



HAL
open science

Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos

Mark Feng Teng

► **To cite this version:**

Mark Feng Teng. Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, 2024, 124, 10.1016/j.system.2024.103381 . hal-04850290

HAL Id: hal-04850290

<https://hal.science/hal-04850290v1>

Submitted on 20 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Please Cite as: Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, 103381. <https://doi.org/10.1016/j.system.2024.103381>

Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos

Mark Feng Teng markteng@mpu.edu.mo

Faculty of Languages and Translation, Macao Polytechnic University, Macau SAR, China
<https://doi.org/10.1016/j.system.2024.103381>

Keywords:

Incidental vocabulary learning
Retention
Listening
Reading
Reading while listening
Viewing captioned videos
Working memory
Prior vocabulary knowledge

Abstract: This study explores how certain input modes (i.e., listening, reading, reading while listening, and viewing captioned videos) affect incidental vocabulary learning in a foreign language context. It also examines the roles of learners' prior vocabulary knowledge and working memory in incidental vocabulary learning using the examined input modes. A total of 150 EFL students at a Chinese university were randomly and equally assigned to the four input modes, as well as a control group only took tests. Forty-eight words were chosen as target words. Participants either listened, read, read while listening to, or watched transcripts during viewing a documentary video. Incidental vocabulary learning outcomes were assessed through a two-part vocabulary test (i.e., form and meaning recognition). Mixed effects model results showed that incidental learning and retention of form and meaning recognition were superior under the caption-viewing condition followed by the reading-while-listening, reading, and listening conditions. Findings also revealed that prior vocabulary knowledge and working memory play distinct roles in incidental learning and retention of form and meaning recognition for each input mode. Relevant implications for vocabulary instruction are provided.

1. Introduction

Increasing attention has been given to incidental vocabulary learning in a foreign language context. This type of vocabulary learning is a by-product of meaning-focused activities (e.g., reading, listening, or viewing) for interest, information, and enjoyment purposes (Webb, 2020). Scholars have begun exploring incidental vocabulary learning via multiple input modes, including listening, reading, and viewing (Feng & Webb, 2020). This young line of research is contextualizing the amount of input needed and the potential for incidental vocabulary learning.

Most work in this vein has concerned the role of reading input (e.g., Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003). A strong connection exists between reading and the incidental learning of word forms and meanings. Reading can engage learners while helping them develop reading fluency as they consolidate prior lexical knowledge. Researchers are also interested in incidental vocabulary learning from spoken input based on findings supporting the possibility of such learning from listening (e.g., van Zeeland & Schmitt, 2013; Vidal, 2011). Brown et al. (2008) pointed out that listening yields smaller gains than reading and suggested an input mode of reading while listening for incidental vocabulary learning; this mode is similarly effective for the incidental learning of multiword items compared with reading or listening alone (Webb & Chang, 2022). Recent research has described the utility of viewing

second language (L2) TV programs for incidental vocabulary learning (e.g., Peters & Webb, 2018). L2 program videos, featuring multimodal input of print text and images, help learners cultivate the skills required to evaluate multimodal texts that use visuals for vocabulary acquisition. To enhance video's promise for incidental vocabulary learning, empirical investigations have also included captions for L2 videos (e.g., ; Montero Perez et al., 2014; Teng, 2022). Studies on incidental vocabulary learning from viewing are important, as watching TV is the preferred input mode for out-of-class L2 learning (Peters, 2018; Vanderplank, 2016). Studying captions is crucial as well: they can help learners process and remember information for incidental vocabulary learning (Teng, 2021). Dang et al. (2022) innovatively explored this type of learning through input modes such as listening, reading, reading while listening, viewing, and viewing with captions. However, little is known about how individual differences in working memory (WM) and prior vocabulary knowledge affect incidental vocabulary learning using different input sources.

Given the criticality of incidental vocabulary learning from reading, listening, reading while listening, and captions from the perspectives of frequency and prior vocabulary knowledge (Teng, 2024), it is also essential to understand how learners' individual differences in working memory and prior vocabulary knowledge may affect this form of learning. One's prior vocabulary knowledge level may determine incidental vocabulary learning outcomes (e.g., Peters & Webb, 2018). WM, a key component in the ability to maintain and rehearse information (Baddeley, 2000), may further shape one's consolidation of lexical knowledge based on modes of viewing input (Montero Perez, 2020). Therefore, we consider the impacts of different input modes (i.e., reading, listening, reading while listening, and viewing captions) on incidental vocabulary learning. We also assess to what extent learners' prior vocabulary knowledge and WM influence such learning. Examining these two factors across input modes enriches the domain of incidental vocabulary learning.

2. Literature review

2.1. Incidental vocabulary learning from reading

Reading provides rich contexts, exposure to, and interaction with vocabulary, leading to the possible incidental learning of unknown words (Pellicer-Sánchez, 2017; Pellicer-Sánchez & Schmitt, 2010; Teng, 2020; Waring & Takaki, 2003; Webb, 2008). Waring and Takaki (2003) initially tested incidental vocabulary learning from graded readers. Participants recognized and recalled the meanings of 10.6 (42.4%) and 4.6 (18.4%) target words in a 26-word set. The delayed tests, which were administered three months later, demonstrated a substantial decay trend. Frequency was deemed a core element of incidental vocabulary learning from reading. Pellicer-Sánchez and Schmitt (2010) explored incidental vocabulary learning from reading a novel. Study participants progressed in spelling, word class recall, meaning recognition, and meaning recall. Pellicer-Sánchez (2017) subsequently delved into the incidental learning of collocational knowledge from reading. This learning was found to occur at a rate similar to learning single words. Teng (2020) examined learners' retention in recognizing and recalling word forms and meanings incidentally gained from reading. Findings highlighted the power of glosses and repeated target word encounters in maximizing incidental vocabulary learning. Webb (2008) evaluated frequency and contextual clues for incidental vocabulary learning from reading. The quality of the context appeared important for acquiring meaning, whereas frequency tended to affect form learning. Contextual quality may explain "why gains in knowledge of meaning have varied from word to word ... and study to study" (p. 238).

The above studies underline the role of reading input for incidental vocabulary learning. However, outcomes related to this style of learning have been inconsistent. Discrepancies may be due to frequency, the tests employed, contextual quality, or learners' vocabulary knowledge. Incidental vocabulary learning from reading seems cumulative, and more effort is needed to identify how new words' form-meaning links are incidentally learned through this task.

2.2. Incidental vocabulary learning from listening

Aural input provides learners with several types of knowledge required for language learning, including phonology, grammar, and vocabulary. Researchers have paid growing attention to incidental vocabulary learning from spoken input. Vidal (2003) explored incidental vocabulary learning through listening. Results from 116 university students showed that learners can achieve significantly better vocabulary gains from doing so. Their performance on a delayed posttest, administered four weeks after the treatment, was significantly better than on the pretest. van Zeeland and Schmitt (2013) also studied incidental vocabulary learning outcomes from listening. Vocabulary learning was assessed using a dimensional approach spanning form, grammar, and meaning. Of these three dimensions, 29.2% of cases (i.e., an average 7.05 out of 24 target items) were detected upon immediate posttest learning. Nineteen percent of cases, or 4.56 target items, were identified on the delayed test. Participants primarily came to recognize words (followed by grammar and finally word meaning) after listening. Jin and Webb (2020) more recently examined incidental vocabulary learning from listening through a unique medium: teacher talk. Approximately 2.85 words (15.8%) and 2 words (12%) were known at the posttest and delayed posttest, respectively. Listening to teacher talk can thus be a fruitful source of incidental vocabulary learning. Meanwhile, consistent with van Zeeland and Schmitt (2013), input frequency did not significantly affect participants' incidental vocabulary learning from listening. Jin and Webb (2020) also reinforced the importance of explaining target word meanings in one's first language for incidental vocabulary learning from listening.

