The why-what-when-who-how of using captioned videos as an instructional aid in EAL classrooms: Theoretical perspectives and classroom implications

Trinh Thai Van Phuc[1*]

[1]Simon Fraser University, Canada
[*]Corresponding author: phuc_trinh@sfu.ca

ARTICLE INFO

DOI:10.46223/HCMCOUJS.soci.en.12.1.2185.2022

Received: February 13th, 2022

Revised: March 28th, 2022

Accepted: April 18th, 2022

Keywords:

captioned videos; English as an Additional Language (EAL); multimedia learning; subtitled videos

ABSTRACT

This paper seeks to address how captioned videos are used as an instructional aid in classrooms of English as an additional language (EAL). The paper employs a narrative literature review to synthesize prior studies and reach pedagogical conclusions for utilizing the captioned videos. The results argue that captioned videos have pedagogical, cognitive, and psychological benefits to language teaching and learning with the awareness of some influential factors during instructional implementations. First, the paper elaborates on the roles of videos in language teaching and learning (Why), underlying theories of employing videos (How), and influential factors of utilizing the captions (When and Who), followed by an emphasis on captioning effects on language learning (What). Then, it discusses some classroom implications as concluding remarks.

1. Introduction

Videos are employed most frequently as one of the media tools in many classroom contexts (Merkt, Weigand, Heier, & Schwan, 2011). According to Berk (2009), many types of videos are utilized in classroom contexts. They are various from drama, comedy, and actions to documentaries, TV programs, and self-made video projects. Moreover, the availability and widespread interest of numerous online video clips from popular websites have notably brought multimedia materials into the context of language classrooms (Berk, 2009; Watkins & Wilkins, 2011). Importantly, a number of decades have witnessed the use of videos and audiovisual media as a main scaffolding tool in a language classroom (Danan, 2004; Herrero & Vanderschelden, 2019; Vanderplank, 2016).

Particularly, captioned videos are representations of texts and images in which texts are delivered in both auditory channels and visual ones. That means they are resources for three channels of input consisting of the visual pictures, second language (L2) sounds, and L2 on-screen texts. In other words, they consist of sounds and texts being delivered in the same L2. Captions are different from subtitles on account of languages used in a channel of the on-screen texts (D'Ydewalle & Gielen, 1992). Captioned videos are given other names such as "bimodal subtitling" (Markham & Peter, 2003, p. 332, as cited in Yang, 2020), Teletext subtitles, native-language subtitles, and intralingual subtitles (Yang, 2020). In contrast, standard subtitles consist of L2 sounds with L1 texts, whereas the reversed subtitles include L1 sounds with L2 texts.

In Vietnam, C.-D. Nguyen and Boers (2019) conducted a quasi-experimental research design with the employment of TED Talk videos to investigate the English vocabulary uptake of undergraduate students. The findings highlighted the great potential of utilizing audiovisual input in EAL classrooms. Acknowledging the ubiquitous use of captioned videos as an instructional aid in Vietnamese EAL classrooms, H. D. Nguyen (2020) investigated the listening strategies of non-English-major students in watching captioned videos and reported that cognitive strategies were most frequently employed among students. In a similar context of tertiary education instruction, films were used as an innovative tool to enhance the intercultural competence of Vietnamese students (Truong & Tran, 2014). Personally, the author has witnessed significant performance improvements from one of his EAL learners, who spends time extensively watching captioned videos and movies. Also, as an academic manager with the main role of instructing, writing, and developing the curriculum at a public university in Ho Chi Minh City, the author has seen the proliferation of using videos as a learning unit in many commonly used English textbooks in Vietnam. They are popular products of well-known publishers such as the Cambridge English Empower series by Cambridge University Press, Navigate coursebook series by Oxford University Press, Life textbooks series by Cengage Learning, Speakout book series by Pearson English, and Breakthrough Plus by Macmillan Education. Indeed, videos are usually employed in the author's classrooms since they have authentic linguistic inputs and are effective tools to deliver cross-cultural content for his students. Consequently, videos attract the author with their appealing characteristics and encourage his inquiring mind to a narrative review of the literature.

This study seeks to review the roles of using videos as an instructional aid, its theoretical underpinnings, influential factors, and their effects on language learning and development. Consequently, the paper will use a narrative literature review method, which synthesizes previous research findings and provides readers with an overview of the topic (Green, Johnson, & Adams, 2006). The author utilizes electronic search engines to find linguistic-related studies and educational-subjected journal articles to conduct this review paper. 'Captioned videos in English language classrooms', 'subtitled videos', and 'multimedia learning' are mainly used as keywords for searching. Most of the references have been downloaded from Simon Fraser University library databases within the last twenty-five years.

