, $F(1, 58) = 4.56$, $p < .05$, also with a medium effect size ($\eta_p^2 = .073$). The Bonferroni test, *t* tests performed to reduce the possibility of getting a statistically significant result, was used to further examine this interaction effect. As shown in Table 5, the test indicated that with static key frames, the participants who viewed key-framed video content with full captions scored 0.625 points higher than the other key-framed video group viewing with partial captions ($p < .05$). To put this gain (0.625 points) into perspective, participants were able to enhance their performance by 125% (3.06/4 points vs. 2.44/4 points) on items assessing local comprehension (extracting details) if they were simultaneously exposed to video content consisting of full captions and static key frames (FC+S condition).

4.3 Results of the cognitive load questionnaire

The Cronbach's alphas, a measure of internal reliability between and among a set of test items, indicated high internal consistency for the subscales of intrinsic ($\alpha = .84$), extraneous ($\alpha = .81$), and germane ($\alpha = .82$) cognitive loads. As shown in Table 6, the amount of perceived intrinsic cognitive load (statements 1–3) did not vary considerably among the four groups, all of which showed a similar rating of approximately 4 (out of 10).

Although the participants' responses to statements 1–3 (which deal with intrinsic load) were quite comparable, their average ratings of the statements focusing on extraneous cognitive load

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

Table 4. Two-way ANOVA results of presentation modes on video comprehension

Source	Dependent variable	*df*	*F*	*p*	η_p^2
Caption mode	Overall	1	.406	.526	.007
	Global	1	1.668	.202	.028
	Local	1	1.209	.276	.020
	Inferential	1	.577	.451	.010
Pictorial mode	Overall	1	4.853	.037*	.073
	Global	1	3.719	.059	.060
	Local	1	.364	.548	.006
	Inferential	1	1.800	.185	.030
Caption mode × Pictorial mode	Overall	1	.624	.433	.011
	Global	1	.556	.459	.010
	Local	1	4.556	.037*	.073
	Inferential	1	.075	.784	.001

Table 5. Bonferroni's post hoc analysis of local comprehension scores

Pictorial mode	Caption mode	Caption mode	Mean difference	Std. error	*p*	95% CI	
						LL	*UL*
S	FC	PC	.625	.269	.024*	.087	1.163
	PC	FC	−.625	.269	.024*	−1.163	−.087
A	FC	PC	−.200	.278	.474	−.756	.356
	PC	FC	.200	.278	.474	−.356	.756

Note. CI = confidence interval; LL = lower limit; UL = upper limit; S = static key frames; FC = full captions; PC = partial captions; A = animation.
*$p < .05$.

(statements 4–6) revealed an interesting picture of the role of nonverbal (pictorial) input in the participants' meaning-making process. Statement 4 shows that the participants who viewed key-framed video content with partial captioning (PC+S) tended to agree that their viewing condition was an effective viewing environment for learning about the video content ($M = 7.88$). Notably, those assigned to seeing the key-framed video with full captions (FC+S) exhibited an even stronger tendency ($M = 8.50$) to agree with the potency of their viewing condition, suggesting that reduction in nonverbal (pictorial) input (seeing videos consisting of static key frames) appeared to leverage their attention to what the verbal input (captions) could offer.

However, when asked to only rate the effectiveness of caption mode (statement 5), the participants viewing with animation all considered their caption mode (both full- and partial-captioning) as an ineffective scaffold ($M = 3.20$). This finding corroborated the previous contention that nonverbal pictorial input was a strong attention-getter insofar as having full access to pictorial details (animation) might have distracted their attention from what the captions could offer, as those who did not have access to pictorial details were more likely to see captions as an effective tool ($M = 7.00$ and $M = 6.00$, for the FC+S and PC+S conditions, respectively).