Overall, the above research implies the potential of listening to input for incidental vocabulary learning, including single words and collocations. Yet participants' learning gains were quite small—perhaps because of challenges in speech segmentation while listening. For example, learners may struggle to balance demands for faster meaning processing of spoken words because it allows less time to process linguistic information than reading input (van Zeeland & Schmitt, 2013). Moreover, the quality of context may affect

vocabulary meaning comprehension from listening more than reading. Aural input is nonetheless integral for optimizing incidental vocabulary learning and warrants a closer look.

2.3. *Incidental vocabulary learning from reading while listening*

Along with the aforementioned types of incidental vocabulary acquisition, scholars have started to scrutinize reading while listening to an audio recording. Some learners tend to break sentences into small, incoherent parts when reading. Reading while listening can help learners retain sentence integrity, resulting in better comprehension. Webb and Chang (2012) explored vocabulary learning through assisted and unassisted repeated reading. Eighty-two students read or read and listened to 28 short texts several times. Reading while listening significantly influenced vocabulary learning. Chang (2011) specified the effect of reading while listening to audiobooks. During a 26-week study, seven students voluntarily took part in the reading-while-listening treatment while 12 received the usual formal instruction (control group). The reading-while-listening group gained 17 marks, whereas learners in the control group only gained four. The aural-written verification of reading while listening is particularly beneficial for incidental vocabulary acquisition among students learning English as a foreign language (EFL). In an empirical study, Webb and Chang's (2014) participants read and listen to the same graded readers in class and then worked on language activities with teacher involvement. Students' vocabulary knowledge increased significantly after the reading-while-listening treatment. For instance, this group learned 19.68 words on average from pre- to posttest, while the comparison group only learned 4.43 words. Webb et al. (2013) considered participants' incidental learning of collocations from reading while listening to a graded reader. Target words consisted of 1, 5, 10, and 15 encounters. Receptive and productive knowledge of collocations could be gained incidentally through reading while listening to a graded reader; repeated encounters with target collocations positively affected participants' incidental vocabulary acquisition.

In summary, studies indicate that reading while listening generates pronounced vocabulary learning outcomes. Brown et al. (2008) justified the benefits of reading while listening as follows. First, this treatment might help learners segment information into meaningful chunks, leading to effective vocabulary acquisition. Second, learners must read at the pace of the audio input when reading while listening, and this pace is likely faster than students' own. Third, reading while listening may help EFL students match a word's spoken and written forms to establish more robust auditory discrimination and word recognition.

2.4. *Incidental vocabulary learning from viewing*

Peters and Webb (2018) emphasized incidental vocabulary learning from viewing L2 TV programs. Upon controlling word- and learner-related factors, findings conveyed the potential of viewing a long TV program for incidental vocabulary learning based on meaning recall and meaning recognition. Word-related aspects (e.g., occurrence frequency and cognateness) and learner-related features (e.g., prior vocabulary knowledge) partly predicted incidental vocabulary learning from viewing. No captioning group was included. Captions, which were first used to facilitate video content comprehension among the deaf and hard of hearing (Vanderplank, 2016), are now gaining traction in a foreign language context. Captioned videos promote EFL students' incidental vocabulary learning because simultaneously presenting visual and verbal input stimulates information processing and recall. Incidental vocabulary learning then becomes feasible (Teng, 2021). Several empirical studies have confirmed the role of captions in vocabulary learning. For example, Peters et al. (2016) compared the use of captions and subtitles. Results from 31 secondary school EFL students showed that captions led to significantly better outcomes than subtitles: participants in the captioning group achieved correct responses of 19.3% for meaning recall and 48.2% for form recognition, whereas the subtitling group achieved 20.8% for meaning recall and 32.4% for form recognition. Teng (2022) further verified captions' utility for incidental vocabulary learning; participants in the captioning group outperformed the non-captioning group in terms of word form and meaning recognition and recall. Learner-related factors, including proficiency level and aptitude, may influence incidental vocabulary learning from captioned videos. Some scholars have compared captioning conditions in this regard (Montero Perez et al., 2014, 2018), with positive results on the keyword captioning and full captioning with highlighted keywords groups for meaning recognition (Montero Perez et al., 2014) and students in the glossed keyword captions group scored best on the form recognition and meaning recall tests (Montero Perez et al., 2018). Recently, Teng (2023a) supported the effectiveness of glossed captions for young learners' incidental vocabulary learning and Teng (2023b) suggested the full captions and keyword captions made significant contributions to incidental learning and retention of form recognition and incidental learning of meaning recall but not of delayed meaning recall. The effects of different captioning groups remain inconclusive, possibly because of test modality and video genre (Teng, 2023c). However, captions do seem to play a part in vocabulary acquisition.

Despite the impacts of captions, Winke et al. (2010) contended that learners experience a split-attention effect when processing verbal and nonverbal input. The merits of captioning are constrained by differences in script, vocabulary knowledge, and learners' language proficiency. It is accordingly necessary to explore how individual differences affect captioning's contributions to learners' incidental vocabulary acquisition.

2.5. *Comparing input modes for incidental vocabulary learning*

Researchers have empirically compared input modes for incidental vocabulary learning. Vidal (2011) did so from listening and reading, with incidental vocabulary learning from reading being superior. Learners need more repetitions while listening (e.g., at least 5–6) than while reading (e.g., 2–3) to achieve marked vocabulary gains. Teng (2018) compared reading and reading while listening for incidental vocabulary learning. Participants who read while listening performed much better on four tests of vocabulary knowledge than participants who only read. The tests measured several types of vocabulary knowledge in L2 students: form recognition, grammar

recognition, meaning recall, and collocation recognition. Webb and Chang (2012) compared reading and reading while listening as well. Participants in the reading-while-listening condition acquired significantly more vocabulary knowledge incidentally compared with their counterparts in the reading condition. Feng and Webb (2020) compared how listening, reading, and viewing a TV program affected incidental vocabulary learning. While these three input modes indeed influenced such learning, no significant differences were detected between them. Two other studies (Brown et al., 2008; Webb & Chang, 2022) also compared reading, listening, and reading while listening and documented the great impact of reading while listening on incidental vocabulary learning, such as for single words (Brown et al., 2008) and collocations (Webb & Chang, 2022). Dang et al. (2022) compared listening, reading, reading while listening, viewing, and viewing with captions on incidental collocation learning. Reading, viewing, and viewing with captions each led to evident learning of form recognition. Even so, these modes' effectiveness did not vary significantly. Teng (2024) compared reading, listening, reading while listening, and captioned viewing for incidental vocabulary learning. Results supported the pronounced effects of captioned viewing while highlighting the effects of frequency and prior vocabulary knowledge.

In general, incidental vocabulary learning occurs via multiple input modes. Combining written and aural input might be particularly useful, but this assumption is tentative. Conflicting findings point to the need to better compare input modes' potential for incidental vocabulary learning.

2.6. Working memory (WM) and prior vocabulary knowledge

Individual differences (e.g., WM and prior vocabulary knowledge) must be accounted for when exploring incidental vocabulary learning from input modes. Such differences highlight the need to consider learners' cognitive abilities and linguistic background when investigating the efficacy of captioned viewing for vocabulary acquisition. By considering factors such as WM capacity and prior vocabulary knowledge, researchers can gain a deeper understanding of the underlying mechanisms and determine how to maximize input exposure for different students.

WM crucially influences learners' caption reading (Gass et al., 2019). It refers to one's ability to briefly maintain and operate on a limited amount of information while completing mentally demanding tasks (Wen et al., 2015). WM can be understood in light of Baddeley and Hitch's (1974) multicomponential model. This model asserts that WM has three components: (a) the central executive, which directs attention, maintains task goals, makes decisions, and retrieves memory; (b) the phonological loop, which is responsible for temporarily storing verbal information; and (c) the visuospatial sketchpad, which stores information in visual and spatial forms. Baddeley (2000) added episodic buffer as the fourth component. This feature stores and integrates visual, spatial, and verbal information; it also connects information with long-term memory. Malone (2018) verified WM's role in form recognition outcomes from reading while listening. Although a captioning group was not included, Montero Perez (2020) supported the potential of viewing a documentary video in incidental vocabulary learning. In addition, participants' prior vocabulary knowledge and complex WM positively correlated with incidental vocabulary learning from viewing. Teng and Zhang (2023) explored vocabulary learning using multimodal input. They attended to phonological short-term memory and executive WM, two popular components of WM in L2 acquisition research. Both components influenced vocabulary learning. There were some recent studies that supported the role of WM in incidental vocabulary learning in the captioned viewing context (Teng, 2023a-c; Teng & Cui, 2023), documenting the influence of WM in either learning single words or collocations. However, WM's impact on incidental vocabulary learning from caption viewing stands to be confirmed.