The guiding research questions for this review are:

1/ Why are captioned videos employed as an instructional aid in EAL classrooms?

2/ How are they processed?

3/ When are they utilized, and who are suitable for learning with the captioned ones?

4/ What effects do they have on the development of language aspects and skills?

In what follows, the paper begins with the elaboration on the why-element, and the roles of videos in language teaching and learning.

2. Roles of videos in EAL classrooms (Why-element)

Like two sides of a coin, employing videos in an EAL classroom has its own advantages and disadvantages. On the one hand, videos are largely used for illustration rather than instruction (Schulte, 1991). According to Chandler and Sweller (1991), the utilization of multiple input modes of sounds, texts, and images split listeners' attention and make working memory (WB) overloaded, so information intake and comprehension are not processed well. In a similar sense, the use of on-screen captions in a fast-paced video of medical reactions does not help to improve the performance and comprehension of non-native American college students because of the cognitive

overloading (Mayer, Lee, & Peebles, 2014). In addition, videos' benefits for the whole of language learning are probably not reviewed thoroughly and theoretically (Fehlman, 1996). Moreover, Mayer and Moreno (1998) revealed that learning ability is diminished due to redundant representations of images, captions, and sounds in which sounds are read rather than heard with images. Similarly, Danan (2016) criticized visual texts as distractions if learners focus on the onscreen texts instead of listening. Besides, Weyers (1999) highlighted the benefits of videos for reading comprehension are not studied much. Last but not least, since learning requires active participation and engagement, watching videos is, however, probably a passive experience (Stigler, Geller, & Givvin, 2015).

On the other hand, videos are employed frequently in EAL language classrooms because of their numerous facilitative roles. Firstly, they have some pedagogical benefits in an EAL classroom. That is, they can be used in a video-based curriculum (Herron, Corrie, Cole, & Dubreil, 1999). More specifically, they are utilized frequently in a communicative classroom (Ciccone, 1995), a task-based one (Fukushima, 2002), content and language integrated learning one (Cinganotto & Cuccurullo, 2015), and a project-based one (Spring, 2020). In addition, the use of videos facilitates skill-learning integration and enriches instruction (Hanley, Herron, & Cole, 1995). They are a productive tool to develop learners' awareness of intercultural communication (Ito, 1996; Truong & Tran, 2014) and of their own error self-correction (Hanson-Smith, 2007, as cited in Toohey, Dagenais, & Schulze, 2012), and improve learners' foreign cultural knowledge and skills (Herron, Dubreil, Corrie, & Cole, 2002). Moreover, they provide contexts and discourses for language learning (Bal-Gezegin, 2014) and real-like language lessons with authentic language input (Flynn, 1988; Sherman, 2003). Besides, paralinguistic features such as gestures, voices, and facial expressions are instructional represented in videos (Stempleski & Arcario, 1992). Last but not least, they are pedagogically employed as an assessment instrument (Coniam, 2001), an instructional aid for the deaf and learners having hearing difficulties (Vanderplank, 2016), and a profoundly effective tool for language learning in informal contexts (Vanderplank, 2019).

Secondly, the use of videos also has cognitive and psychological benefits (Yang, 2020). Cognitively, language processing is more easily accessed on account of the facilitation and benefits of language chunks in visual texts (Vanderplank, 1990). In this sense, the utilization of videos enriches students' understanding and remembering more greatly than that of only print materials or auditory texts (Gardner, 2006, as cited in Yang, 2020). In addition, Berk (2009) revealed that employing videos enhances the retention of learners' visual cognition since ideas and concepts are illustrated with images and pictures rather than texts. In the same vein, the use of videos helps to enhance information retention and motivate language learners (Sherman, 2003). Psychologically, the use of interesting videos with meaningful contexts strengthens learners' motivation twice as much as written words (Kirk, 1992, as cited in Yang, 2020). As Bird (2005, as cited in Vanderplank, 2016, p. 24) showed, the employment of videos can be named an "edutainment system" since learning and watching for leisure happen concurrently. In brief, the use of videos in EAL classrooms has pedagogical, cognitive, and psychological benefits to language learning and teaching. As a consequence, their advantageous roles outweigh the disadvantages. In the following session, the how-element is presented.

3. Underlying theories of using videos (How-element)

These above-mentioned definitions of captioned videos illustrate an underlying theory of bi-channel learning by Hartman (1961). That is, learning is reinforced and facilitated when relevant information is illustrated redundantly between the two channels of sounds and visuals. This theory highlights the beneficial roles and the growing number of relevant cues (sounds, texts, images)

reinforcing one another to promote learning and enhance comprehension.