As for the ratings for pictorial mode (statement 6), the participants assigned to the FC+S condition (seeing key-framed video with full captions) tended to *strongly* agree that their viewing

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

Table 6. Descriptive statistics of the cognitive load questionnaire

	Statements	FC+A	FC+S	PC+A	PC+S
		M			
Intrinsic cognitive load	1. The topic (business clustering) covered in the video was very complex.	4.27	4.56	4.33	4.38
	2. The story used to explain the topic was very complex for me.	3.47	3.63	3.40	3.56
	3. The terms used to explain the topic were very complex for me.	4.67	4.88	4.73	4.94
	Total	4.14	4.36	4.15	4.29
Extraneous cognitive load	4. The concurrent presentation of captions and dynamic/static images was effective for learning this topic.	4.54	8.50	5.06	7.88
	5. The full/partial captions in the video were effective for learning this topic.	3.20	7.00	3.20	6.00
	6. The animation/images in the video was/were effective for learning this topic.	2.87	7.50	4.00	6.76
	Total	3.54	7.67	4.09	6.88
Germane cognitive load	7. I made an effort to understand the details and overall context of the video.	7.53	8.19	7.33	8.44
	8. My point while watching the video was to understand everything correctly.	6.87	8.31	7.80	8.06
	9. The video consisted of elements supporting my comprehension of the topic.	7.87	8.19	7.80	8.06
	Total	7.42	8.23	7.64	8.19

Note. FC = full captions; A = animation; S = static key frames; PC = partial captions.

environment was an effective scaffold ($M = 7.50$); those in the PC+S condition (seeing key-framed video with partial captions) did not give such a high rating for the same statement ($M = 6.76$). Both findings from the participants' self-perception data again confirmed that reduction in pictorial images played a prominent role in deciding the participants' perceptions of the usefulness of captions.

It is worth noting that when invited to comment on their responses to statement 6 in the informal post-study interview, more than half the participants assigned to the animation condition indicated that they felt they could not appropriately process all the verbal and nonverbal input because "everything happened so fast" and "they only had a fleeting glimpse of the animation" – indirect evidence that some L2 learners might not have sufficient capacity to process multimodal input simultaneously presented from different channels, in particular the transient input (Kam *et al.*, 2020; Lee *et al.*, 2021).

The last three statements targeting germane cognitive load demonstrate that the static key frames might have induced higher germane cognitive load than the animation, irrespective of the concurrent caption mode. As revealed by statements 7–9, the PC+S and FC+S conditions led to more effort invested in correctly understanding the details and overall content of the video.

5. Discussion

5.1 RQ1: Effects of caption presentation mode

Based on the quantitative data, full captions – regardless of the concurrent pictorial mode – revealed only a marginal advantage on L2 viewers' overall comprehension as compared with partial captions (FC = 8.13 vs. PC = 7.84). This appears to be aligned with the findings of some

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

L2 captioning research (Mirzaei *et al*., 2017; Yang *et al*., 2010). In spite of this marginal difference, we did observe a significant impact of full-captioning video when the purpose of the multimodal learning material targeted local comprehension. In the informal post-study interview, nearly two thirds of the participants viewing the videos with full captions indicated that they preferred full captioning for extracting details, as it provided more semantic and syntactic details not available in partial captioning. One may recall that the participants' local comprehension performance was enhanced by 125% when they viewed the key-framed video that came with full captions. This enhanced gain in this viewing condition is by no means insignificant. Accordingly, when the objective of using multimodal video materials concerns the learning of details, key-framed videos with full captions will be the advised viewing material. This result seems to contradict the original stipulations of CTML (which were proposed for the learning of science and math), namely that full transcription of spoken narrative would impose undesirable cognitive loads on learners. Mayer, Fiorella and Stull (2020) noted that "when learning from video lessons in a second language ... [many CTML] principles are reversed" (p. 848). This phenomenon was also pointed out by Plass and Jones (2005), who observed that full captions may not be redundant for L2 learners, because "reading and listening ..., in many cases one is used as input enhancement for the other" (p. 480). As a result, for L2 learners, any meaningful verbal input in the full captions, including conjunctions omitted in the partial captions, may have potential value for providing more detailed, comprehensible understanding.