Prior vocabulary knowledge is another aspect of interest. Horst et al. (1998) underscored its role in vocabulary learning from reading. Later, Webb and Chang (2015) indicated its significance during extensive reading. Peters and Webb (2018) also argued for the role of prior vocabulary knowledge when viewing L2 programs. Dang et al. (2022) explored this attribute via input modes including reading, listening, reading while listening, viewing, and viewing with captions. Participants' prior vocabulary knowledge did not significantly contribute to their incidental learning of collocations. Puimège and Peters (2020) noted that prior vocabulary knowledge influenced vocabulary learning from viewing L2 TV programs. In addition, Teng and Mizumoto (2023) highlighted that depth of vocabulary knowledge can make a unique contribution to the prediction of incidental vocabulary learning at the form and meaning recognition level, in addition to the prediction afforded by the breadth of vocabulary knowledge. These inconclusive findings may be attributable to input modes' characteristics, hence the need for greater scrutiny.

3. The present study

The present study tested students' incidental learning of form and meaning recognition across control, listening, reading, reading-while-listening, and captioned video viewing groups. The present study also examined prior vocabulary knowledge and WM in incidental vocabulary learning gains. Two research questions were addressed.

1. Do different input modes lead to incidental learning of single words? If so, to what extent?
2. What relationships exist between incidental vocabulary learning through different input modes, prior vocabulary knowledge, and working memory?

4. Methods

4.1. Participants

The study sample consisted of 150 students at a university in China. All participants were English majors, but English was learned as a foreign language (EFL). Their ages ranged from 18.1 to 19.8 ($M = 19.1$, $SD = 1.01$), and the students were from six classes. They were gathered and then equally and randomly assigned to one of five conditions (i.e., listening, reading, reading while listening, caption viewing, and a control group that only took the tests). Their first language was Mandarin Chinese, and all were EFL students. The participants described themselves as intermediate English learners (e.g., the B1–B2 level based on the Common European Framework of Reference for Languages).

Participants signed a consent form prior to joining the study voluntarily. They were briefly told they would need to watch, listen to, or read a text or video and complete some exercises. The study's true purpose, namely to test incidental vocabulary learning from different input modes, was disclosed after the study. The vocabulary tests thus came as a surprise to students and reflected incidental vocabulary learning. Each participant received a supermarket coupon worth 50 Chinese Yuan as a token of gratitude. No participants withdrew from the study.

4.2. Video selection

The chosen video was a documentary titled *Ancient World* available on YouTube (<https://www.youtube.com/watch?v=Ml7lgPw-X3E>). This video was selected based on several criteria, following previous studies (e.g., Montero Perez et al., 2018). First, it needed to be appealing enough to maintain participants' interest. Second, its language dimensions (e.g., lexical coverage and speed of dialogue) had to be suitable for students. Third, the video needed to contain some words with which learners were unlikely to be familiar. *Ancient World* described the top 10 enigmas of the ancient world. A pilot group of 10 English majors chose this topic after watching the video and finding it interesting and suitable for L2 learning. The video runs for 1 h, 6 min, and 27 s. Its spoken and written languages align. We used VocabProfile (<https://www.lexutor.ca/>) to determine its lexical profile. The script contained 8039 running words. The 1000-, 2000-, 3000-, and 4000-word families covered 73.67%, 81.20%, 89.01%, and 95.8% of all running words in the script, respectively. Following the cut-off point of mastery (24/30; Hu & Nation, 2000), updated Vocabulary Levels Test (VLT) results (see Results section) showed that participants had reached the 4000-word level. Thus, the target learners could understand this video.

4.3. Target words

We took 48 words as test items based on VocabProfile. Approximately 56% of the target words (nouns, verbs, and adjectives) were beyond the 3000-word level. All test items occurred only once (see Table 1).

Table 1
Target words.

Items	Frequency	Item	Frequency
Zigzag	1	Meddled	1
Worship	1	Incarnation	1
Withstand	1	Illusion	1
Vicinity	1	Fierce	1
Verdict	1	Fascinating	1
Unreinforced	1	Explosion	1
Tribute	1	Execution	1
Thunderbolt	1	Erosion	1
Testament	1	Equivalent	1
Symbolic	1	Magnificent	1
Staggering	1	Enslaved	1
Speculate	1	Elaborate	1
Sophisticated	1	Distraction	1
Shipwreck	1	Depiction	1
Sculptor	1	Decipher	1
Screw	1	Crushed	1
Sacrilege	1	Companion	1
Revolt	1	Combat	1
Resilient	1	Coalition	1
Renaissance	1	Brutal	1
Reenactment	1	Besiege	1
Ransack	1	Alignment	1
Perplexing	1	Accuse	1
Mock	1	Acoustic	1

4.4. Learner-related factors

We explored incidental vocabulary learning through reading, listening, reading while listening, and viewing captioned videos. We also considered the roles of prior vocabulary knowledge and WM in participants' incidental vocabulary learning.

4.4.1. Prior vocabulary knowledge

Prior vocabulary knowledge was evaluated via the updated VLT (Webb et al., 2017), which has a paper-and-pencil format. This test mainly concerns receptive vocabulary knowledge across the 1000-, 2000-, 3000-, 4000-, and 5000-word levels. Each level includes 30 test items. The full test contains 150 items worth one point each. Test takers must match each definition with the word it defines. The measure's Cronbach's alpha value was 0.91, indicating sound item reliability.

4.4.2. WM

The assessment of WM was based on a reading span task (RST) adapted from Daneman and Carpenter (1980) and van den Noort et al. (2008). An RST is a verbal memory test often used to examine WM, cognitive processing, and reading comprehension. The participants were required to read a series of unconnected sentences aloud and judge whether each sentence made sense. This section captured the processing element of WM. Participants were also instructed to recall the end-of-sentence words in their original order at the end of a series, representing the storage element of WM. This RST therefore served as a complex verbal WM test. The number of sentences in a series increased incrementally. A sentence-word sequence is called a "set size." Each trial included 3–7 set sizes, totaling 80 target words and 80 sentences. The words to be recalled were unrelated. All sentences to be processed were in participants' first language of Mandarin Chinese. This parameter minimized the potential for individual differences in language proficiency and reading comprehension to counterbalance the results. Half of the sentences were plausible and half were not.

We followed Conway et al. (2005) in granting partial credit during scoring. For example, each item recalled in the correct order was awarded one point, even though participants could not remember all items in the trial. As in Unsworth et al. (2005), we set an 85% accuracy criterion: only when the accuracy rate of sentence judgment reached this threshold were the items in that trial calculated. This per-trial accuracy criterion mitigated the possibility that participants might sacrifice their sentence judgment to deliberately memorize the target words. This test was administered through E-prime. Its Cronbach's alpha value was 0.86, indicating good reliability.

4.5. Vocabulary test and scoring

Incidental vocabulary learning was measured with a two-part test (i.e., form and meaning recognition) in paper-and-pencil format. It was administered via a pretest, immediate posttest, and delayed posttest. Each test included a different set of 20 high-frequency words within the 1000 word level from the BNC/COCA word list (Nation, 2017). The aim was to encourage participants to focus on the assessment. The added words were not scored. This test included written and aural input so that a specific mode of exposure was not favored.

The form recognition measure was based on the yes/no EFL vocabulary test (Meara, 1992). Participants had to check off whether they recognized a word after hearing it read twice by a native English speaker. The meaning recognition test contained four response options: the correct meaning, three distractors, and an "I don't know this word" option. The "I don't know" option was meant to reduce wild guessing. Participants were required to choose one option after hearing the target word. Table 2 presents sample items for each test section.