Mayer (2009) highlighted that ones learn better when information is presented with words and visuals rather than with words separately. That "learning from words and pictures" represents multimedia learning (Mayer, 2009, p. 03). He further elaborated on the parts of speech if multimedia is referred to as a noun or an adjective. When viewing it as a noun, it means a technological device to present words and visuals, while an adjective refers to three contexts. That is, (i) multimedia learning illustrates the learning from two modes (words and visuals), while (ii) multimedia messages are presentations of these modes, and (iii) multimedia instructions refer to these presentations to promote learning. In his premises, multimedia messages are viewed in three categories, namely "the delivery media (e.g., amplified speaker and computer screen), presentation mode (e.g., words and pictures), or sensory modalities (e.g., auditory and visual)" (Mayer, 2009, p. 03). The first category focuses on the technological device, while the remaining ones accentuate the learners' cognitive information processing. The differences between the second category and the third one are (i) separate systems for processing information and (ii) chronological processing sequence in learners' cognitive systems. That means the presentation modes refer to the processing of "verbal and pictorial knowledge" in a later sequence of the cognitive process, while the sensory ones refer to the "auditory and visual processing" early happening in the cognitive system (Mayer, 2009, p. 09).

Multimedia learning has been illustrated with various presentations, including a study of multimedia support and vocabulary learning (Chun & Plass, 1996), L2 reading on the web (Chun, 2001), and captioning effects of television programs on EFL learners' comprehension (Rodgers & Webb, 2017). Consequently, the multimedia messages and instructions represent the employment of captioned videos since they consist of sounds, texts, and images in alignment with the fast development of technological devices as a media delivery channel. Further, captions are growing globally because of the development of learners' virtual communities and multimedia (Vanderplank, 2016).

Norton and Toohey (2004) revealed that the meaning-making process needs to be reconceptualized since language learning closely correlates with social change. In other words, texts to facilitate this process do not predominantly exist in linguistic modes, but they can be illustrated in other modes such as gestural and visual ones. In this sense, Stein (2004, p. 98) advocated the assumption of multimodality theory in which "meaning-making is multimodal" with different modes of communication (i.e., texts, visuals, sounds) being connected and working together to reinforce meaning-making. In a similar vein, Zhang (2016) underscored the importance of three variables affecting the meaning-making process, namely (i) characteristics of texts, (ii) characteristics of viewers or readers, and (iii) social contexts. Also, multimodal meaning-making consists of other representations of the body and mind, such as emotions, feelings, anticipations, etc. (Kramsch, 2009). As a result, multimodal communications lead to a pedagogical shift from language teaching to instructions of multiple modes to maximize learning (Stein, 2004). In a study by Toohey et al. (2012), video-making projects are employed as a tool to develop English language learners' multimodal literacies and contribute to second language learning. Similarly, Choi and Yi (2016), further revealed that multimodal projects and practices are mainly employed "in the form of videos and webpages" (p. 310). Indeed, video games and captioned videos are viewed as a multimodal platform for learning (Dressman, 2020).

In brief, the widespread interest and exponential growth for audiovisual materials with the delivery of media have made a closer connection and strong integration between multimedia learning and multimodality practices. Indeed, "multimodality is directly relevant to pedagogical computer-mediated communications in many ways" (Kern, Ware, & Warschauer, 2016, p. 550).

The employment of online and offline videos to improve language learning is critically reflected and strongly connected in three learning theories of bi-channel, multimedia, and multimodality. The next part of this paper seeks to present the influential factors of using the videos and its effects on language learning.

4. Influential factors of using captions (When-and-Who elements)

This part illustrates when videos are employed and who is suitable for learning with videos. There are numerous reports about the influential factors of using captioned videos. Probably, captions were first employed as a tool to enhance comprehension for deaf and hearing-impaired individuals in the early 1990s; then they became a productive tool for normal hearers in their additional language learning (Vanderplank, 2016).

Winke, Gass, and Sydorenko (2010) investigated how twenty-six English-speaking learners (of Arabic, Chinese, Russian, and Spanish) and their proficiency levels have an influence on their L2 learning in repeated caption viewings. The investigation used an eye-tracking methodology, which is supposed to give a complete record of learners' cognitive behaviors of textual viewing. The results showed that watching captions twice was most beneficial and effective for language learning. Interestingly, when only watching captions once, the learners of Russian and Spanish found watching captions first more beneficial whereas the learners of Chinese and Arabic performed better when watching captions second. This indicates that the L1-L2 orthographic differences result in the effective sequences of watching captions. In other words, learners with L1-L2 orthographic similarity utilize captions first, whereas those with L1-L2 orthographic dissimilarity employ sounds first.