Note that this study found that full captioning, which was found to be beneficial for the participants' local comprehension (getting the details), was not as effective in terms of promoting their *overall* comprehension; in this study, the participants seemed to rely heavily on nonverbal pictorial input – in particular, the static key frames – for all aspects of comprehension (as borne out by the main effect of the pictorial mode in the ANOVA). This observation is generally aligned with Tragant and Pellicer-Sánchez's (2019) eye-tracking study, which also deals with L2 learners of a similar (high-intermediate) proficiency profile. Tragant and Pellicer-Sánchez found that although their participants could exploit a variety of meaning-making strategies to extract the gist of multimodal content, they seemed to rely more on the pictorial input – which was probably more effective in terms of providing a coherent gist of the video content – than on text (captions) for overall comprehension.

5.2 RQ2: Effects of pictorial presentation mode

While the ANOVA analysis did not detect any main effect from the caption mode on the participants' overall comprehension, it did detect a main effect arising from the pictorial mode. The above findings suggest that the (nonverbal) pictorial presentation mode played a more prominent role than the (verbal) caption presentation mode in enhancing the participants' well-rounded performance in global, local, and inferential comprehension.

Note that this does not unequivocally endorse the use of all types of pictorial input. One may recall that the ANOVA also showed that the participants were better able to take advantage of what full captioning could offer in extracting details (local comprehension) only when it was accompanied by static key frames (Table 4), but such an advantage disappeared when full captioning was accompanied by animation. It was possible that pictorial input was more effective than verbal (captions) input in terms of attracting L2 learners' attention (see Tragant & Pellicer-Sánchez, 2019); in this case, the transiency of fast-changing dynamic animation might have imposed a high cognitive load that stopped the participants – in particular those who did not have sufficient working memory capacity – from taking full advantage of both the verbal and nonverbal input due to inherent cognitive constraint (see Kam *et al*., 2020; Lee *et al*., 2021). In contrast, the key-framed video, which was less cognitively demanding, allowed the participants to process the nonverbal (pictorial) input without taxing their mental resources, which in turn enabled them to further process concurrent verbal (captions) input. Accordingly, key-framed

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

video content may be more helpful than animation for promoting L2 learners' overall comprehension. This speculation is also corroborated by the participants' qualitative reports; one may recall that the participants' qualitative comments in the post-study interview also revealed that reduction in pictorial input (seeing key-framed video content) determined the participants' perceptions of their capacity to handle the verbal input.

The above contention does not mean that reduction in both nonverbal (pictorial) input and verbal input presents the most desirable learning environment in all cases. Specifically, we observed that the static key frames led to the lowest score in local comprehension when the participants were presented with key-framed video content with partial captions. The participants' less desirable performance after viewing key-framed videos with partial captions might be attributed to the simultaneous reduction in both verbal and pictorial input, which might have discarded too much key information from the video content. In this regard, the information was probably too fragmented for the participants to obtain the details of the video content (Teng, 2019).

Taken together, the above discussion lends some support to the effectiveness of key-framed videos in terms of overall comprehension (Table 1); however, if the learning purpose is concerned with extracting details, the key-framed video should be supplied with full captions. Reduction in both verbal and nonverbal (pictorial) input is probably only desirable when the objective of a class is concerned with inference making; as can be seen in Table 2, the participants' inferencing performance was the best when they viewed the key-framed video with partial captioning. With the help from software such as VLC, the extraction of static key frames can be made easy, without too much manual work from the instructor.

5.3 RQ3: Perceived cognitive load under different presentation modes

The last research question sheds light on the intrinsic, extraneous, and germane cognitive loads perceived by the participants in different viewing conditions. As reflected by the very close ratings on intrinsic cognitive load, the difficulty level of the video content might have been equivalent for all participants. The similarly perceived intrinsic load also indicated that the participants could maintain comparative cognitive capacity for dealing with extraneous and germane processing, which showed a larger discrepancy among the ratings of the four viewing groups.