The test took 40 min. Each correct answer was given 1 point, each incorrect answer was given 0 points, and the maximum score on each section was 48 points. The Cronbach's alpha value was 0.92, demonstrating strong reliability.

4.6. Procedure

This study was completed over three sessions. The first session involved a pretest and a VLT completed during the first week. The second session occurred two weeks later when participants were gathered and then equally and randomly assigned to a group. They also completed the informed consent form at that time. Each experimental treatment took approximately 1 h and 50 min. All participants used a separate computer. Learners in the reading group read the text online at a similar pace as in other conditions. Learners in the listening group listened to the text without visual support. Learners in the reading-while-listening group read the text online with audio support. Learners in the viewing group watched the captioned video. The presentation pace for the three groups was similar.

Table 2

Sample items for the vocabulary test.

	Form recognition test	Meaning recognition test
Items	Have you ever heard of the word? Do you recognize the word? If you are sure, please choose "yes". If you are not sure, choose "no".	Please choose the appropriate meaning for each word you have heard. If you are not sure, please choose the "I don't know" option.
Zigzag	* Yes * No	a. 开心前进 b. 曲折行进 c. 照顾 d. 恳求 e. 我不知道
Worship	* Yes * No	a. 崇拜 b. 出名 c. 高尚 d. 愤怒 e. 我不知道
Withstand	* Yes * No	a. 站住 b. 一起 c. 抵挡 d. 煎熬 e. 我不知道

Participants were told to focus on content comprehension and were allowed to take notes during the treatment. Learners in the control group only took the test. The third session, the delayed test, took place two weeks after the second session.

4.7. Data analysis

The Kolmogorov–Smirnov test of normality showed reasonable normality ($p > 0.05$ in all cases). The first question pertained to how different input modes lead to greater increases in incidental vocabulary learning. Linear mixed effects models were performed using the lme4 package (Bates & Maechler, 2010) in the R language and environment (R Development Core Team, 2009). Mixed effects models are preferable to analyses of variance (ANOVA) and covariance because they include time (pretest, immediate posttest, and delayed posttest) and groups (control vs. experimental) in a single model. These models also account for potential variance due to individual differences through random effects. The fixed effects consisted of group, time, and the interaction between group and time; the random effects were participants. Group and time were categorical variables. The control group acted as the reference group, and the pretest was the reference group for time. The variance inflation factors for Time and Group were approximately 1.0. Multicollinearity was thus not a problem.

The second question referred to the extent to which prior vocabulary knowledge and WM explain incidental vocabulary learning outcomes for each input mode. A separate logistic regression was conducted for the immediate posttest and the delayed posttest. The analysis was based on the number of cases instead of total test scores or overall learning gains per participant. The odds ratio was calculated to predict the odds of a correct response. Prior vocabulary knowledge and WM were entered into the model as predictors.

5. Results

5.1. Question 1: Incidental vocabulary learning across input modes

To begin, descriptive statistics were compiled for the baseline test (prior vocabulary knowledge and WM) and inferential analyses (i.e., ANOVA) were performed to determine whether groups' test outcomes were different (Table 3).

Table 3 indicates variations in participants' prior vocabulary knowledge and WM. The individual differences between groups were not significant ($p > 0.05$). Figs. 1 and 2 present similar graphical results for VLT and WM.

Table 4 displays statistics for the form and meaning recognition test. In all cases, the experimental groups' mean scores increased from the pretest to the immediate posttest. A loss in form and meaning recognition scores was detected on the delayed posttest.

Table 5 reflects the first mixed effects model comparing form recognition scores across the five groups and three test times.

A significant main effect for the listening, reading, reading-while-listening, and viewing groups on form recognition was noted ($p < 0.001$). A significant main effect was not found for time ($p > 0.05$). A significant group-by-time (posttest) interaction for all experimental groups was noticed ($p < 0.001$). The results further showed a significant group-by-time (delayed test) interaction for the reading, reading-while-listening, and viewing groups ($p < 0.001$). Fig. 3 illustrates that time and group interaction significantly influenced word form recognition. Taking the control group and pretest as reference points, all experimental groups demonstrated better performance in word form recognition on the posttest. The reading, reading-while-listening, and viewing groups performed best on the delayed test.

A series of pairwise comparison tests based on the emmeans package in R (Lenth, 2019) was conducted. Bonferroni adjustment was adopted for multiple comparisons. For each group, the estimate of the mean of the pretest scores was lower than for the posttest scores ($p < 0.001$). The results identified significant differences between the estimate of the mean of the immediate posttest scores and the delayed posttest scores for the reading, reading-while-listening, and viewing groups ($p < 0.05$). No significant differences appeared between the estimate of the mean of the pretest scores for all five groups ($p > 0.05$); that is, the groups possessed similar knowledge of the target words at the form recognition level prior to treatment. The viewing group earned significantly higher scores than the other groups on both the immediate posttest ($p < 0.001$) and the delayed posttest ($p < 0.001$). The viewing group showed the most pronounced vocabulary learning at the form recognition level.

Table 6 lists the mixed effects model for comparing meaning recognition across the five groups over three test times.

Findings revealed significant main effects of the listening, reading, reading-while-listening, and viewing groups on meaning recognition ($p < 0.001$). A significant main effect did not emerge for time ($p > 0.05$). The results also showed a significant group-by-time (posttest) interaction for all experimental groups ($p < 0.001$). Only a significant group-by-time (delayed test) interaction applied

Table 3
Results for VLT and WM.

Group	VLT		WM	
	<i>M</i>	Std.	<i>M</i>	Std.
Control	105.73	17.02	43	6.76
Listening	110.67	20.26	44.83	8.3
Reading	109.37	19.2	45.87	13.66
Reading while listening	105.8	20.33	42.87	8.39
Caption viewing	111.5	19.34	49.09	13.68
	$F = 0.593, p = 0.669$		$F = 1.741, p = 0.144$	

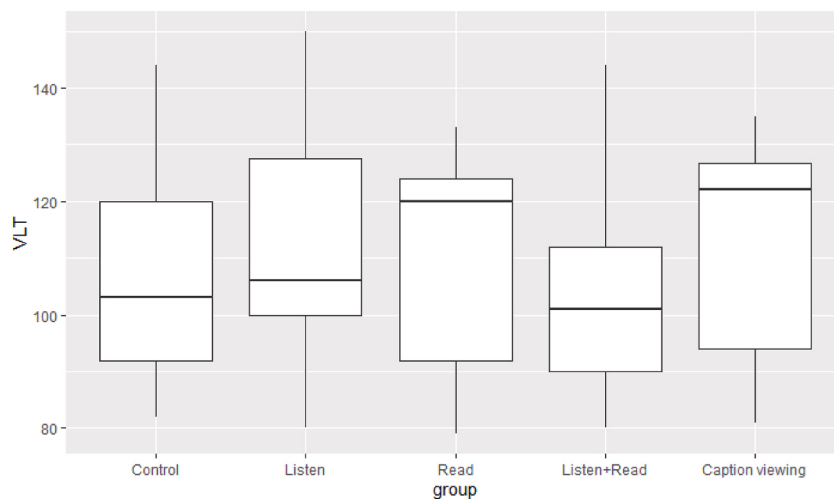


Fig. 1. Graphical results for VLT.

