Research Article

Mark Feng Teng* and Danyang Zhang

The associations between working memory and the effects of multimedia input on L2 vocabulary learning

https://doi.org/10.1515/iral-2021-0130
Received February 19, 2021; accepted October 31, 2021; published online November 11, 2021

Abstract: The efficient use of working memory (WM) increases the potential of a learner's cognitive abilities in learning through multimedia. The present study aims to explore the role of working memory in vocabulary learning through multimedia input. In particular, we explore the possible associations between two components of WM – executive WM and phonological short-term memory (PSTM) – and the effects of three types of input conditions (Definition + Word information + Video, Definition + Word information, and Definition) on second language (L2) vocabulary learning. A total of 95 students completed learning under the three conditions and took two WM tests: a reading span test, which measures complex executive WM, and a non-word span test, which gauges PSTM. We administered a vocabulary knowledge test, which included receptive and productive vocabulary knowledge, immediately and after two weeks. Our findings, based on repeated-measures analysis of covariance (ANCOVA), support the pronounced effects of the Definition + Word information + Video condition in vocabulary learning and retention, as well as the significant role of complex and phonological WM in vocabulary learning and retention under the three conditions. Theoretical and pedagogical implications concerning the role of WM in vocabulary learning through multimedia input are discussed.

Keywords: multimedia input; productive vocabulary knowledge; receptive vocabulary knowledge; vocabulary learning; working memory

***Corresponding author: Mark Feng Teng**, Center for Linguistic Sciences, Beijing Normal University, Zhuhai, China, E-mail: markteng@bnu.edu.cn. https://orcid.org/0000-0002-5134-8504
Danyang Zhang, School of Foreign Languages, Shenzhen University, Shenzhen, Guangdong, China

Open Access. © 2021 Mark Feng Teng and Danyang Zhang, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

1 Introduction

Vocabulary knowledge is of great importance in foreign language (FL) and second language (L2) teaching and learning. Vocabulary knowledge is multifaceted in nature (Milton 2013). Nation (1990) and Schmitt (2014) categorized vocabulary knowledge into two dichotomous aspects: receptive and productive vocabulary knowledge. The former requires learners to recognize word form and understand word meaning. The latter refers to learners' ability to correctly express word meaning and appropriately use it in certain contexts (Laufer 1998; Laufer and Paribakht 1998). Given the multifaceted nature of vocabulary knowledge, language teachers have faced challenges in making vocabulary instruction effective. Multimedia program in the form of text, pictures, videos, and sound may help teachers deal with the challenge.

According to Paivio's (1990) dual-coding theory, visual and verbal aids, which come in the form of multimedia presentations, initiate, stimulate, and reinforce learning sensors. Owing to the recent boom in multimedia technologies, researchers have attempted to make vocabulary instruction more effective by integrating online visual and verbal aids into teaching and learning (Boers et al. 2017; Ramezanali and Faez 2019; Yanguas 2009; Yoshii and Flaitz 2002). Although visual and verbal aids can be part of traditional vocabulary instruction, the provision of online resources or technological aids may help learners acquire new word knowledge and develop strategies to enable them to take control of their learning and thus increase the depth of that knowledge (Teng 2018). In addition, understanding a word involves much more than knowing its definition, and simply memorizing a definition does not guarantee the ability to use a word in reading or writing. The use of multimedia technologies may benefit word knowledge acquisition (Teng 2021).

Based on the cognitive theory of multimedia learning (Mayer 2001), learners store incoming information in sensory memory and then select and transfer relevant visual and auditory information to modality-specific subsystems of working memory (WM), where it can be maintained and processed. Each of these modality-specific WM stores is limited in capacity. Learners need to construct incoming information in WM before integrating the information with their prior knowledge. Learning achievement takes place when such integration has occurred. According to Mayer (1997, 2001), WM, which involves conscious awareness, is vital for holding and manipulating multimedia input and contributing to knowledge acquisition. WM overload during processing may influence learners' perceptions or interpretations of multimedia input, thus affecting their vocabulary learning outcomes through multimedia input. Thus, it is worthwhile to explore how WM affects multimedia learning (Schüler et al. 2011), particularly in the context of L2 vocabulary learning.

WM, a cognitive device for online information retrieval and processing, is necessarily implicated in a process of coordinating cognitive and linguistic resources for multimedia learning (Mayer 2001). Such a process may impose cognitive burden and pressure on learners' WM resources when processing information for vocabulary learning from multimedia learning. Related to this, it is essential to explore the role of WM on vocabulary learning through multimedia input. However, this issue has been neglected in the context of English education in China. Despite China's educational reforms (which have been underway for a number of years) that are designed to facilitate efforts to develop learners' communicative competence, English teaching in China is still dominated by grammar-translation teaching approaches, a situation which arises in a large part due to the exam-orientated nature of the Chinese education system (Rao 2013). Under the exam pressure, teachers and learners may depend on grammar-translation teaching approaches for intentional learning of word form and meaning. Multimedia technologies, which provide learners with different vocabulary learning resources, including websites, apps, and online learning platforms, have received less attention in the Chinese context. It is thus meaningful to use multimedia resources, which involve texts, audio recordings, pictures and videos, to foster learners' acquisition of different aspects of vocabulary knowledge (Ramezanali and Faez 2019).

Therefore, vocabulary learning from multimedia input requires the coordination of cognitive and linguistic resources. Learners with different WM capacities could be expected to execute and orchestrate these processes with different degrees of efficacy, and thus vary in how they benefit from multimedia input. Despite the acknowledged role of WM in vocabulary learning, previous studies have not explored the effects of multimedia input on vocabulary learning and retention from the perspective of learners' WM in the Chinese context. We have attempted to fill this gap by examining whether two different components of WM – phonological short-term memory (PSTM) and complex WM – are associated with the impact of three different input conditions (Definition + Word information + Video, Definition + Word information, and Definition) on vocabulary learning. The findings have implications for both theoretical understanding and pedagogical practice.

2 Literature review

2.1 Multimedia input in L2 vocabulary learning: cognitive theory of multimedia learning

It is necessary to examine relevant multimedia learning theories for an understanding of L2 vocabulary learning through multimedia inputs. According to

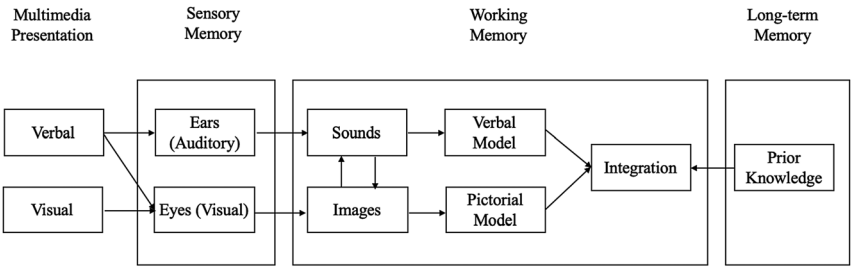


Figure 1: Cognitive theory of multimedia learning (Mayer 1997, 2001).

Paivio's (1972, 1986, 1990) dual-coding theory, humans, who possess various sensory modalities, can process information through two channels. One channel is responsible for verbal inputs from speech or writing, while the other channel is responsible for processing non-verbal information from images (Sadoski and Paivio 2001). From a cognitive perspective, multimedia learning features two information processing subsystems: verbal and visual information (Paivio 1972, 1986, 1990). Information received through the two channels may deepen learners' understanding of different types of information for L2 vocabulary learning.