Further exploring influential factors of using captions, Winke, Gass, and Sydorenko (2013) studied how orthography differences and content familiarity affected caption viewings. The authors investigated thirty-three English-speaking sophomore students (of Arabic, Chinese, Russian, and Spanish) in the USA. They employed two three-to-five-minute videos about different animal documentaries in which salmons are familiar to the learners whereas bears are unfamiliar to them. The findings first showed that the participants spent 68 percent of the total on-screen time on reading captions. Second, Russian and Spanish learners spent less time on captions than Arabic ones due to the differences in L2 orthography. Additionally, learners of Chinese found reading Chinese captions challenging and demanding for their eye fixations. Third, only learners of Chinese spent more time reading captions in the unfamiliar context, while there are no differences in the other groups of learners. The findings highlighted that those learners experienced cognitive overload when they encountered unfamiliar contents and shallow depth of lexical knowledge. The onscreen images and texts in the visual working memory and sounds in the auditory working memory were not reinforced one to promote information intake because learners' attention between the two visual modes was split. The authors elaborated that captions significantly facilitated the process of listening in many cases where L2 captions were similar to L1 orthography.

Another influential factor for caption viewing is speed. Kruger (2013) reported that the increasing speed of the captions led to the shifting focus on lecturers' presentations of pictures and slides. In the same vein, Mayer et al. (2014) shed light on a cognitive load theory that comprehension of non-native English speakers was enhanced when adding a video representation to a slow-paced narration about wildlife documentaries. In contrast, adding captions in a fast-paced video about chemical reactions did not support learners' performance since their cognitive processing was overloaded, and captions were unnecessarily redundant.

Muñoz (2017) examined the influences of forty ESL learners' age and proficiency on watching videos with Spanish L1 subtitles and English L2 captions. The study was also conducted by the eye-tracking methodology. She categorized the participants into three age groups (namely, nineteen children ranging from 10 to 12.7 years old, nine adolescents from 13.1 to 16, and twelve adults from 19 to 41) with six proficiency levels from A1 to C2 according to the Common European Framework of Reference (CEFR). There are 18 participants in the elementary level group (A1 and A2 levels), 12 participants in the intermediate one (B1-B2 levels), and 12 participants in the advanced one (C1 and C2 levels). The findings showed that the children group utilized the subtitles much more than the other groups and spent more time reading L2 captions than L1 subtitles. Interestingly, intermediate and advanced learners skipped more L1 subtitles than L2 captions. The findings further explained that captions were challenging tasks for elementary learners and children. It took longer periods of time for them to comprehend captions. The results shed light on the various influences of captions on different learners from different age groups and proficiency levels.

In a related study, Du (2015, as cited in Yeldham, 2018) used another technique called paused transcriptions to check whether the participants watching captioned videos are listening or reading after an intermittent pause. This technique as well as eye-tracking are viewed as a direct tapping of the real-time cognitive processing of learners (Yeldham, 2018). The participants are upper-intermediate Taiwanese learners of English. The findings revealed that the learners paid more attention to sounds than texts. This indicated that the higher proficiency levels the learners reach, the more auditory mode is processed. Similarly, other studies by Taylor (2005) and Pujola (2002, as cited in Yeldham (2018)) used other instruments such as interviews, observations, verbal reports, questionnaires, and written reflections. They investigated multi-proficiency learners using captions for comprehension. The results showed that less-proficient learners utilized captions as a facilitative tool for their comprehension. Further, Taylor (2005) elaborated that beginning proficiency learners found multiple inputs of sounds, texts, and visuals demanding for cognitive processing. In the same vein, low-proficient learners generally employed captions while high-proficient ones used multiple modes of sounds, visuals, and captions more fully (Gass, Winke, Isbell, & Ahn, 2019).

In a recent study by Yang (2020), he investigated the influences of various YouTube visual text support on fourteen American college students learning the upper-intermediate and advanced German language. Four kinds of videos are employed in three experiments, namely (i) no captions/ no subtitles, (ii) English L1 subtitles, (iii) German L2 captions, and (iv) annotated captions (i.e., L2 captions with some L1 translations of key new words). The findings showed that 57 percent of students preferred reading captions to reading subtitles. Around 80 percent considered captions helpful resources for learning German. Subtitles were beneficial for content comprehension but failed to support and enhance their L2 learning. The most distinguishing results in this study compared to previous ones were the occasional annotations of L1 English in the L2 German captions, which helped to facilitate vocabulary learning, reduce learners' language barriers, and can be functioned as an efficient instructional aid.