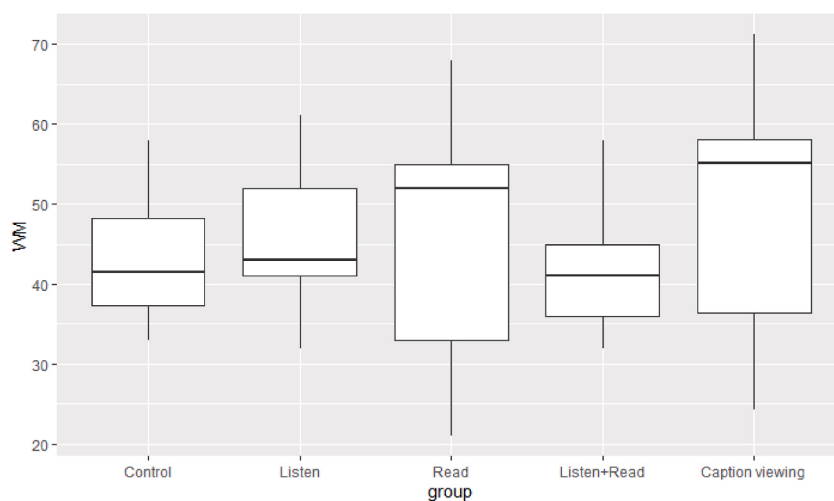


Fig. 2. Graphical results for WM.

Table 4

Descriptive statistics for dependent variables.

	Form recognition						Meaning recognition					
	Pre		Post		Delayed		Pre		Post		Delayed	
	M	Std.	M	Std.	M	Std.	M	Std.	M	Std.	M	Std.
Control	4	2.07	3.4	1.96	3.37	1.83	3.97	1.87	3.53	2.05	3.43	1.92
Listening	5.93	1.8	13.73	4.88	6.17	2.94	5.9	1.73	11.13	3.59	4.9	2.31
Reading	6.73	1.7	23.23	10.42	12.4	6.77	6.63	1.65	14.73	6.83	7.9	3.71
Reading while listening	7.33	2.2	23.43	7.9	12.17	5.85	7.33	2.23	16.17	4.73	8.37	3.02
Caption viewing	9.7	2.6	32.83	8.64	20.03	7.32	9.57	1.57	22.27	5.58	14.1	3.82

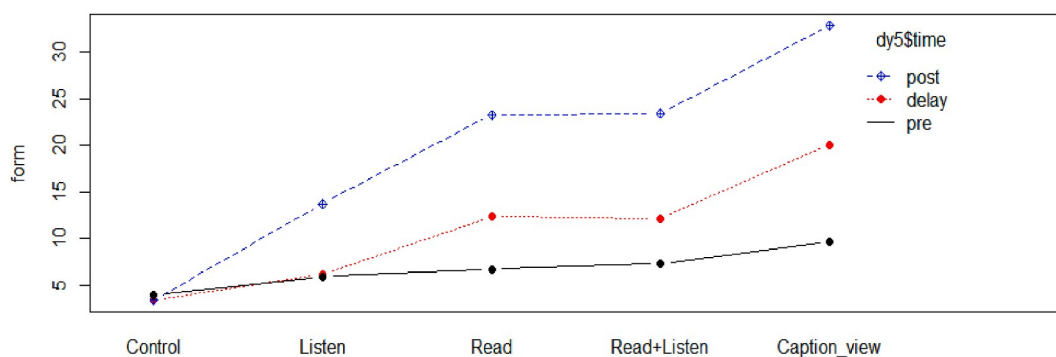
to the viewing group ($p < 0.001$). As Fig. 4 shows, the group and time interaction significantly affected word meaning recognition. When the control group and pretest were taken as references, all experimental groups performed better in word meaning recognition on the posttest. The viewing group exhibited the best performance on the delayed test.

A series of pairwise comparison tests were also conducted. The estimate of the mean of the pretest scores for each group was always lower than that of the posttest scores ($p < 0.001$). Significant differences between these estimates for immediate posttest scores and delayed posttest scores only applied to the viewing group ($p < 0.05$). No group's estimate of the mean of the pretest scores varied

Table 5

Comparisons of form recognition for the five groups over three test times.

	Estimate	Std. Error	z	Value	Pr ($> z $)
(Intercept)	1.38629	0.09128	15.187	<2.00E-16	***
Group (listening)	0.39429	0.11811	3.338	0.000843	***
Group (reading)	0.52078	0.11525	4.519	6.23E-06	***
Group (listening + reading)	0.60614	0.11348	5.341	9.23E-08	***
Group (viewing)	0.88583	0.10849	8.165	3.20E-16	***
Time (post)	-0.16252	0.13467	-1.207	0.227499	
Time (delay)	-0.17237	0.13503	-1.277	0.201758	
Group (listening) \times Time (post)	1.00176	0.1618	6.191	5.97E-10	***
Group (reading) \times Time (post)	1.40104	0.15659	8.947	<2.00E-16	***
Group (listening + reading) \times Time (post)	1.32425	0.15525	8.53	<2.00E-16	***
Group (viewing) \times Time (post)	1.38184	0.15029	9.195	<2.00E-16	***
Group (listening) \times Time (delay)	0.21094	0.17104	1.233	0.217473	
Group (reading) \times Time (delay)	0.783	0.16084	4.868	1.13E-06	***
Group (listening + reading) \times Time (delay)	0.67864	0.15974	4.248	2.15E-05	***
Group (viewing) \times Time (delay)	0.89764	0.15275	5.877	4.19E-09	***

**Fig. 3.** Time and group interaction for form recognition.**Table 6**

Comparisons of meaning recognition for the five groups over three test times.

	Estimate	Std. Error	z	Value	Pr ($> z $)
(Intercept)	1.37793	0.09167	15.031	<2.00E-16	***
Group (listening)	0.39703	0.11855	3.349	0.000811	***
Group (reading)	0.51418	0.11588	4.437	9.12E-06	***
Group (listening + reading)	0.6145	0.11379	5.4	6.66E-08	***
Group (viewing)	0.88036	0.10903	8.074	6.78E-16	***
Time (post)	-0.11568	0.13356	-0.866	0.386389	
Time (delay)	-0.14439	0.13458	-1.073	0.283308	
Group (listening) \times Time (post)	0.75068	0.16273	4.613	3.97E-06	***
Group (reading) \times Time (post)	0.91369	0.15851	5.764	8.20E-09	***
Group (listening + reading) \times Time (post)	0.90621	0.15635	5.796	6.79E-09	***
Group (viewing) \times Time (post)	0.96049	0.15106	6.358	2.04E-10	***
Group (listening) \times Time (delay)	-0.04132	0.17483	-0.236	0.813151	
Group (reading) \times Time (delay)	0.31915	0.1654	1.93	0.053659	.
Group (listening + reading) \times Time (delay)	0.27622	0.16322	1.692	0.090591	.
Group (viewing) \times Time (delay)	0.53228	0.15479	3.439	0.000584	***

significantly ($p > 0.05$); put simply, all groups possessed similar knowledge of the target words at the meaning recognition level prior to treatment. The viewing group earned significantly higher scores than other groups on the immediate posttest ($p < 0.001$) and the delayed posttest ($p < 0.001$). The viewing group also displayed the most pronounced vocabulary learning at the meaning recognition level.

5.2. Question 2: Relationships between incidental vocabulary learning through different input modes and prior vocabulary knowledge and working memory

The second question concerned how prior vocabulary knowledge and WM contribute to incidental vocabulary learning by input

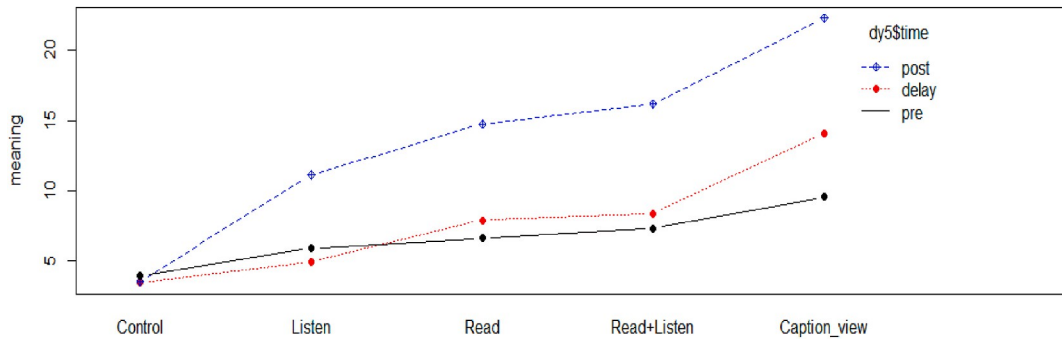


Fig. 4. Time and group interaction for meaning recognition.