On this basis, Mayer (1997, 2001) proposed the well-known cognitive theory of multimedia learning (Figure 1), which contains three important constructs that may influence learning. The first refers to "dual channels," reflecting similar ideas of dual-coding theory (Mayer 2001; Paivio 1972, 1986, 1990). The second, called "limited capacity," indicates that learners can only process a limited amount of either visual or verbal information in WM. The third highlights that learners are active in constructing knowledge, including (1) selecting relevant information, (2) organizing information, and (3) integrating information with prior knowledge (Mayer 2001). From the explanations of the three key constructs, we know that the information process is enhanced by multimedia input. As Figure 1 shows, learners first notice multimedia information before processing verbal and visual information via their ears and eyes in sensory memory. After that, the selected information comes into learners' WM. The information is then organized and becomes coherent verbal/pictorial models. Finally, the information, together with learners' previous knowledge, is connected, integrated, and stored in their long-term memory.

In Figure 2, the colored boxes illustrate how input is cognitively processed only through *textual* input. When learners are provided with textual input such as word definitions and example sentences, they use their eyes to receive the information and then bring that information into their sensory memory. After that, they select input and send it to their WM, formed as verbal models. Based upon learners' prior knowledge, information is actively organized and integrated.

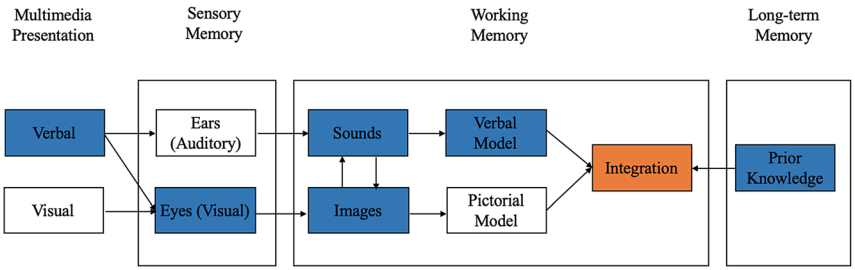


Figure 2: Processing of textual inputs.

As Figure 3 shows, the processing of multimedia (i.e., textual, pictorial, and auditory) input is more complex. For example, when learners are given both verbal and visual input (e.g., text and video), both their ears and eyes should be used to receive the information before delivering it to their sensory memory. Then, the two types of information are selected and transferred to learners' WM. The sounds and images in WM are not separate. The information mutually interacts and is converted, and then stored in verbal and pictorial models. The final step is to integrate learners' prior knowledge with the two types of information.

2.2 Research on multimedia input in L2 vocabulary learning

A number of previous studies have focused on the influence of different types of multimedia input (textual, auditory, and visual) on learners' vocabulary knowledge development (e.g., Akbulut 2007; Chun and Plass 1996; Ramezanali and Faez 2019; Yanguas 2009; Yoshii 2006; Yoshii and Flaitz 2002). Many of these studies have found that the combination of textual and visual input is more beneficial than one type of input. For example, as Plass et al. (1998) summarized, various kinds of

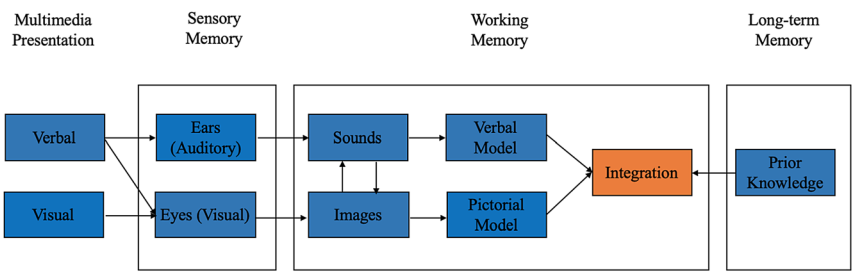


Figure 3: Processing of multimedia inputs.

input (auditory, visual, and pictorial) can make connections with (1) the target L2 word, (2) the image that represents the concept of the word, and (3) the first language (L1) equivalent. According to the questionnaire and interview data of Ramezanali and Faez (2019), learners indicated a more positive attitude towards the dual glossing mode of L2 definition and video animation.

More specifically, some studies have attempted to measure and compare the effects of different types of input. For instance, in Al-Seghayer's (2001) study, 30 English as a second language (ESL) learners were allocated to different input groups with the provision of (1) only printed text, (2) a printed text definition plus still pictures, and (3) a printed text definition plus video clips. The group with textual input and video clips performed better than the other two groups. In Yanguas' (2009) study, although the findings revealed significant differences between the experimental groups (the picture group and the text plus picture group) and the control group (the text-only group) regarding learners' receptive vocabulary knowledge development, no difference was found in terms of productive vocabulary knowledge. Likewise, Çakmak and Erçetin (2018) assigned 88 students who had a low English proficiency level to four groups according to gloss (i.e., the explanations of words that accompany a text) – no gloss, textual gloss, pictorial gloss, and textual plus pictorial gloss. This study showed that the type of gloss had no significant effect on learners' receptive or productive vocabulary acquisition.

In a meta-analysis that included the examination of gloss mode in vocabulary learning (Yanagisawa et al. 2020), glossed reading mode led to significantly greater learning gains than the non-glossed reading condition. The greatest effect was yielded by multiple-choice glosses, followed by marginal glosses, then hyperlinked glosses. Similarly, Ramezanali et al. (2021) conducted a meta-analysis of multimodal input and L2 vocabulary learning. They found many possible factors that may shape the effects of multimedia input on L2 learners' vocabulary development. These factors include learners' L2 proficiency, the language of instruction, and research design. As a result of these outcomes, it appears that further research is needed to explore the possibility of L2 vocabulary learning from multimedia input. In addition, individual differences in WM capacity may impact multimedia learning outcomes, as in one study, students with high WM capacity were better able to recall and could transfer more information during multimedia learning than students with low WM capacity (Anmarkrud et al. 2019). It is thus essential to explore the role of WM in L2 vocabulary learning through multimedia input.

2.3 Working memory (WM) and L2 vocabulary learning

WM refers to the cognitive system for storing, processing, and manipulating information for the temporary maintenance of task-relevant details in the face of other distracting information (Baddeley 1998, 2003). WM is operationalized as learners' constrained cognitive capacity, which allows them to simultaneously store and process information to gain awareness for completing mental tasks (Baddeley 2003). With regard to WM, there are two major research traditions: one is British, and the other is North American. The British tradition advocates the simple and storage dimensions of WM, such as the non-word repetition span task (Gathercole et al. 1994). The North American tradition suggests the implementation of complex memory span tasks to tap into the dual functions of WM. However, Williams (2012) claimed that the distinctions between the British and North American traditions are not always clear; hence, the definition of WM should be based on storage and processing functions.

Individuals vary a great deal in their cognitive skills and this influences their vocabulary learning outcomes (Teng and Zhang 2021). In relation to learners' cognitive skills, WM capacity is one of the most extensively investigated factors in individual differences in cognition. A perusal of the literature on WM reveals that most researchers refer to Baddeley's (1998, 2003) model of WM as the most influential framework for understanding WM. Based on this model, WM comprises the central executive, the phonological loop, and the visuospatial sketchpad. The central executive function is concerned with the control of information required to carry out complex tasks (Baddeley and Hitch 1974). The phonological loop and the visuospatial sketchpad, which are assigned for short-term memory, play important roles in retaining information. In particular, the phonological loop stores phonological information (e.g., remembering a phone number), while the visuospatial scratchpad maintains visual and spatial information (e.g., memorizing chess configurations) (Baddeley and Hitch 1974). Information in the two systems is assumed to decay rapidly. Engle et al. (1999) challenged Baddeley and Hitch's (1974) original conception of WM and highlighted that WM should be more connected to complex cognition in general. The central executive is crucial to maintaining task-relevant information.