For the perceived extraneous cognitive load, the participants' responses are generally consistent with the discussion based on the participants' quantitative performance data. First, the high ratings from participants assigned to the FC+S ($M = 8.50$) and PC+S conditions ($M = 7.88$) for statement 4 showed that overall static key frames were deemed more useful than animated content for promoting their learning ($M = 4.54$ and $M = 5.06$ for the FC+A and PC+A conditions, respectively). This pattern is also replicated in the participants' responses to statement 6, which examined their perceptions of the role of animation/images in learning about a given topic. As noted earlier, several participants who viewed the animated video content did not find the transient nature of the pictorial content helpful, especially when they intended to process textual (captions) and pictorial input at the same time. The most interesting picture lies in the participants' responses to statement 5, where they indicated that the usefulness of captions (full and partial) only became manifest when they were accompanied by static frames (rather than animation). This again suggests that pictorial mode mediates the effect of captioning mode, and that the potency of captions is manifested when the participants' attention is not taxed.

With regard to germane cognitive load, the questionnaire result indicated that the participants in the key-framing groups perceived higher germane cognitive load than those in the animation groups. In contrast, captioning modes did not seem to be the main factor that determined the participants' germane processing. The participants' ratings also concurred with their inferential comprehension outcome, where the static key frames led to higher scores (Tables 3 and 5). The result lends support to the theoretical assumption of CTML that more germane cognitive

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

load can elicit more cognitive efforts on deeper processing, including inference making based on available information.

6. Pedagogical implications

The key-framed video yielded better overall comprehension with full captions, as revealed in this study. Similarly, if the pictorial backdrop of a video is less dynamic and rather monotonous (e.g. TED Talks and slow-paced documentaries), full captions would be cognitively manageable and more supportive for L2 viewers. In particular, for educational programs that often need to build up linguistic scaffolds for L2 learners within limited instructional hours, such as remedial courses and content and language integrated learning lessons, key-framed videos with full captions can be a useful multimedia learning material. When the instructional goal of a program aims to promote learners' overall comprehension, the results of this study suggest that key-framed videos or equivalent pictorial presentations would be more desirable; key-framed videos, in particular those that are accompanied by full captioning, would be helpful for promoting learners' balanced performance in all aspects of comprehension. With the slow-changing backdrop of the video, L2 (remedial) learners can focus on the captions in a more consistent manner, which may help them to process the text more deeply and achieve better content learning. Notably, as commented by the participants, such pictorial reduction via static key frames was preferable to the verbal reduction via partial captions.

While employing key-framed videos, L2 instructors can enhance students' extraction of details through the inclusion of full captions. It may involve a manual process, but if this can significantly enhance learners' performance in extracting and retaining more details, this additional effort, which can be made easy through the use of VLC or MonkeyLearn, is probably worth considering.

Finally, if the objective of a class is concerned with inferential comprehension or learning (e.g. speculating about the speakers' attitudes and predicting the subsequent development), instructors may want to use key-framed video materials with partial captions. The reduction in both verbal and nonverbal input may create desirable difficulty for L2 learners to generate reasonable inferences and guesses and fill in the comprehension gaps. Knowing how learners should be benefited through different verbal and nonverbal input presentation modes will help L2 instructors better prepare multimedia materials for different comprehension purposes.

7. Conclusions and limitations

Since multimedia learning requires very complex mental processes, the results of this one-time experiment based on manipulating a captioned animation should be interpreted with care. The major limitation is that this study focused only on the effects of animation and captions (and their reduced modes), so other types of multimedia presentation can be adopted. For nonverbal display, videos that include real-life scenarios may not draw the viewer's attention in the same way as animation and static key frames do. Likewise, other types of textual support, such as annotations and bilingual subtitles, may activate different verbal processing strategies. Future research can thus shed light on the efficacy of these verbal/nonverbal elements commonly seen in videos.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344022000088

Ethical statement and competing interests. The authors declare that there is no conflict of interest. The participants were informed about the experiment details and agreed to participate without any duress. All participants understood their rights, and their participation consent was obtained before this study. Results of experiments were immediately substituted with ID codes after the information was analyzed. Confidentiality of data is thus ensured.