mode. Regarding form recognition, Table 7 indicates that prior vocabulary knowledge made a significant difference to the model for the delayed posttest ($p < 0.001$) but not for the immediate posttest ($p > 0.05$) in the caption-viewing condition. The odds ratios (Exp [B]) revealed that, when participants' prior vocabulary knowledge increased by one unit, the odds of a correct response on the delayed posttest rose by 9.6%. Prior vocabulary knowledge significantly contributed to the model for the delayed posttest ($p < 0.05$) but not for the immediate posttest ($p > 0.05$) in the reading-while-listening condition. The odds ratio values (Exp [B]) demonstrated that a one-unit increase in learners' prior vocabulary knowledge raised the odds of a correct response on the delayed posttest by 7.4%. Prior vocabulary knowledge significantly contributed to the model for the immediate posttest ($p < 0.05$) and delayed posttest ($p < 0.05$) in the reading condition. As evidenced by the odds ratios (Exp [B]), as participants' prior vocabulary knowledge increased by one unit, the odds of a correct response on the immediate and delayed posttests improved by 10.1% and 9.7%, respectively. Prior vocabulary knowledge significantly contributed to the model for the immediate posttest ($p < 0.001$) and delayed posttest ($p < 0.001$) in the listening condition. The odds ratio values (Exp [B]) revealed that a one-unit rise in learners' prior vocabulary knowledge caused the odds of a correct response on the immediate and delayed posttests to increase by 10.2% and 10.3%, respectively. WM also significantly contributed to the model for the immediate posttest and delayed posttest. Such results were consistent across all input modes with the exception of the immediate posttest in the reading-while-listening condition ($p > 0.05$).

Table 8 presents results for meaning recognition. Prior vocabulary knowledge significantly contributed to the model for the delayed posttest ($p < 0.001$) but not for the immediate posttest ($p > 0.05$) in the caption-viewing condition. The odds ratios (Exp [B]) showed that as learners' prior vocabulary knowledge increased by one unit, the odds of a correct response in the delayed posttest increased by 10.5%. Prior vocabulary knowledge significantly contributed to the model for the immediate posttest ($p < 0.05$) and delayed posttest ($p < 0.001$) in the reading-while-listening condition. The odds ratio values (Exp [B]) confirmed that a one-unit rise in participants' prior vocabulary knowledge led the odds of a correct response on the immediate and delayed posttest to improve by 8.1% and 6%, respectively. Prior vocabulary knowledge significantly contributed to the model for the immediate posttest ($p < 0.05$) but not the delayed posttest ($p > 0.05$) in the reading condition. Odds ratios (Exp [B]) revealed that, as learners' prior vocabulary knowledge increased by one unit, the odds of a correct response on the immediate posttest increased by 10.2%. Prior vocabulary knowledge significantly contributed to the model for the immediate posttest ($p < 0.05$) but not the delayed posttest ($p > 0.05$) in the listening condition. The odds ratio values (Exp [B]) showed that as participants' prior vocabulary knowledge rose by one unit, the odds of a correct response on the immediate posttest rose by 10.1%. WM significantly contributed to the model for the immediate posttest and

Table 7
Logistic regression for form recognition.

			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. EXP(B)	
									Lower	Upper
Caption viewing	Immediate posttest	VLT	-0.002	0.007	0.091	1	0.763	0.998	0.984	1.012
		WM	0.064	0.011	35.672	1	0.000	1.066	1.044	1.088
	Delayed posttest	VLT	-0.034	0.007	20.480	1	0.000	0.967	0.953	0.981
		WM	0.087	0.011	57.559	1	0.000	1.091	1.067	1.116
Reading while listening	Immediate posttest	VLT	-0.094	0.074	1.611	1	0.204	0.910	0.787	1.053
		WM	0.290	0.180	2.596	1	0.107	1.336	0.939	1.902
	Delayed posttest	VLT	-0.295	0.096	9.500	1	0.002	0.745	0.618	0.898
		WM	0.791	0.233	11.586	1	0.001	2.207	1.399	3.480
Reading	Immediate posttest	VLT	0.015	0.007	4.223	1	0.040	1.015	1.001	1.030
		WM	0.043	0.011	16.805	1	0.000	1.044	1.023	1.066
	Delayed posttest	VLT	-0.025	0.009	8.450	1	0.004	0.976	0.959	0.992
		WM	0.082	0.013	42.805	1	0.000	1.085	1.059	1.112
Listening	Immediate posttest	VLT	0.023	0.004	36.020	1	0.000	1.023	1.015	1.030
		WM	0.023	0.009	6.844	1	0.009	1.023	1.006	1.041
	Delayed posttest	VLT	0.035	0.009	16.562	1	0.000	1.036	1.018	1.053
		WM	0.060	0.019	10.290	1	0.001	1.062	1.024	1.102

Table 8
Logistic regression for meaning recognition.

			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. EXP(B)	
									Lower	Upper
Caption viewing	Immediate posttest	VLT	-0.005	0.007	0.442	1	0.506	0.995	0.981	1.009
		WM	0.050	0.011	22.271	1	0.000	1.051	1.030	1.073
	Delayed posttest	VLT	-0.023	0.009	6.629	1	0.010	0.978	0.961	0.995
		WM	0.077	0.013	34.924	1	0.000	1.080	1.053	1.107
Reading while listening	Immediate posttest	VLT	-0.208	0.083	6.320	1	0.012	0.813	0.691	0.955
		WM	0.552	0.200	7.579	1	0.006	1.736	1.172	2.572
	Delayed posttest	VLT	-0.504	0.134	14.179	1	0.000	0.604	0.465	0.785
		WM	1.284	0.325	15.555	1	0.000	3.610	1.907	6.832
Reading	Immediate posttest	VLT	0.023	0.008	7.390	1	0.007	1.023	1.006	1.040
		WM	0.025	0.012	4.522	1	0.033	1.025	1.002	1.048
	Delayed posttest	VLT	0.008	0.012	0.397	1	0.529	1.008	0.984	1.032
		WM	0.054	0.016	11.040	1	0.001	1.056	1.022	1.090
Listening	Immediate posttest	VLT	0.015	0.004	11.770	1	0.001	1.015	1.006	1.023
		WM	0.027	0.010	7.205	1	0.007	1.028	1.007	1.048
	Delayed posttest	VLT	-0.002	0.019	0.008	1	0.927	0.998	0.961	1.037
		WM	0.151	0.057	7.109	1	0.008	1.163	1.041	1.299

delayed posttest as well. These outcomes were consistent across all input modes.

6. Discussion

The discussion centers on incidental vocabulary learning across various input modes. The focus was then turned to the roles of WM and prior vocabulary knowledge. Moreover, by comparing the findings with those of previous studies, I derive new arguments and possible research ideas.

6.1. Incidental vocabulary learning across input modes

The first research question considered certain input modes' potential to promote incidental vocabulary learning via form and meaning recognition. The results identified listening to spoken input as a source of such learning: participants exhibited greater word form knowledge of about 13.73 (28.6%) of the new words and word meaning knowledge of approximately 11.13 (23.18%) new words immediately after listening. These encouraging results were similar to those of van Zeeland and Schmitt (2013). As in earlier work on incidental vocabulary learning from listening (Jin & Webb, 2020; Vidal, 2003), listening to spoken input is essential to building an initial form-meaning link. Second, the findings provide evidence for the power of reading in incidental vocabulary learning. Such results were not surprising considering that previous studies have supported this role (e.g., Horst, 2005; Horst et al., 1998; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Teng, 2020; Waring & Takaki, 2003; Webb, 2008). Incidental vocabulary learning at the form and meaning recognition level appears to occur through reading, which is encouraging and expected. Being able to choose a word's meaning from a list of plausible choices (as in a multiple-choice test) shows that at least some knowledge of form and meaning has been retained. The capacity to recognize which words occurred in the text and which did not indicates that learners showed some familiarity with word form. This step is important in incidental vocabulary learning: managing to recognize a word's form from reading is a substantial step.