As noted above, different notions of WM abound in the literature. However, these are differences in emphasis, rather than overall conception. WM is a multicomponent system that consists of domain-specific storage systems and domain-general executive components. Previous models on WM center on two areas: Baddeley and his colleagues focused on the storage components, while Engle and his colleagues emphasized executive functions. These different perspectives account for individual differences in the efficiency of L1 and L2 processing.

Researchers have paid attention to L2 vocabulary learning from the perspective of WM. Cheung (1996) attempted to study the correlation between phonological memory and natural vocabulary development among young learners by adopting non-word span to measure phonological memory. The participants in Cheung's study comprised a group of 84 seventh-grade high school students in Hong Kong. The results showed that phonological memory underlies L2 vocabulary acquisition. Researchers have also separately measured the predictive role of PSTM and executive WM in L2 vocabulary learning. Although some findings have suggested that young learners might not be able to rehearse transforming novel verbal materials into long-term memory (Gathercole et al. 1994), Cheung (1996) argued that WM and the rehearsal process are important determinants for young learners to pass on information to register it in long-term memory. Martin and Ellis (2012) looked at PSTM based on non-word repetition, non-word recognition, and listening span. Their study's participants were 50 native English speakers who learned single vocabulary words and sentences in a foreign language. Their results implied that PSTM is correlated with learners' vocabulary learning performance ($r = 0.33–0.45$). Their regression analyses suggested that PSTM made independent and significant contributions to their participants' vocabulary learning gains ($\beta = 0.39$). However, the actual mechanisms underlying PSTM and WM were not fully understood in their study. As a result, the findings may not be extended to L2 or FL learning. In a recent study (Karousou and Nerantzaki 2020), the focus was on assessing the effectiveness of a phonological memory training educational intervention on the vocabulary development of young L2 learners. A total of 97 learners were divided into two groups: an experimental group and a control group. The phonological working memory test was an English-sounding non-word repetition test. Vocabulary learning was evaluated through receptive and productive knowledge tests. The training included 33 sessions, which lasted for 12 weeks. The results supported the significant relationship between phonological working memory and L2 vocabulary size. Although phonological working memory did not significantly affect L2 receptive vocabulary knowledge, it significantly predicted productive vocabulary gains.

Engel and Gathercole (2012) explored the relationship between WM and L1, L2 and third language (L3) vocabulary, grammar, and literacy learning proficiency with 119 Luxembourgish-speaking children from 34 primary classes in 16 state schools. The learners completed complex span and verbal short-term storage WM tasks, as well as a series of vocabulary, grammar, and literacy tests. The results showed that PSTM was correlated with L1 and L2 vocabulary, grammar, and literacy learning outcomes. However, the study suggests that executive WM is a weak predictor of L2 vocabulary learning. The findings, based on controlling phonological awareness, indicated a non-significant relationship between PSTM and L3-French vocabulary acquisition. Such outcomes contradict those of Cheung (1996). One reason may be the different use of WM tasks. Specifically, while

Cheung (1996) used non-word repletion, Engel and Gathercole (2012) employed digit span. Another reason might be that the acquisition of unfamiliar phonology for the learners in the Luxembourgish-speaking context was influenced by their capacity to discern the sound system of the target language. Yang et al. (2017) explored vocabulary learning outcomes under different involvement load conditions. They also examined the role of WM in vocabulary learning outcomes. Data were collected from 85 first-year English major university students in China. WM was based on the dual tasks of sentence-final word recall and semantic similarity judgment. They found that WM was a significant predictor of the vocabulary learning scores of the comprehension-only group and the gap-fill group, but not the sentence-writing group. In addition, WM did not influence the delayed vocabulary test scores. However, the scoring of the WM test was based on the composite score of the reading span test. Such a scoring system ignores the different roles of PSTM and executive WM (Engel et al. 1999).

In sum, the above studies suggest that PSTM, rather than executive WM, significantly predicts vocabulary learning outcomes. However, in Linck et al.'s (2014) meta-analysis the WM executive control component was significantly correlated with L2 proficiency outcomes, including receptive and productive vocabulary knowledge. Yet despite the insight generated by these studies gaps in our understanding remain. For example, the role of WM in the vocabulary learning rate through multimedia input, to the best of our knowledge, has not been examined. In a recent study involving 63 L2 learners of French (Montero Perez 2020), individual differences in complex WM (measured by a backward digit span and an Ospan task) were used to predict learners' performance in picking up new words from watching videos. The findings on the captioned videos, which require comprehension of bimodal inputs (Teng 2019), can be extended to multimedia input. However, the validity of the claims concerning the role of WM in vocabulary learning through multimedia input is open to question because such claims are based on limited evidence. As a result, more research is warranted to fill the gaps, such as investigating the associations between WM and vocabulary learning from multimedia input.

2.4 Rationale of the current study

Despite the sufficient attention given to the effects of multimedia input on vocabulary learning (Teng 2021), few studies have looked into the role of WM in the context of multimedia input-guided vocabulary learning (Schüler et al. 2011). Adding WM as a variable may enhance the theoretical and practical understanding of vocabulary learning outcomes through multimedia input. In addition, there has been a call to examine the interface between cognitive variables and treatment

type in vocabulary acquisition (Alzahrani 2017). Hence, we aimed to determine whether WM is associated with the effects of multimedia input on vocabulary learning. We have attempted to fill a research gap in vocabulary research by investigating how WM (e.g., complex WM and PSTM) impacts the effects of three types of input conditions (Definition + Word information + Video, Definition + Word information, and Definition) on vocabulary learning. The three learning conditions have not been explored in previous studies. The findings provide insights into WM and L2 vocabulary learning from multimedia input. To achieve our goals, we developed two research questions:

(1) To what extent do the three multimedia conditions differ from each other in learners' L2 vocabulary acquisition?
(2) Do complex WM and PSTM predict vocabulary learning outcomes under different input conditions?

3 Methods

3.1 Participants

We recruited participants from the Department of English at a Chinese university, and conducted this study in three intact classes. We invited a total of 105 first-year students from three classes to participate. They ranged in age from 18 to 20 years old. Their native language was Chinese, and they were learning English as a FL. All participants reported that they had no other learning language experiences apart from Chinese and English. We randomly assigned each class to one of the three types of input conditions described above. We excluded 10 participants who failed to complete the post-test. Therefore, the final dataset included 95 students, with 32 assigned to the Definition condition, 33 to the Definition + Word information condition, and 30 to the Definition + Word information + Video condition. The participants completed the Vocabulary Levels Test (VLT) developed by Schmitt et al. (2001). Learners under the condition of Definition achieved a mean score of 26.21 (SD = 1.35) for the 2,000 word level, 16.34 (SD = 1.02) for the 3,000 word level, and 4.52 (SD = 0.78) for the 5,000 word level. Learners under the condition of Definition + Word information achieved a mean score of 25.35 (SD = 1.31) for the 2,000 word level, 14.31 (SD = 0.93) for the 3,000 word level, and 3.22 (SD = 0.59) for the 5,000 word level. Learners under the condition of Definition + Word information + Video achieved a mean score of 26.89 (SD = 1.39) for the 2,000 word level, 15.11 (SD = 1.08) for the 3,000 word level, and 3.05 (SD = 0.51) for the 5,000 word level. None of the participants had any knowledge of words at the 10,000 level.