In short, factors such as age, speed, L2 orthographic scripts, content familiarity, types of captions, viewing sequential conditions/ times, and proficiency levels have a profound impact on when captions are utilized and who are appropriate with captioning. As a result, careful preparation of videos with consideration of influential factors should be required. In the next part of this paper, the captions' effects on language aspects and learning of language skills will be reviewed.

5. Effects of captioned videos on language learning (What-element)

As mentioned in previous parts, videos have pedagogical, cognitive, and psychological roles in language learning and teaching. This part aims to review what effects videos have on language aspects (vocabulary, grammars, and pronunciation) and language skills.

With respect to language aspects, the employment of audiovisual materials ostensibly has profound effects on enhancing lexical resources and enriching opportunities for being immersed in L2 exposure (Perez & Rodgers, 2019; Webb, 2015). Particularly, Winke et al. (2013) perceived captions as a useful aid for form-meaning connections of unknown words. On-screen texts play an essential role in boosting vocabulary learning (Vanderplank, 2016). Peters (2019) studied the effects of on-screen visuals on vocabulary learning of 142 English learners in the Netherlands. The data was conducted in three viewing conditions, namely English L2 captions, Dutch L1 subtitles, and no captions/ no subtitles. The results demonstrated that captions made the most effective lexical learning gains and that word learning has a positive relation with images, which notably corresponds to the multimedia learning theory. Further, Dang (2020) revealed that TV programs related to academic disciplines are useful resources for incidental learning of academic vocabulary. This correlates with the explanations of Nation (2011) that deliberate vocabulary learning (i.e., learning from exercises and tasks to enhance word retention in both short-and-long-term memories) should be "one part of a well-balanced learning program" (p. 533). A combination of deliberate/ intentional and incidental vocabulary learning should be implemented (Nation, 2013). In order to make incidental learning happen, it is important for learners to be exposed in rich L2 exposure (Schmitt, 2008, as cited in Dang, 2020). Consequently, captioned videos can represent supplementary and rich input resources to do incidental learning work (Peters & Webb, 2018).

Less noticeably, captioned videos also have significant effects on learning of grammatical aspects. In a study by Plotnikova (2017), she revealed that captioned videos facilitated both intentional and incidental types of grammatical learning when captioned videos were viewed twice. Besides, not only are L2 word recognition, and meaning recalls promoted in the videos, but grammatical functions of L2 words are also enhanced greater in the use of videos compared to the utilization of reading blogs (Arndt & Woore, 2018). With regard to pronunciation aspects, Wisniewska and Mora (2020) investigated the influences of captioned videos on L2 pronunciation. The findings showed that participants improved their pronunciation with the support of captions or, with uncaptions when they directly focused on phonological aspects. These results are interestingly different from previous studies in which captioned videos are more beneficial for speech perception and speech segmentation compared to uncaptioned ones.

Regarding their effects on learning four language skills, Mitterer and McQueen (2009, as cited in Yeldham, 2018) found that captions help to improve listeners' bottom-up processing because listeners can perceive the hearing words more clearly. This process is important for listening comprehension since it is processed incrementally from smallest units of phonemes to larger ones of words and phrases (Buck, 2001, as cited in Ling & Kettle, 2011). Similarly, Vanderplank (2016) noted that captions help viewers to improve segmentation speech processing, that is, the ability to identify word boundaries and language chunks. Additionally, Kam, Liu, and Tseng (2020) studied the influences of working memory capacity and modality preferences on captioned viewings to enhance L2 listening performances. The findings first showed that captions were beneficial for processing L2 listening comprehension. Second, they revealed that for learners with a low capacity of working memory, their preferences for visuals or sounds did not have an effect on L2 listening performances; whereas, for learners with high capacity, their modality preferences had a profound influence on L2 listening outcomes. That is to say, visual learners

perform best with captioned viewing while auditory ones achieve the best listening outcomes with uncaptioned ones.

For the skill of reading, viewing captions helps to facilitate the reading skill of low-proficient learners (Muñoz, 2017; Taylor, 2005) and that of L2 learners with L1-L2 orthographic differences (Winke et al., 2010; Winke et al., 2013). With regard to productive skills, Spring (2020) explained that videos are an effective instructional tool to improve learners' oral proficiency in a project-based classroom. He added that among three oral proficiency aspects (namely syntactic complexity, accuracy, and fluency), the syntactic complexity and the accuracy had the greatest improvement. For the writing one, Hsu (2013) investigated videos' effects on the use of advanced words of fifty Taiwanese undergraduate students in their EFL writing class. The researcher employed four kinds of videos (captions, non-captions, captions with silent sounds, and sounds only). The findings showed that (i) students found pre-writing video activities relaxing and encouraging to free writing, and (ii) non-captioned viewing resulted in the highest percentage of using advanced vocabulary. These findings further suggested that videos were also beneficial for top-down processing to activate learners' background knowledge and that high proficient students paid more attention to auditory modes.