Third, encountering words during reading while listening contributed to participants' incidental vocabulary learning. Other studies have come to similar conclusions regarding single words (Teng, 2018; Webb & Chang, 2012, 2014) and collocations (Webb et al., 2013). The participants in the present study made sizeable gains in receptive knowledge of the form-meaning link through reading and listening to transcripts. The participants' scores increased by 16.1 words from 7.33 to 23.43 for form recognition and 8.84 words from 7.33 to 16.17 for meaning recognition. The size of these gains contrasts with the relatively small ones identified in prior research on reading (e.g., Horst et al., 1998; Waring & Takaki, 2003). Reading while listening provides greater opportunities to consolidate one's knowledge of unknown and partially known words.

Finally, the results (e.g., recognition of word form knowledge = 68.3% and word meaning knowledge = 46.3% at the immediate posttest) highlight the role of viewing captioned videos in incidental vocabulary learning (Montero Perez et al., 2014, 2018; Teng, 2022). This benefit of watching L2 programs is in line with studies that did not include a captioning group (Peters & Webb, 2018). We have thus confirmed the use of L2 captioned videos as a preferred input mode for incidental vocabulary learning (Teng, 2021; Teng, 2023a-c; Teng & Cui, 2023). It seems that these videos (i.e., combining verbal and visual mental representations) help learners organize and store information in WM in addition to activating prior knowledge. Incidental vocabulary acquisition therefore increases.

The identified incidental vocabulary learning and retention profile can be summarized as follows: caption viewing > reading while listening > reading > listening > control. These results remained consistent across the form and meaning recognition tests, mirroring studies aiming to depict acquisition profiles across input modes. For instance, reading is preferable to listening for incidental vocabulary learning (Vidal, 2011), reading while listening is better than reading only (Teng, 2018), and reading while listening is more effective than both reading and listening for incidentally learning single words (Brown et al., 2008) and collocations (Webb & Chang,

2022). Consistent with Teng (2024), captioned viewing yielded better incidental vocabulary learning performance than reading-while-listening, followed by it were reading and listening. Dang et al. (2022) found that reading, viewing, and viewing with captions each yielded significant differences on an immediate form recognition test; however, the groups' performance on the delayed posttest did not vary significantly. The present results partially echo these outcomes but do not support Feng and Webb's (2020) lack of significant differences between three input modes (e.g., viewing without captions, reading, and listening) for incidental vocabulary learning. Several arguments can be put forth based on these inconsistencies. First, EFL learners rely more on written than spoken input when processing information for incidental vocabulary learning. Second, reading while listening reinforces the benefits of navigating demands to comprehend content solely through spoken input. Finally, compared with reading while listening, on-screen text in captioned videos may better help learners understand L2 programs. Incidental vocabulary learning can then be maximized.

6.2. *Prior vocabulary knowledge and incidental vocabulary learning*

The findings imply that prior vocabulary knowledge plays distinct roles in form and meaning recognition for each input mode. These outcomes somewhat support research revealing significant relationships between prior vocabulary knowledge and incidental vocabulary learning through reading (Horst et al., 1998), listening (Vidal, 2011), viewing L2 TV programs (Peters & Webb, 2018), and captioned viewing (Teng & Mizumoto, 2023). Feng and Webb (2020) identified a partial effect of prior vocabulary knowledge on incidental vocabulary learning (e.g., they noticed a significant correlation between the two for reading and viewing but not for listening). In the present study, prior vocabulary knowledge partially predicted incidental vocabulary learning performance in the caption-viewing, reading-while-listening, reading, and listening conditions. Exceptions included the immediate meaning recognition posttest in the caption-viewing condition, the delayed meaning recognition posttest in the reading and listening conditions, the immediate form recognition posttest in the caption-viewing condition, and the immediate form recognition posttest in the reading-while-listening condition. According to Feng and Webb (2020), prior vocabulary knowledge may not have significantly influenced incidental vocabulary learning from listening because a written test measuring such knowledge might not reflect students' familiarity with the spoken form-meaning link. I argue that EFL learners' prior vocabulary knowledge could help them segment the amount of connected speech and the speech rate in spoken input, but immediate and delayed posttests may pose different demands.

Dang et al. (2022) assessed how prior vocabulary knowledge affects incidental learning of collocations after watching an academic lecture. This knowledge did not significantly contribute to participants' incidental learning of single words and collocations. Different from Dang et al. (2022), Teng and Cui (2023) argued for the importance of prior vocabulary knowledge in learning single words and collocations. Peters and Webb (2018) also underlined the importance of prior vocabulary knowledge in incidental vocabulary learning from viewing L2 TV programs. Puimège and Peters (2020) found similar results for learning collocations. In the present study, this type of knowledge significantly contributed to delayed posttest scores in form and meaning recognition under the caption-viewing condition. These contradictory findings could be attributed to several factors. First, the present study and that by Peters and Webb (2018) examined nonacademic genres (e.g., L2 TV programs), whereas Dang et al. (2022) considered an academic genre. Learners may need specialized vocabulary to follow academic lectures. The updated VLT did not include specialized terms but tapped into knowledge of form-meaning links for words occurring between the 1000-word and 5000-word levels. Second, Dang et al. (2022) explored the effects of prior knowledge on recognizing collocations. The present study and Peters and Webb (2018) investigated the form and meaning of single words.

6.3. *WM and incidental vocabulary learning*

The results determined that WM significantly affected incidental vocabulary learning via different input modes apart from the immediate form recognition test in the reading-while-listening condition. Overall, WM was crucial for incidental vocabulary learning in all input conditions. Gass et al. (2019) stated that WM's impact on captioned video comprehension varied between learners (e.g., Spanish L2 learners' WM capacity did not influence comprehension; this capacity had a moderate effect on ESL learners). The authors contended that although participants used captions irrespective of WM capacity, caption use partly depended on learners' WM capacities and L2 proficiency levels. WM's significant effects on incidental vocabulary learning from listening, reading, reading while listening, and caption viewing can be rationalized thusly: WM assessment is a verbal memory task that involves learners' storage and processing ability while reading. Such tasks may be more directly related to people's language proficiency and reading comprehension. Therefore, individual differences in WM account for variation in incidental vocabulary learning from the examined input modes.

Montero Perez (2020) also documented WM's role in incidental vocabulary learning from viewing a documentary. Complex WM positively correlated with incidental vocabulary learning from viewing. The chosen RST in the present study measured complex WM as well, wherein learners were expected to hold information (i.e., end-of-sentence words) in phonological memory while manipulating it. This task calls for applying background information to judge whether a sentence is plausible. Participants who performed better on the complex WM tests were more likely to score higher on incidental vocabulary learning from the examined input modes because their higher complex WM scores presumably correlated with a "greater ability to focus, divide, and switch attention among various task demands" (Re'vész, 2012, p. 123). Complex WM tasks, which measure one's capacity to store and manipulate information in memory, appear essential to allocating one's attentional resources to input for incidental vocabulary learning. People can then notice word forms and infer new words' meanings. However, Malone (2018) pointed out that reading while listening places higher WM demands on learners than reading alone. Including an aural component, as in the reading-while-listening condition, increases this memory load. The findings have complemented prior studies by showing that complex WM significantly affects incidental vocabulary learning from reading, listening, reading while listening, and caption viewing. The present study did not include phonological short-term memory,

the results cannot verify its role in incidental vocabulary learning through different input modes. Montero Perez (2020) asserted that this type of short-term memory is less important for incidental vocabulary acquisition from viewing. However, Teng and Zhang (2023) highlighted the significance of phonological short-term memory and complex WM for vocabulary learning via multimodal input. The different varied roles of WM in incidental vocabulary learning in the captioned viewing context were also noted in Teng (2023ab) and Teng and Cui (2023). Future research on this topic is necessary.