3.2 Target words

We selected all 24 target words from the *Word of the Day* section of the *Merriam-Webster Online Dictionary* (available at https://www.merriam-webster.com/word-of-the-day). *Merriam-Webster* is one of the most well-known and trusted dictionaries across the world and is supported by a professional dictionary editing and writing team. In the *Word of the Day* section in the online dictionary, readers can enjoy one carefully chosen word every day. The section not only includes the word's definition, but also provides extra information (e.g., background stories, synonyms, antonyms, example sentences, and etymology). Unlike typical dictionaries, this section offers learners multimedia inputs with a video explaining word pronunciation, meaning and use, part of speech (PoS), and example sentences.

To ease the difficulty of understanding and acquiring a word, we used the *VocabProfile* section (http://www.lextutor.ca/vp/comp) of *Compleat Lexical Tutor* (www.lextutor.ca) to check word frequency. The lexical profiles of the definitions of the target words were as follows: 2,000 word level (65.37%), 3,000 word level (77.38%), 4,000 word level (83.24%), and 5,000 word level (91.5%). The definitions were deemed appropriate for learners' comprehension. The 24 target words (Table 1) were all beyond the K-10 level, and can therefore be categorized as low-frequency words. As all the participants had reached certain vocabulary level, but none knew any words at the 10,000-word level (see the section above on the participants), the target words were likely to be unfamiliar to them. Indeed, as the pre-test results showed, none of the participants had any previous knowledge of the 24 target words.

3.3 Treatment

We divided all the participants into three treatment groups. Specifically, we only offered students in Group 1 word definitions, which are explanations of word meanings. In Group 2, students had further access to the background information, which usually contains a story that introduces the target word's origin. In Group 3, students received multimedia inputs (e.g., videos) chosen from the *Word of the Day*

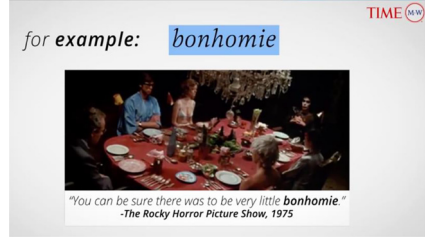
Table 1: The 24 target words.

Bonhomie	Parvenu	Tractable	Surfeit	Foible	Aught
k-14	k-18	k-15	k-13	k-12	k-15
Ersatz	Maunder	Sedulous	Ineffable	Insuperable	Bombast
k-14	k-21	k-20	k-12	k-14	k-13
Glom	Miscible	Hoodwink	Gormless	Passim	Satiate
k-23	k-18	k-14	k-19	k-14	k-13
Irascible	Nosegay	Emote	Mettlesome	Asperity	Hirsute
k-13	k-19	k-18	k-23	k-17	k-17

Table 2: The treatment details in the six groups.

Group	Treatment details	Example of treatment	Number
1 (Definition)	Input: definition only	**Target word: bonhomie** – Definition: good-natured easy; friendliness	32
2 (Definition + extra information)	Input: definition + extra information about the word	**Target word: bonhomie** – Definition: good-natured easy; friendliness – Did you know? English speakers borrowed bonhomie from the French, where the word was created from bonhomme, which means "good-natured man" and is itself a composite of two other French words: bon, meaning "good," and homme, meaning "man." That French compound traces to two Latin terms, bonus (meaning "good") and homo (meaning either "man" or "human being"). English speakers have warmly embraced bonhomie and its meaning, but we have also anglicized the pronunciation in a way that may make native French speakers cringe. (We hope they will be good-natured about it!) – Examples of bonhomie in a sentence: The bonhomie of strangers singing together around a campfire	33
3 (Definition + extra information + video)	Input: definition + extra information about the word + video	**Target word: bonhomie** – Definition: good-natured easy; friendliness – Did you know? English speakers borrowed *bonhomie* from the French, where the word was created from *bonhomme*, which means "good-natured man" and is itself a composite of two other French words: *bon*, meaning "good," and *homme*, meaning "man." That French compound traces to two Latin terms, *bonus* (meaning "good") and *homo* (meaning either "man" or "human being"). English speakers have warmly embraced *bonhomie* and its meaning, but we have also anglicized the pronunciation in a way that may make native French speakers cringe. (We hope they will be good-natured about it!)	30

Table 2: (continued)

Group	Treatment details	Example of treatment	Number
		– Examples of *bonhomie* in a sentence: The *bonhomie* of strangers singing together around a campfire – Video: 	

section of the *Merriam-Webster Online Dictionary*. Table 2 shows the treatment details, number of students, and treatment examples in each group.

3.4 Instruments

3.4.1 Vocabulary tests

The vocabulary test used in this study was designed based on the Vocabulary Knowledge Scale (VKS) adapted from Wesche and Paribakht (1996). The VKS is regarded as a widely accepted and commonly used framework to measure learners' vocabulary knowledge and to report learners' acquisition from complete unfamiliarity to a status where they are able to correctly and appropriately use a word in a particular context (Paribakht and Wesche 1997). Because the original VKS can be applied to evaluate whether learners have seen a word before, we modified this option to better fit our study. Figure 4 contains an example of the VKS for the target word *parvenu*.

The participants did not necessarily report every option from A to F. For Option A, they were required to state whether they knew the word. If so, they had to give a general explanation of the word (Option B) and/or a definition (Option C). To assess productive knowledge, participants were asked to report whether they could use the word in a sentence (Option D) and, if so, to write down the sentence in English (Option E). Finally, they were asked to translate the sentence into Chinese (Option F). Learners who provided a negative answer for Option A did not proceed with the following options. Likewise, learners who provided a negative answer for Option D did not proceed with the options after that.

Please make a choice(s) and fill in the blank when needed.
Word: parvenu

A. I don't know what it means.
Tick (√) if true: _____ (do not proceed)
B. I think it is related to the following word/idea:
_____ (answer may be given in English/Chinese)
C. I think it means:
_____ (answer may be given in English/Chinese)
D. I can't use this word in a sentence.
Tick (√) if true: _____ (do not proceed)
E. Write down your sentence in English.
_____
F. Translate your sentence into Chinese.
_____

Figure 4: An example of the pre-test: parvenu.

3.4.2 Scoring system for VKS

In terms of the VKS, we designed Options A, B, and C to evaluate participants' receptive knowledge (i.e., word meaning), while we designed Options D, E, and F to assess their productive knowledge (i.e., word use). As Table 3 shows, the scores of each option differed depending on the quality of the participants' answers. The maximum score for each item is 5. Here, each item refers to receptive or productive knowledge, and each aspect has five points. The total mark of either receptive or productive knowledge is thus 120 points.

Table 3: The marking criteria of the VKS.

Options & descriptions			Score
Receptive knowledge	A	√	0
	B	Related word(s) or idea(s)	2
		Unrelated or incorrect idea(s)	1
		No answer	0
	C	Provided correct synonym(s) or definition(s) that reflect all of the meanings	3
		Provided correct synonym(s) or definition(s) that reflect some of the meanings (only applicable for words with multiple meanings)	2
		An incorrect answer	1
		No answer	0
Productive knowledge	D	√	0
	E	A grammatically and semantically correct sentence	3
		(1) A sentence that demonstrates a very good knowledge of the target word but makes minor grammatical and syntactic error(s); or (2) A sentence that demonstrates satisfactory knowledge of the target word but uses correct grammar	2
		(1) A sentence that demonstrates very good/satisfactory knowledge of the target word but makes major grammatical and syntactic error(s); or (2) A sentence that makes no/minor grammatical and syntactic error(s) but demonstrates little knowledge of the target word	1
		No English sentence or the sentence is incomprehensible	0
	F	A translated sentence that shows a clear understanding of the meaning of the target word	2
		A translated sentence that shows a relatively vague understanding of the meaning of the target word	1
		No translation or a translated sentence that demonstrates no/an incorrect understanding of the meaning of the target word	0

With regard to scoring, two experienced teachers who were not teaching the participants were invited to independently score the participants' answers to minimize the risk of scoring bias. A third teacher was also available if the two teachers gave different scores. The interrater reliability for the VKS was 91.4%. Disagreements were solved based on majority opinion.