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

References

Ayres, P. & Paas, F. (2007) Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology*, 21(6): 811–820. https://doi.org/10.1002/acp.1351

Guillory, H. G. (1998) The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal*, 15(1–3): 89–108. https://doi.org/10.1558/cj.v15i1-3.89-108

Hegarty, M. (1992) Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5): 1084–1102. https://doi.org/10.1037/0278-7393.18.5.1084

Hegarty, M. (2004) Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction*, 14(3): 343–351. https://doi.org/10.1016/j.learninstruc.2004.06.007

Höffler, T. N. & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722–738.

Hsieh, Y. (2020) Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 33(5–6): 567–589. https://doi.org/10.1080/09588221.2019.1577898

Kam, E. F., Liu, Y.-T. & Tseng, W.-T. (2020) Effects of modality preference and working memory capacity on captioned videos in enhancing L2 listening outcomes. *ReCALL*, 32(2): 213–230. https://doi.org/10.1017/S0958344020000014

Klepsch, M., Schmitz, F. & Seufert, T. (2017) Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8: 1–18. https://doi.org/10.3389/fpsyg.2017.01997

Lee, P. J., Liu, Y.-T. & Tseng, W.-T. (2021) One size fits all? In search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, 25(3): 400–430. https://doi.org/10.1177/1362168819856451

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T. & Van Merriënboer, J. J. G. (2013) Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4): 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Martin, A. J. & Evans, P. (2018) Load reduction instruction: Exploring a framework that assesses explicit instruction through to independent learning. *Teaching and Teacher Education*, 73: 203–214. https://doi.org/10.1016/j.tate.2018.03.018

Mayer, R. E. (2002) Multimedia learning. In Ross, B. H. (ed.), *Psychology of learning and motivation* (Vol. 41). Academic Press, 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2005) Cognitive theory of multimedia learning. In Mayer, R. E. (ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, 31–48. https://doi.org/10.1017/CBO9780511816819.004

Mayer, R. E. (2014) Cognitive theory of multimedia learning. In Mayer, R. E. (ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, 43–71. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., Fiorella, L. & Stull, A. (2020) Five ways to increase the effectiveness of instructional video. *Educational Technology Research & Development*, 68(3): 837–852. https://doi.org/10.1007/s11423-020-09749-6

Mayer, R. E., Lee, H. & Peebles, A. (2014) Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28(5): 653–660. https://doi.org/10.1002/acp.3050

Mayer, R. E. & Moreno, R. (2010) Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In Plass, J. L., Moreno, R. & Brünken, R. (eds.), *Cognitive load theory*. Cambridge University Press, 131–152. https://doi.org/10.1017/CBO9780511844744.009

Meyer, O. (2011) Introducing the CLIL-pyramid: Key strategies and principles for quality CLIL planning and teaching. In Eisenmann, M. & Summer, T. (eds.), *Basic issues in EFL teaching and learning*. Heidelberg: Universitätsverlag Winter, 295–313.

Mirzaei, M. S., Meshgi, K., Akita, Y. & Kawahara, T. (2017) Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL*, 29(2): 178–199. https://doi.org/10.1017/S0958344017000039

Mohsen, M. A. & Mahdi, H. S. (2021) Partial versus full captioning mode to improve L2 vocabulary acquisition in a mobile-assisted language learning setting: Words pronunciation domain. *Journal of Computing in Higher Education*, 33(2): 524–543. https://doi.org/10.1007/s12528-021-09276-0

Montero Perez, M., Peters, E. & Desmet, P. (2014) Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, 26(1): 21–43. https://doi.org/10.1017/S0958344013000256

Montero Perez, M., Van Den Noortgate, W. & Desmet, P. (2013) Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3): 720–739. https://doi.org/10.1016/j.system.2013.07.013

Moreno, R. (2007) Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology*, 21(6): 765–781. https://doi.org/10.1002/acp.1348

Othman, J. & Vanathas, C. (2017) Topic familiarity and its influence on listening comprehension. *The English Teacher*, 34: 19–32.