In a nutshell, captioned videos have profound impacts on enhancing students' acquisition of language aspects and developing their language skills, albeit overloading the cognitive processing of multimodal inputs. Consequently, some pedagogical implications will be presented in the concluding part of the paper.

6. Concluding remarks

This paper attempts to review Why, When, Who, What, and How captioned videos affect additional language learning and teaching. Generally, the reasons for employing videos and their facilitative roles in pedagogy, cognition, and psychology outweigh their drawbacks. They can be employed as an effective aid for teaching language aspects and skills in both formal and informal contexts.

The review also underscores the importance of influential factors for some appropriate classroom implications. First and foremost, learners' proficiency levels will affect the cognitive processing of viewing and/ or hearing of the videos. That is, the higher proficiency levels students reach, the more multimodal input they fully attend to and engage in. Consequently, a professionally reliable placement test of learners' proficiency levels should be utilized before implementing videos for language learners. In addition, instructors can use 'the 5-finger rule' to quickly elicit their students' proficiency levels during the listening/ viewing (Ivone & Renandya, 2022, p. 162). More particularly, instructors play the selected video for ONE minute; students are asked to raise their fingers to indicate difficulty levels in which number 5 represents the most difficult level, and instructors are consequently supposed to change a different material. Additionally, number 2 is the perfect choice of material and proficiency level, while number 0 and 1 represents the easy ones and number 3 and 4 indicates more instructional scaffoldings.

Secondly, when proficiency levels are ensured, instructors should think of other factors such as age, speed, content familiarity, and orthographic differences so that times and sequences of viewings with captions or/and non-captions are employed carefully and appropriately. That means, viewing captions once is never sufficient for the learners to promote their L2 learning. Besides, individual differences in cognitive styles of learning should be noticed for the best performance of L2 learners. Third, since the meaning-making process occurs beyond the boundaries of processing linguistic inputs, other modes of emotions, motivations, learner

characteristics, cultures, and social contexts definitely merit consideration. Last but not least, follow-up activities such as reflecting, retelling, transferring, and tweaking after the viewing/ listening becomes of paramount importance to know whether the students are doing it (Ivone & Renandya, 2022). In other words, students should have rooms during (synchronously) or after (asynchronously) classes for (i) sharing their opinions about the viewing topics (reflecting), (ii) articulating their own words to retell their viewing (retelling), (iii) using graphic organizers to visualize their comprehension (transferring), and (iv) adding modifications and changing something in the viewing contents to make it more interesting (tweaking).

In conclusion, videos are representations of multimedia learning and multimodality practices. Indeed, videos and captions do surround our life. Employing it as an instructional tool is an inevitable consequence. Careful selections of the video content and thoughtful implementations with purposeful delivery steps become of paramount importance. The paper needs to be studied further in the author's context of teaching pre-intermediate and intermediate students at postsecondary institutions in Vietnam.

ACKNOWLEDGEMENTS

The author wishes to express his deepest appreciation to two anonymous reviewers for their valuable comments. Additionally, he respectfully acknowledges that he lives, works, studies, plays, and writes this paper on the unceded and traditional territories of the xʷməθkʷəy̓əm (Musqueam), Sḵwx̱wú7mesh Úxwumixw (Squamish), səl̓ilw̓ətaʔɬ (Tsleil-Waututh), q̓íc̓əy̓ (Katzie), kʷikʷəƛ̓əm (Kwikwetlem), Qayqayt, Kwantlen people.

References

Arndt, H. L., & Woore, R. (2018, October). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning & Technology, 22*(3), 124-142. doi:10125/44660

Bal-Gezegin, B. (2014). An investigation of using video vs. audio for teaching vocabulary. *Procedia - Social and Behavioral Sciences, 143*, 450-457. doi:10.1016/j.sbspro.2014.07.516

Berk, R. A. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning, 5*(1), 1-21.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Choi, J., & Yi, Y. (2016). Teachers' integration of multimodality into classroom practices for English language learners. *TESOL Journal, 7*(2), 304-327. doi:10.1002/tesj.204

Chun, D. (2001). L2 reading on the web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning, 14*(5), 367-403. doi:10.1076/call.14.5.367.5775

Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198. doi:10.2307/328635

Ciccone, A. (1995). Teaching with authentic video: Theory and practice. In F. Eckman, D. Highland, P. Lee, J. Milman & R. Weber (Eds.), *Second language acquisition theory and pedagogy* (pp. 203-218). Mahwah, NJ: Lawrence Erlbaum.