3.4.3 Working memory

The measure of WM includes complex WM and PSTM. We examined complex WM through a reading span test, which was a complex memory span task adapted from Daneman and Carpenter (1980). The purpose of complex memory span tasks is to measure learners' executive WM (Wen 2015). The test we chose focuses on the processing and storage components of short-term memory. The reading span test required the learners to judge whether each sentence was plausible, while at the same time remembering the final word in the sentence. Learners then recalled the sentence-final words in the order in which the sentences were presented after the entire set of sentences. This test included 60 sentences in Chinese. Although Mackey et al. (2010) argued that WM is an individual cognitive variable that is not related to language, we decided to use sentences in the participants' native language (rather than English) to minimize the influence of some learners' lower English proficiency on judging the English sentences and memorizing the final word. Among the 60 sentences, 10 were practice items and 50 were target sentences. We divided the 50 target sentences into 12 sets consisting of two, three, four, and five sentences. Each set (also called a span) was repeated three times. All sentences were in an affirmative and active form and contained 12–16 words. Half the sentences were semantically plausible, while the other half were not plausible.

We investigated PSTM through a non-word repetition task that we adapted from Gathercole et al. (2001). This test only focuses on the storage component of short-term memory. The learners were required to listen to 22 sequences of non-words. Each sequence included 4–7 one-syllable non-words. The learners were required to repeat all the words after each sequence. In total, the 22 sequences included 120 items. We created all the non-words based on the phonotactic rules of English. We used non-words rather than real words because learners' L2 vocabulary knowledge had the potential to impact their test performance, which would lead to an inaccurate evaluation of PSTM (Gathercole et al. 2001). A native speaker was invited to read the words and record the stimuli. The two WM tests were administered through E-Prime, a software that assesses learners' psychological behavior. Each participant was tested individually in a lab. The first WM test was conducted in the morning, while the second test took place in the afternoon to ease potential pressure related to memory overload placed on the participants.

3.4.4 Scoring for working memory

We adopted different scoring systems for the two WM tests to fit the nature of each test. In terms of the reading span test, we focused on three components: (a) the number of correctly recalled sentence-final words, (b) the number of correctly judged sentences, and (c) the mean reaction time for correctly judged sentences. We first transformed the raw scores for the three components into *z*-scores. We then summed the *z*-scores and divided them by three to obtain a composite score. With regard to reaction time, higher reaction times represented slower responses, for which we multiplied the *z*-scores for the reaction time by –1. This step ensured that a higher score would reflect better performance for the three components. Overall, a composite score represents the processing and storage components of short-term memory. Li and Roshan (2019) pointed out that a score for recalling sentence-final words only, rather than a composite score for the three components, might be inaccurate due to a possible trade-off between the storage and processing components of WM. For example, a learner may sacrifice accuracy of sentence judgment to achieve a better recall of the stimulus.

In terms of the non-word repetition test, the participants had to remember and repeat the 22 sequences of target words, including the 120 target words. The learners earned one point for each correctly recalled item. The total scores were based on the total words correctly recalled in the 22 sequences.

3.5 Data collection

We conducted the entire study in a computer classroom. Three teachers were responsible for the three conditions. We randomly assigned the three teachers to a condition after they attended a training session to help them understand the procedures of the treatment session. The teachers provided the participation instructions to the students about how to complete the experiment. The participants completed the training and the respective requirements under each condition in a computer classroom.

This study included a pre-test, a treatment, a post-test, and a delayed post-test (Table 4). The participants completed the pre-test four weeks before the treatment. Given that four weeks is a relatively long period of time, we assumed that learners would not commit the words to deliberate memory. The treatment session was completed in Week 5. We conducted the first post-test immediately after the treatment session. The participants completed the WM test (i.e., the complex WM test and the PSTM test) in Week 6. The reason for administering the WM test one week after the treatment session, rather than immediately after the treatment

Table 4: Procedures.

Conditions	Definition (n = 32)	Definition + word information (n = 33)	Definition + word information + video (n = 30)
Week 1	Pre-test (60 min)		
Week 5	Treatment: definition input (10 min)	Treatment: definition plus background information for the word (20 min)	Treatment: definition plus background information and watching a video about the word (30 min)
	Immediate test (60 min)		
Week 6	Complex WM test (reading span test; 15 min)		
	Phonological short term memory test (non-word span test; 5–8 min)		
Week 7	Delayed test (60 min)		

session, was to reduce the possibility that the cognitive load imposed on learners by the treatment session may influence their WM. The participants then finished the delayed post-test in Week 7. The VKS served as the pre-test, post-test, and delayed post-test. The differences were that we added different sets of non-target words to the test, and the order of test items was different. The purpose was to minimize learners' deliberate tendency to memorize the target words, because we expected their vocabulary learning outcomes mainly come from the training sessions. The participants had to finish the online VKS within 1 h. During the treatment session, each video was played only once. Learners needed to complete each question and could not move back to previous ones.

We obtained ethical permission for this study from the internal research committee of the experimented university. Participants signed a consent form to indicate their agreement to participate. They were assured anonymity and confidentiality, and were allowed to withdraw at any time. Participants received a coupon for their effort and time.

3.6 Data analysis

We analyzed all the data using SPSS Version 26. We ran a two-way analysis of covariance (ANCOVA), followed by Bonferroni post hoc comparisons, to test the influence of the independent variables on the dependent variables at the immediate and delayed tests while controlling for the covariate. The dependent variables were the two dimensions of the vocabulary test (receptive and productive vocabulary knowledge), which we administered twice. The independent variables were the three groups of input conditions. The covariates were the two components

of WM. The test results allowed us to examine the possible impact of (1) different types of input conditions on vocabulary knowledge acquisition and (2) the impact of WM on vocabulary test scores.

4 Results

Table 5 reports the average performance of the participants (Mean), the variation in the average performance (Std. Deviation), and the sample size (*N*) for each of the vocabulary knowledge and WM tests. Learners in the Definition + Word information + Video group achieved the best performance on receptive knowledge immediate test scores (*M* = 48.33), productive knowledge immediate test scores (*M* = 39.67), receptive knowledge delayed test scores (*M* = 44.23), and productive knowledge delayed test scores (*M* = 33.77). There were slight variations in complex

Table 5: Descriptive statistics.

		M	Std.	*N*
RK (immediate)	Definition	15.66	4.04	32
	Definition + word information	35.45	3.82	33
	Definition + word information + video	48.33	4.79	30
	Total	32.85	14.03	95
PK (immediate)	Definition	10.78	4.01	32
	Definition + word information	27.82	4.27	33
	Definition + word information + video	39.67	4.69	30
	Total	25.82	12.56	95
RK (delayed)	Definition	12.50	3.76	32
	Definition + word information	32.55	3.75	33
	Definition + word information + video	44.23	4.75	30
	Total	29.48	13.69	95
PK (delayed)	Definition	7.19	3.02	32
	Definition + word information	21.82	4.36	33
	Definition + word information + video	33.77	4.41	30
	Total	20.66	11.51	95
Complex WM	Definition	5.54*	0.61	32
	Definition + word information	4.56*	0.71	33
	Definition + word information + video	3.65*	0.97	30
	Total	7.28	0.78	95
Phonological WM	Definition	33.88	9.52	32
	Definition + word information	41.42	6.07	33
	Definition + word information + video	48.64	7.46	30
	Total	41.46	9.77	95

RK = receptive knowledge; PK = productive knowledge; *Z-score.