Paas, F., Van Gerven, P. W. M. & Wouters, P. (2007) Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology*, 21(6): 783–793. https://doi.org/10.1002/acp.1349

Plass, J. L. & Jones, L. C. (2005) Multimedia learning in second language acquisition. In Mayer, R. E. (ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press, 467–488. https://doi.org/10.1017/CBO9780511816819.030

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press

Rooney, K. (2014) The impact of keyword caption ratio on foreign language listening comprehension. *International Journal of Computer-Assisted Language Learning and Teaching*, 4(2): 11–28. https://doi.org/10.4018/ijcallt.2014040102
Sweller, J., van Merriënboer, J. J. G. & Paas, F. (2019) Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2): 261–292. https://doi.org/10.1007/s10648-019-09465-5
Takacs, Z. K. & Bus, A. G. (2016) Benefits of motion in animated storybooks for children's visual attention and story comprehension. An eye-tracking study. *Frontiers in Psychology*, 7: 1–12. https://doi.org/10.3389/fpsyg.2016.01591
Takacs, Z. K., Swart, E. K. & Bus, A. G. (2015) Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research*, 85(4): 698–739. https://doi.org/10.3102/0034654314566989
Teng, F. (2019) Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos. *Computer Assisted Language Learning*, 32(7): 665–691. https://doi.org/10.1080/09588221.2018.1532912
Tragant, E. & Pellicer-Sánchez, A. (2019) Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, 80: 212–223. https://doi.org/10.1016/j.system.2018.12.002
Winke, P., Gass, S. & Sydorenko, T. (2010) The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1): 65–86.
Winke, P., Gass, S. & Sydorenko, T. (2013) Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1): 254–275. https://doi.org/10.1111/j.1540-4781.2013.01432.x
Yang, J. C., Chang, C. L., Lin, Y. L. & Shih, M. J. A. (2010, November) A study of the POS keyword caption effect on listening comprehension. In Wong, S. L., Kong, S. C. & Yu, F.-Y. (eds.), *Proceedings of the 18th International Conference on Computers in Education (ICCE 2010)*. Putrajaya: Asia-Pacific Society for Computers in Education, 708–712.
Yeldham, M. (2018) Viewing L2 captioned videos: What's in it for the listener? *Computer Assisted Language Learning*, 31(4): 367–389. https://doi.org/10.1080/09588221.2017.1406956

About the authors

Chen Chi is an MA student in the Department of English at National Taiwan Normal University. Her research focuses on technology-enhanced language learning. She can be contacted at National Taiwan Normal University, No. 162, Sec. 1, HePing E. Rd., Da'an Dist., Taipei 106, Taiwan.

Hao-Jan Howard Chen is a professor in the Department of English at National Taiwan Normal University. His research revolves around computer-assisted language teaching and learning. He can be contacted at National Taiwan Normal University, No. 162, Sec. 1, HePing E. Rd., Da'an Dist., Taipei 106, Taiwan.

Wen-Ta Tseng is a professor in the Department of Applied Foreign Languages at National Taiwan University of Science and Technology. His research focuses on second/foreign language assessment, language learning strategies, and quantitative research methods. He can be contacted at National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd., Da'an Dist., Taipei City 106, Taiwan.

Yeu-Ting Liu is the corresponding author of this paper. He is a professor in the Department of English at National Taiwan Normal University. His research focuses on bilingual processing and cognitive development of second language learners. He can be contacted at National Taiwan Normal University, No. 162, Sec. 1, HePing E. Rd., Da'an Dist., Taipei 106, Taiwan.

Author ORCiD. Chen Chi, https://orcid.org/0000-0001-7757-721X
Author ORCiD. Hao-Jan Howard Chen, https://orcid.org/0000-0002-8943-5689
Author ORCiD. Wen-Ta Tseng, https://orcid.org/0000-0001-5673-5944
Author ORCiD. Yeu-Ting Liu, https://orcid.org/0000-0001-8055-0587

https://doi.org/10.1017/S0958344022000088 Published online by Cambridge University Press