Cinganotto, L., & Cuccurullo, D. (2015). The role of videos in teaching and learning content in a foreign language. *Journal of e-Learning and Knowledge Society, 11*(2), 49-62.

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System, 29*(1), 1-14. doi:10.1016/S0346-251X(00)00057-9

D'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415-427). New York, NY: Springer-Verlag

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *META, 49*(1), 67-77. doi:10.7202/009021ar

Danan, M. (2016). Enhancing listening with captions and transcripts: Exploring learner differences. *Applied Language Learning, 26*(2), 1-24.

Dang, Y. T. N. (2020, June). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly, 54*(2), 436-459. doi:10.1002/tesq.552

Dressman, M. (2020). Multimodality and language learning. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 39-55). Chichester, UK: Wiley-Blackwell.

Fehlman, R. (1996). Viewing film and television as whole language instruction. *The English Journal, 85*(2), 43-50.

Flynn, K. F. (1998). Bring language to life! Using video in your ESL/EFL Program. *ESL Magazine, 1*(2), 1-20.

Fukushima, T. (2002). Promotional video production in a foreign language course. *Foreign Language Annals, 35*(3), 349-355. doi:10.1111/j.1944-9720.2002.tb01860.x

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology, 23*(2), 84-104. doi:10125/44684

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5*(3), 101-117. doi:10.1016/S0899-3467(07)60142-6

Hanley, J. E., Herron, C. A., & Cole, S. P. (1995). Using video as an advance organizer to a written passage in the FLES classroom. *The Modern Language Journal, 79*(1), 57-66. doi:10.1111/j.1540-4781.1995.tb05416.x

Hartman, F. (1961). Single and multiple channel communication: A review of research and a proposed model. *Audio Visual Communications Review, 9*(6), 235-262.

Herrero, C., & Vanderschelden, I. (2019). Introduction. In C. Herrero & I. Vanderschelden (Eds.), *Using film and media in the language classroom: Reflections on research-led teaching* (pp. 14-20). Bristol, UK: Channel View Publications.

Herron, C., Corrie, C., Cole, S. P., & Dubreil, S. (1999). The effectiveness of a video-based curriculum in teaching culture. *The Modern Language Journal, 83*(4), 518-533. doi:10.1111/0026-7902.00038

Herron, C., Dubreil, S., Corrie, C., & Cole, S. P. (2002). A classroom investigation: Can video improve intermediate-level French language students' ability to learn about a foreign culture? *The Modern Language Journal, 86*(1), 36-53. doi:10.1111/1540-4781.00135

Hsu, W. (2013). The effects of audiovisual support on EFL learners' productive vocabulary. *ReCALL, 26*(1), 67-79. doi:10.1017/S0958344013000220

Ito, Y. (1996). Communication between high school and college japanese language education: Implications from a survey on the use of video materials in the United States. *Foreign Language Annals, 29*(3), 463-479.

Ivone, F. M., & Renandya, W. A. (2022). Bringing extensive listening into the second language classroom. In M. Reed & T. Jone (Eds.), *Listening in the classroom: Teaching students how to listen* (pp. 157-170). Alexandria, Egypt and Virginia, USA: TESOL Press.

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. doi:10.1017/S0958344020000014

Kern, R., Ware, P., & Warschauer, M. (2016). Computer-mediated communication and language learning. In G. Hall (Ed.), *The routledge handbook of english language teaching* (pp. 542-555). London, UK: Routledge.

Kramsch, C. (2009). The embodied self. In C. Kramsch (Ed.), *The multilingual subject* (pp. 53-77). Oxford, UK: Oxford University Press.

Kruger, J. L. (2013). Subtitles in the classroom: Balancing the benefits of dual coding with the cost of increased cognitive load. *Journal for Language Teaching, 47*(1), 29-53. doi:10.10520/EJC143069

Ling, B., & Kettle, M. (2011). Rethinking comprehension and strategy use in second language listening instruction. *TESOL in Context, 21*(1), 5-22.

Mayer, R. E. (2009). The promise of multimedia learning. In R. E. Mayer (Ed.), *Multimedia learning* (2nd ed., pp. 03-27). New York, NY: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

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. doi:10.1002/acp.3050

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction, 21*(6), 687-704. doi:10.1016/j.learninstruc.2011.03.004

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eyetracking study. *System, 67*, 77-86. doi:10.1016/j.system.2017.04.015

Nation, I. S. (2011, October). Research into practice: Vocabulary. *Language Teaching, 44*(4), 529-539. doi:10.1017/S0261444811000267

Nation, I. S. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: CUP.