WM across the three groups. Again, learners in the Definition + Word information + Video group showed the highest phonological WM scores (M = 48.64). The ANOVA results revealed significant differences in phonological WM, F (2, 94) = 27.710, p < 0.001, complex WM, F (2, 94) = 5.907, p < 0.05, the immediate test of receptive knowledge, F (2, 94) = 472.752, p < 0.001, the immediate test of productive knowledge F (2, 94) = 350.695, p < 0.001, the delayed test of receptive knowledge, F (2, 94) = 478.176, p < 0.001, and the delayed test of productive knowledge, F (2, 94) = 347.445, p < 0.001.

We tested the same two-way ANCOVA model twice, firstly to investigate the post test data and secondly the delayed test data. Time was the within-group variable and treatment was the between-group variable. We entered complex and phonological WM in the model as covariates, although we had to exclude the pre-test since the participants did not exhibit any prior knowledge on the tests. Before carrying out the ANCOVA, we examined the assumption of homogeneity of regression slopes via the interaction between treatment and WM when predicting each dependent variable. The p values were larger than 0.05 for the immediate and delayed post-tests of receptive and productive knowledge. In addition, the dependent variables were normally distributed. These findings indicate that the assumption of homogeneity of regression slopes was met for each dependent variable.

There was a main effect of treatment conditions on the immediate post-test score after controlling for WM scores on receptive knowledge, F (2, 95) = 354.494, p < 0.001, η^2_p = 0.887, and productive knowledge, F (2, 95) = 345.294, p < 0.001, η^2_p = 0.885. There was also a main effect of treatment conditions on the delayed post-test score after controlling for WM scores on receptive knowledge, F (2, 95) = 263.317, p < 0.001, η^2_p = 0.853, and productive knowledge, F (2, 95) = 252.131, p < 0.001, η^2_p = 0.849. The results did not reveal a significant time effect of the immediate test on the delayed post-test score after controlling for WM scores in receptive knowledge, F (1, 90) = 0.172, p = 0.691, η^2_p = 0.003, and productive knowledge, F (1, 90) = 0.174, p = 0.691, η^2_p = 0.003. We did not detect a significant Time × Treatment interaction effect on the delayed post-test score after controlling for WM scores in receptive knowledge, F (1, 90) = 0.172, p = 0.691, η^2_p = 0.003, and productive knowledge, F (1, 90) = 0.174, p = 0.691, η^2_p = 0.003.

A Bonferroni post hoc test of the immediate post-test showed that the Definition + Word information + Video group had significantly higher scores than the Definition + Word group in receptive knowledge (p < 0.001) and productive knowledge (p < 0.001). The Definition + Word group had significantly higher scores than the Definition group in receptive knowledge (p < 0.001) and productive knowledge (p < 0.001). In terms of the delayed post-test, the Definition + Word information + Video group demonstrated significantly higher scores than the

Definition + Word information group in receptive knowledge ($p < 0.001$) and productive knowledge ($p < 0.001$). The Definition + Word information group scored significantly higher than the Definition group in receptive knowledge ($p < 0.001$) and productive knowledge ($p < 0.001$). The findings address the first research question concerning the different impact of the three learning conditions on L2 vocabulary learning.

The next step was to examine the role of WM in vocabulary learning outcomes. Overall, using Pillai's trace, we found a significant effect of complex WM on vocabulary learning, $V = 0.207$, $F(4, 87) = 5.694$, $p < 0.001$. Again, using Pillai's trace, we detected a significant effect of phonological WM on vocabulary learning, $V = 0.210$, $F(4, 87) = 5.772$, $p < 0.001$. Table 6 presents the effects of WM as a covariate on the different components of vocabulary knowledge at the two administered times.

Table 6 shows that complex WM, as the covariate, significantly predicted the immediate test scores of productive knowledge, $F = 22.664$, $p < 0.001$, $\eta^2_p = 0.201$, and receptive knowledge, $F = 16.795$, $p < 0.001$, $\eta^2_p = 0.167$, as well as delayed test scores of productive knowledge, $F = 21.387$, $p < 0.001$, $\eta^2_p = 0.192$, and receptive knowledge, $F = 10.858$, $p < 0.05$, $\eta^2_p = 0.108$. Phonological WM, as the covariate, also significantly predicted immediate test scores of productive knowledge, $F = 15.376$, $p < 0.001$, $\eta^2_p = 0.146$, and receptive knowledge, $F = 18.987$, $p < 0.001$, $\eta^2_p = 0.174$, as well as delayed test scores of productive knowledge, $F = 14.967$, $p < 0.001$, $\eta^2_p = 0.143$, and receptive knowledge, $F = 21.478$, $p < 0.001$, $\eta^2_p = 0.193$. The findings address the second research question concerning how complex WM and PSTM predict L2 vocabulary learning outcomes under different input conditions.

Table 6: The effects of WM on vocabulary knowledge scores.

	Vocabulary knowledge	Type III sum of squares	df	Mean square	*F*	Sig.	Partial eta squared
Complex WM	RK immediate	195.448	1	195.448	16.795	0.000	0.157
	PK immediate	270.648	1	270.648	22.664	0.000	0.201
	RK delayed	123.750	1	123.750	10.858	0.001	0.108
	PK delayed	219.843	1	219.843	21.387	0.000	0.192
Phonological WM	RK immediate	220.955	1	220.955	18.987	0.000	0.174
	PK immediate	183.614	1	183.614	15.376	0.000	0.146
	RK delayed	244.777	1	244.777	21.478	0.000	0.193
	PK delayed	153.847	1	153.847	14.967	0.000	0.143

RK = receptive knowledge; PK = productive knowledge.

5 Discussion

We examined the extent to which complex WM and PSTM were associated with the effects of three types of input conditions on vocabulary learning and retention. We found that (1) the learning and retention of vocabulary knowledge was more pronounced in the Definition + Word information + Video condition and (2) complex and PSTM influenced vocabulary learning and retention under the different input conditions. In the following section, we seek to explain and discuss the findings with reference to the nature of the treatment by comparing the results with previous findings and in relation to relevant theories. Theoretically, the findings contribute to existing frameworks such as the cognitive theory of multimedia learning (Mayer 1997, 2001) by further verifying and extending its arguments. Pedagogically, the findings offer insights for developing instructed L2 vocabulary learning, thereby deepening understanding of the affordances of multimedia input on vocabulary learning in the digital era.

5.1 Vocabulary learning from multimedia input

Our first purpose was to explore how vocabulary is learned and retained from multimedia input. The findings support the idea that combining definitions and information of words with associated visuals more effectively facilitates receptive and productive vocabulary knowledge, compared to only providing word definitions or incorporating information. In line with earlier studies (e.g., Akbulut 2007; Chun and Plass 1996; Ramezanali and Faez 2019; Yanguas 2009; Yoshii 2006; Yoshii and Flaitz 2002), the availability of visual input, along with word meanings, could help L2 learners to perform better in acquiring vocabulary knowledge (versus a single type of textual input). One explanation for this is that the availability of multiple types of input for a word may encourage learners to actively look up the word's meaning, thereby reinforcing learning and retention. Another explanation relates to the so-called hypermnesia effect, which predicts better recall of visual input over time than textual input, as textual input tend to be forgotten. This psychological effect may account for the improved performance in vocabulary learning with Definition + Word information + Video, and the lack of pronounced improvement with text definitions alone. We speculate that visual input allows one to develop a mental mode of the information (Teng 2019). A textual definition, on the other hand, may not be sensitive to learners' cognitive constraints, for which learners might not be able to reflect and refresh their short-term memory. In all three groups, scores for the delayed vocabulary tests were lower than those for the tests administered immediately after the treatment. One

possible explanation is that the learners may have demonstrated attrition for the words they memorized during treatment, as the findings also showed the influence of learners' WM on vocabulary learning and retention. However, for words where Definition + Word information + Video were provided were recalled significantly better on the delayed tests, whereas words where only a text definition or Definition + Word information were provided were recalled less well on the delayed tests.

These findings are in line with Mayer's cognitive theory of multimedia learning (1997, 2001), particularly with dual-coding theory (Paivio 1986, 1990). In the present study, the findings highlight the potential of presenting an explanation in words and visually, rather than solely textually. Mayer (2001) claimed that the effects of multimedia learning can be evaluated through transfer and retention. Transfer refers to learners' ability to use the material in a multimedia input to solve new problems and retention refers to learners' ability to remember important verbal information from multimedia input. In the present study, evidence for transfer was shown when learners demonstrated significantly better gains in receptive and productive vocabulary knowledge through processing information in the presented multimedia input. Evidence for retention was exhibited when learners demonstrated relatively modest gains on a delayed post-test. Reflecting the theory, the combination of definitions, word information, and videos may help learners to reinforce the referential connections between form and meaning, leading to better vocabulary knowledge learning and retention. The reinforcement of their vocabulary learning outcomes is probably due to the availability of multiple types of inputs for the target words. According to Mayer's cognitive theory of multimedia learning (1997, 2001), the presentation of verbal and visual information can attract learners' attention, helping them to build mental images that depict connections or provide gestalt. In the present study, students who received word information and definitions with videos, which contained narration and animation, were able to better retain information because they received the same information at least twice, either verbally or through visual inputs. Consistent with Paivio's (1972, 1986, 1990) dual-coding theory, dual channels may be an aid rather than a hindrance. In the present study, we argue that dual channels may have a constant, fixed quality, allowing for the development of a more enduring mental model of the information, lessening the cognitive load of the EFL learners in processing information, and increasing their short-term recall of vocabulary knowledge.

5.2 WM and vocabulary learning from multimedia input

As mentioned above, there are a range of factors which affect learners' L2 vocabulary learning (Ramezanali et al. 2021). In the present study, WM was shown to be one of those factors, which is in accordance with the work of Anmarkrud et al. (2019). The findings suggest that WM predicted the participants' learning and

retention of vocabulary knowledge in the three groups. In the study, the multimedia input in the form of a word definition and video information required learners to process the received input, while at the same time retrieving information from long-term memory and encoding new information in it. Based on the results, we argue that executive WM – which involves the storage and manipulation of information in the service of cognition – is needed in processing information when learning words from multimedia input. Complex executive WM, which assumes both storage and processing functions, is correlated with vocabulary learning performance (Cheung 1996; Yang et al. 2017). In line with Montero Perez (2020), complex WM predicted learners' performance in picking up new words from viewing videos. However, the findings were not consistent with those of Engel and Gathercole (2012), who did not find executive WM to be a significant predictor of vocabulary learning outcomes. The positive outcome in vocabulary learning in the present study can be explained from two angles: testing and learning. In terms of a testing perspective, the low-WM learners may have retrieved word knowledge due to the repeated tests rather than from the treatment. With regard to a learning perspective, learners with better WM were probably better at internalizing multimedia input for automatic use.

Consistent with previous studies (e.g., Engel and Gathercole 2012; Karousou and Nerantzaki 2020; Martin and Ellis 2012), the findings also support the role of phonological WM in vocabulary learning performance. One possible explanation is that PSTM – which supports the consolidation of stable phonological representations in long-term memory (Ellis 1996) – may facilitate the maintenance of relevant information in a multimedia input, as well as the regulation of processing during complex operations, allowing one to notice linguistic features and the possible integration into vocabulary learning and retention. Engel and Gathercole (2012) explored the differential effects of executive and phonological memory on vocabulary learning. Révész (2012) also distinguished PSTM and complex WM, and argued that PSTM helps learners to perform better on oral tests, while complex WM is essential for performance on written tests. We assert that vocabulary learning – which involves the sequential sound patterns of words and their arbitrary mapping to meaning – requires learners to harness their executive and phonological WM in processing input to acquire receptive and productive vocabulary knowledge. This argument partially aligns with Wen's (2015) proposal concerning (a) the role of executive WM in production and comprehension, and (b) the role of PSTM in affecting the final product of learning. However, our findings have to be interpreted cautiously, as we are still not sure how the different components of WM might influence the various aspects of vocabulary learning. Montero Perez (2020) offered insights into associating WM and multimedia input (e.g., videos), namely, that it is essential to explore the role of WM in multimedia input treatments based on a clear

theoretical account of the mechanism through which it affects vocabulary learning. Clearly, the relationship between WM and vocabulary knowledge type in multimedia input needs to be further investigated.

6 Conclusions

The present study highlights the effectiveness of combining definition, word information, and videos for the learning and retention of vocabulary knowledge. However, individual differences in complex and phonological short-term memory predicted students' vocabulary learning and retention under different input conditions. Despite the value of this study's findings there are some limitations which must be noted. First, the findings should be approached with caution because the target population consisted of English major students at a Chinese university. Given this, the study should be replicated in other language learning contexts to generalize the findings. In addition, we did not examine learners' pre-existing differences in English proficiency level. Future research can therefore study whether high-level learners could benefit more from multimedia inputs than low-level learners. Second, the duration of treatment was short. As such, the claims relating to vocabulary learning and retention can be better supported through a longitudinal study. Third, the assessment of vocabulary learning only focused on receptive and productive knowledge. Milton and Fitzpatrick (2013) asserted that vocabulary learning is an incremental, dynamic process that includes knowledge of form, meaning, use, word association, word parts, and collocations. Future studies can employ more vocabulary assessments to tap into different components of vocabulary learning. Fourth, during treatment, learners may have enhanced their vocabulary learning performance because of repeated exposure to the target words. Word exposure frequency, an important variable in vocabulary research (Teng 2020), should therefore be investigated. Finally, the target words included adjectives, nouns, and verbs. It would be interesting to see whether multimedia input has differential effects on the learning and retention of different types of words.

Despite the limitations, our study provides pedagogical and theoretical implications for teaching and learning vocabulary through multimedia input. The findings demonstrate that exposing learners to multiple modalities of presentation (i.e., definition, background information, and video) leads to effective vocabulary learning and retention. Future studies might consider comparing the effects of different gloss conditions, similar to the approach adopted and the insights generated by Yanagisawa et al. (2020). Pedagogically, it is essential to instruct students to use multimedia input for vocabulary learning. Textbook designers can

include interesting and relevant visual materials to accommodate learners' cognitive constraints in vocabulary learning and retention. Videos are not just for entertainment, but can be used to strengthen the connection between word form and meaning (Teng 2021). In terms of theoretical implications, the study's findings support dual-coding theory (Paivio 1972, 1986, 1990) and highlight the effects of explaining words and providing corresponding videos for better vocabulary learning and retention. The findings also support the cognitive theory of multimedia learning (Mayer 1997, 2001) and emphasize the role of learners' memory resources in processing multimedia input. Multimedia input, in the form of combining visual and verbal input as retrieval cues, can stimulate and encourage learners' engagement in the cognitive processes required for meaningful vocabulary learning and retention.