Nguyen, C.-D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly, 53*(1), 5-29. doi:10.1002/tesq.441

Nguyen, H. D. (2020). Understanding EFL students' use of listening strategies in watching English captioned movies. *Vietnam Journal of Education, 4*(2), 37-46. doi:10.52296/vje.2020.18

Norton, B., & Toohey, K. (2004). Critical pedagogies and language learning: An introduction. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 01-18). Cambridge, UK: Cambridge University Press.

Perez, M. M., & Rodgers, M. P. (2019). Video and language learning. *The Language Learning Journal, 47*(4), 403-406. doi:10.1080/09571736.2019.1629099

Peters, E. (2019). The effect of imagery and on-Screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4), 1008-1032. doi:10.1002/tesq.531

Peters, E., & Webb, S. (2018, September). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551-577. doi:10.1017/S0272263117000407

Plotnikova, A. (2017). *Intentional and incidental learning of English grammar through captioned video exposure* (Unpublished master's thesis). Barcelona, Spain: Universitat de Barcelona.

Rodgers, M. R., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal, 34*(1), 20-38. doi:10.1558/cj.29522

Schulte, G. (1991). Murder in the classroom: An audiovisual alternative for intermediate level language instruction. *Die Unterrichtspraxis/Teaching German, 24*(2), 176-182.

Sherman, J. (2003). *Using authentic video in the language classroom.* Cambridge, UK: Cambridge University Press.

Spring, R. (2020). Can video-creation project work affect students' oral proficiency? An analysis of fluency, complexity and accuracy. *TESL-EJ, 24*(2), 1-17.

Stein, P. (2004). Representation, rights, and resources: Multimodal pedagogies in the language and literacy classroom. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 95-115). Cambridge, UK: Cambridge University Press.

Stempleski, S., & Arcario, P. (1992). *Video in second language teaching: using, selecting and producing video for the classroom.* Alexandria, Egypt and Virginia, USA: TESOL Publications.

Stigler, J. W., Geller, E. H., & Givvin, K. B. (2015). Zaption: A platform to support teaching, and learning about teaching, with video. *Journal of E-Learning and Knowledge Society, 11*(2), 13-25.

Taylor, G. (2005). Perceived processing strategies for students watching captioned video. *Foreign Language Annals, 38*(3), 422-427. doi:10.1111/j.1944-9720.2005.tb02228.x

Toohey, K., Dagenais, D., & Schulze, E. (2012). Second language learners making video in three contexts. *Language and Literacy, 14*(2), 75-96. doi:10.20360/G2S59R

Truong, L. B., & Tran, L. T. (2014). Students' intercultural development through language learning in Vietnamese tertiary education: A case study on the use of film as an innovative approach. *Language and Intercultural Communication, 14*(2), 207-225. doi:10.1080/14708477.2013.849717

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System, 18*(2), 221-234. doi:10.1016/0346-251X(90)90056-B

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching. Subtitles for the deaf and hard-of-hearing as tools for lanuage learning.* London, UK: Palgrave Macmillan.

Vanderplank, R. (2019). Video and informal language learning. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 181-201). Chichester, UK: Wiley-Blackwell.

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL classroom. *Language Education in Asia, 2*(1), 113-119.

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. Richards (Eds.), *Language learning beyond the classroom* (pp. 159-168). London, UK: Routledge.

Weyers, J. (1999). The effect of authentic video on communicative competence. *Modern Language Journal, 83*(3), 339-349. doi:10.1111/0026-7902.00026

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. doi:10.1111/j.1540-4781.2013.01432.x

Wisniewska, N., & Mora, J. C. (2020, July). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition, 42*(3), 599-624. doi:10.1017/S0272263120000029

Yang, P. (2020). The cognitive and psychological effects of Youtube video captions and subtitles on higher-level german language learners. In M. R. Freiermuth & N. Zarrinabadi (Eds.), *Technology and the psychology of second language learners and users* (pp. 83-111). Switzerland: Palgrave Macmillan.

Yeldham, M. (2018). Viewing L2 captioned videos: what's in it for the listener? *Computer Assisted Language Learning, 31*(4), 367-389. doi:10.1080/09588221.2017.1406956

Zhang, L. J. (2016). Teaching reading and viewing to L2 learners. In W. A. Renandya & H. P. Widodo (Eds.), *English language teaching today: Linking theory and practice* (pp. 127-141). Switzerland: Springer International Publishing.