References

Akbulut, Yavuz. 2007. Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science* 35(6). 499–517.

Al-Seghayer, Khalid. 2001. The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language, Learning & Technology* 5(1). 202–232.

Alzahrani, Saad. 2017. Linking multimedia vocabulary CALL research to SLA cognitive theories. *American Journal of Educational Research* 5(7). 821–827.

Anmarkrud, Øistein, Anette Andresen & Ivar Bråten. 2019. Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist* 54(2). 61–83.

Baddeley, Alan. 1998. Working memory. *C.R. Académie des sciences Paris, Sciences de la Vie/Life Sciences* 321. 167–173.

Baddeley, Alan. 2003. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4. 829–839.

Baddeley, Alan & Graham Hitch. 1974. Working memory. In Gordon H. Bower (ed.), *The psychology of learning and motivation*, vol. 8, 47–90. New York: Academic Press.

Boers, Frank, Paul Warren, Lin He & Julie Deconinck. 2017. Does adding pictures to glosses enhance vocabulary uptake from reading? *System* 66. 113–129.

Çakmak, Fidel & Gülcan Erçetin. 2018. Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL* 30(1). 24–47.

Cheung, Him. 1996. Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology* 32(5). 867–873.

Chun, Dorothy M. & Jan L. Plass. 1996. Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal* 80(2). 183–198.

Daneman, Meredyth & Patricia A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour* 19. 450–466.

Ellis, Nick C. 1996. Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18. 91–126.

Engel, Pascale P. & Susan Gathercole. 2012. Executive and phonological processes in second language acquisition. *Journal of Educational Psychology* 104. 974–986.
Engle, Randall W., Stephen W. Tuholski, James E. Laughlin & Andrew R. A. Conway. 1999. Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General* 128. 309–331.
Gathercole, Susan E., Catherine S. Willis, Alan Baddeley & Hazel Emslie. 1994. The children's test of nonword repetition: A test of phonological working memory. *Memory* 2(2). 103–127.
Gathercole, Susan E., Susan J. Pickering, Melanie Hall & Sarah M. Peaker. 2001. Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology* 54. 1–30.
Karousou, Alexandra & Theodora Nerantzaki. 2020. Phonological memory training and its effect on second language vocabulary development. *Second Language Research*. https://doi.org/10.1177/0267658319898514.
Laufer, Batia. 1998. The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics* 19(2). 255–271.
Laufer, Batia & T. Sima Paribakht. 1998. The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning* 48(3). 365–391.
Linck, Jared A., Peter Osthus, Joel T. Koeth & Michael F. Bunting. 2014. Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review* 21(4). 861–863.
Li, Shaofeng & Saeed Roshan. 2019. The associations between working memory and the effects of four different types of written corrective feedback. *Journal of Second Language Writing* 45. 1–15.
Mackey, Alison, Rebecca Adams, Catherine Stafford & Paula Winke. 2010. Exploring the relationship between modified output and working memory capacity. *Language Learning* 60. 501–533.
Martin, Katherine I. & Nick C. Ellis. 2012. The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition* 34. 179–413.
Mayer, Richard E. 1997. Multimedia learning: Are we asking the right questions? *Educational Psychologist* 32. 10–19.
Mayer, Richard E. 2001. *Multimedia learning*. New York: Cambridge University Press.
Milton, James. 2013. Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In Camilla Bardel, Christina Lindqvist & Batia Laufer (eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*, 57–78. Amsterdam: European Second Language Association.
Milton, James & Tess Fitzpatrick. 2013. *Dimensions of vocabulary knowledge*. New York: Springer.
Montero Perez, Maribel. 2020. Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition* 42(4). 749–773.
Nation, I. S. Paul. 1990. *Teaching and learning vocabulary*. Boston: Heinle and Heinle.
Paivio, Allan 1972. *Imagery and verbal processes*. New York, NY: Holt, Rinehart & Wilston.
Paivio, Allan 1986. *Mental representations*. New York: Oxford University Press.
Paivio, Allan 1990. *Mental representation: A dual-coding approach*. Oxford, UK: Oxford University Press.
Paribakht, T. Sima & Marjorie Wesche. 1997. Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In James Coady & Thomas Huckin (eds.),

Second language vocabulary acquisition: A rationale for pedagogy, 174–200. Cambridge: Cambridge University Press.

Plass, Jan L., Dorothy M. Chun, Richard E. Mayer & Detlev Leutner. 1998. Supporting visual and verbal learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology* 90. 25–36.

Ramezanali, Nasrin & Farahnaz Faez. 2019. Vocabulary learning and retention through multimedia glossing. *Language, Learning & Technology* 23(2). 105–124.

Ramezanali, Nasrin, Takumi Uchihara & Farahnaz Faez. 2021. Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly* 55. 105–133.

Rao, Zhenhui. 2013. Teaching English as a foreign language in China: Looking back and forward: Reconciling modern methodologies with traditional ways of language teaching. *English Today* 29(3). 34–39.

Révész, Andrea. 2012. Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning* 62. 93–132.

Sadoski, Mark & Allan Paivio. 2001. *Imagery and text: A dual coding theory of reading and writing*. Mahwah, New Jersey; London: Lawrence Erlbaum Associates, Publishers.

Schmitt, Norbert. 2014. Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64(4). 913–951.

Schmitt, Norbert, Diane Schmitt & Caroline Clapham. 2001. Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language Testing* 18. 55–88.

Schüler, Anne, Katharina Scheiter & Erlijn van Genuchten. 2011. The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Education Psychology Review* 23. 389–411.

Teng, Feng. 2018. A learner-based approach of applying online reading to improve learner autonomy and lexical knowledge. *Spanish Journal of Applied Linguistics* 31. 104–134.

Teng, Feng. 2019. The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education* 142. 103655.

Teng, Feng. 2020. Retention of new words learned incidentally from reading: Word exposure frequency, L1 marginal glosses, and their combination. *Language Teaching Research* 24(6). 785–812.

Teng, Feng. 2021. *Language learning through captioned videos: Incidental EFL vocabulary acquisition*. New York: Routledge.

Teng, Feng & Danyang Zhang. 2021. Task-induced involvement load, vocabulary learning in a foreign language, and their association with metacognition. *Language Teaching Research*. https://doi.org/10.1177/13621688211008798.

Wen, Zhisheng. 2015. Working memory in second language acquisition and processing: The phonological/executive model. In Zhisheng Wen, Mailce Borges Mota & Arthur McNeill (eds.), *Working memory in second language acquisition and processing*, 41–62. Bristol: Multilingual Matters.

Wesche, Marjorie & T. Sima Paribakht. 1996. Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review* 53(1). 13–40.

Williams, John N. 2012. Working memory and SLA. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 427–441. London: Routledge.

Yanagisawa, Akifumi, Stuart Webb & Takumi Uchihara. 2020. How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition* 42. 411–438.

Yang, Yingli, Natsuko Shintani, Shaofeng Li & Yingyi Zhang. 2017. The effectiveness of post-reading word-focused activities and their associations with working memory. *System* 70. 38–49.
Yanguas, Iñigo. 2009. Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language, Learning & Technology* 13(2). 48–67.
Yoshii, Makoto. 2006. L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language, Learning & Technology* 10(3). 85–101.
Yoshii, Makoto & Jeffra Flaitz. 2002. Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal* 20(1). 33–58.