It is also crucial to acknowledge that we may have oversimplified WM by concentrating on verbal WM. To enhance theoretical understanding, scholars should contemplate how the visuospatial sketchpad operates in conjunction with verbal WM. Integrating verbal and visuospatial components is vital to fully grasp learners' WM capacity. Considering the interplay between these WM components will provide a more nuanced view of how learners process information through different modalities. By acknowledging the roles of verbal and visuospatial WM, researchers can more fully unveil the cognitive processes involved in incidental vocabulary learning. Conclusions may yield a more robust theoretical framework regarding learners' WM capacity.

7. Conclusions

Overall, the results underscore the effectiveness of incidental vocabulary learning through diverse modes of input, such as listening, reading, reading while listening, and viewing videos with captions. Notably, the data highlighted the significant advantage provided by the use of on-screen text in captioned videos for facilitating incidental vocabulary learning. Furthermore, the study delved into the nuanced impact that prior vocabulary knowledge and working memory capacity have on the learning process across different input modes. Prior vocabulary knowledge and working memory capacity play critical and distinct roles in the learners' ability to learn and retain both the form and meaning of new words across different input modes. The interplay between prior vocabulary knowledge, working memory, and the mode of input presents a complex yet insightful picture of how incidental vocabulary learning occurs.

This study has several limitations. First, repeated viewing was not assessed across input modes but may influence how WM and prior vocabulary knowledge affect incidental vocabulary learning. Second, word-related factors (e.g., cognateness, word relevance, and contextual clues surrounding target items) were not evaluated; adding these characteristics may enhance the understanding of incidental vocabulary learning. Consideration must also be given to the nature of target words. The acquisition of different word types, such as abstract nouns, verbs, and adjectives, may vary. Certain words might be imperative for video comprehension (and for capturing students' interest); other words could affect conceptual understanding less and cause students to be less attentive to them. Third, the chosen video was an L2 documentary. Future studies can include additional video genres, as lexical coverage may differ by category. Fourth, follow-up work can examine productive knowledge, which is also a crucial part of vocabulary learning. Finally, learners' look-up behavior when encountering target words was not addressed. Eye-tracking technology could be deployed in the future to expand our sense of incidental vocabulary learning from different input modes.

Despite its limitations, this study offers theoretical and pedagogical implications for incidental vocabulary learning. Theoretically, the findings offer insights into learners' WM resources when processing audio and visual input. Exploring dual processing mechanisms reveals how people concurrently engage with auditory and visual stimuli. The interplay between WM and dual processing is vital to the effectiveness of incidental vocabulary learning from audiovisual materials. This theoretical framework encourages contemplation of how learners use their WM resources when simultaneously processing information from different modalities. Findings could inform later research on the cognitive aspects of language acquisition.

In terms of pedagogical implications, the results underscore the need to combine spoken, written, and audiovisual input with captions to improve incidental vocabulary learning. This recommendation aligns with the literature stressing reading as a primary avenue for such knowledge acquisition. Teachers can leverage this trend by integrating audio support for reading and captions for video viewing; this approach should foster incidental vocabulary learning in foreign language contexts. With the growing accessibility of audiobooks as well as captioned videos (e.g., on YouTube and Netflix), educators can easily bring these resources into classroom and independent EFL learning. Second, even with advances in incidental vocabulary learning, the participants' vocabulary gains were still relatively limited. Teachers may need to diversify their instructional materials across input modes by tailoring items to students' current vocabulary knowledge. Finally, the present study emphasizes the importance of considering learners' WM resources in incidental vocabulary learning. Teachers should account for students' WM capacities when it comes to recalling word forms and meanings. While learners use WM resources across various input modes, implementing these modes does not necessarily exclude students with lower WM abilities. Scholars should further explore incidental vocabulary learning across input modes. Investigating high- and low-WM participants separately could fortify our understanding of how WM shapes vocabulary acquisition.

CRedit authorship contribution statement

Mark Feng Teng: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The author declares that there are no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by National Social Science Fund of China, entitled cross-sectional effects and longitudinal development of working memory and vocabulary acquisition (Grant number: 22BYY182).

References

- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York, NY: Academic Press.
- Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using Eigen and Eigen. URL <http://CRAN.R-project.org/package=lme4>. R package version 0.999375-33.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136–163.
- Chang, A. C. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching*, 21, 43–64.
- Conway, A., Kane, M., Bunting, M., Hambrick, Z., Wilhelm, D., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Dang, T., Lu, C., & Webb, S. (2022). Incidental learning of collocations in an academic lecture through different input modes. *Language Learning. Online advance publication*. <https://doi.org/10.1111/lang.12499>
- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42, 499–523.
- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language, Learning and Technology*, 23(2), 84–104.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61, 355–382.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223.
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal*, 104(3), 550–565.
- Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squared means. Retrieved from <https://CRAN.R-project.org/package=emmeans>.
- Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40, 651–675.
- Meara, P. (1992). EFL vocabulary tests. *Clearing House*.
- Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*. <https://doi.org/10.1017/S0272263119000706>. Online preprint.
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language, Learning and Technology*, 18, 118–141.
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2), 1–26.
- Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists [Data. file] Version 1.0.0. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21, 381–402.
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22, 31–55.
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 169, 142–168.
- Peters, E., Heynen, E., & Puimege, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148.
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2. television and factors that affect learning. *Studies in Second Language Acquisition*, 40, 551–577.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1–28.
- Puimege, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 42, 525–549.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>. ISBN 3-900051-07-0.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62, 93–132.
- Teng, M. F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, 12(3), 274–288.
- Teng, M. F. (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. <https://journals.sagepub.com/doi/10.1177/1362168819829026>.
- Teng, M. F. (2021). *Language learning through captioned videos: Incidental EFL vocabulary acquisition*. New York: Routledge.
- Teng, M. F. (2022). *Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors*. *System*. <https://doi.org/10.1016/j.system.2022.102736>.
- Teng, M. F. (2023a). *Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory*. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2023.2173613>.
- Teng, M. F. (2023b). Incidental vocabulary learning from captioned videos: Learners' prior vocabulary knowledge and working memory. *Journal of Computer Assisted Learning*, 39(2), 517–531. <https://doi.org/10.1111/jcal.12756>
- Teng, M. F. (2023c). Incidental vocabulary learning from captioned video genres: Vocabulary knowledge, comprehension, repetition, and working memory. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2023.2275158>
- Teng, M. F., & Cui, Y. (2023). Comparing incidental learning of single words and collocations from different captioning conditions: The role of vocabulary knowledge and working memory. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12910>.
- Teng, M. F., & Mizumoto, A. (2023). The role of spoken vocabulary knowledge in language minority students' incidental vocabulary learning from captioned television. *Australian Review of Applied Linguistics*, 46(2), 253–278. <https://doi.org/10.1075/aryl.22033.ten>.
- Teng, M. F., & Zhang, D. (2023). The associations between working memory and the effects of multimedia input on L2 vocabulary learning. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 61(3), 1021–1049. <https://doi.org/10.1515/iral-2021-0130>.
- Teng, M. F. (2024). Incidental vocabulary learning from listening, reading, and viewing captioned videos: Frequency and prior vocabulary knowledge. *Applied Linguistics Review*. <https://doi.org/10.1515/applirev-2023-0106>
- Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498–505.
- van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the Reading Span Test in different languages. *European Journal of Psychological Assessment*, 24, 35–42.
- Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave.
- van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41, 609–624.

- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24, 56–89.
- Vidal, K. (2011). A Comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61, 219–258.
- Waring, R., & Takaki, M. (2003). *At what rate do learners learn and retain new vocabulary from reading a graded reader? Reading in a Foreign Language*, 15 pp. 130–163).
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245.
- Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). London: Routledge.
- Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review*, 68, 267–290.
- Webb, S., & Chang, A. C. (2014). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667–686.
- Webb, S., & Chang, A. C. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37, 651–675.
- Webb, S., & Chang, A. C. (2022). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 44, 35–56.
- Webb, S., Newton, J., & Chang, A. C. (2013). Incidental learning of collocation. *Language Learning*, 63, 91–120.
- Webb, S., Sasao, Y., & Balance, O. (2017). The updated vocabulary levels test. *ITL - International Journal of Applied Linguistics*, 168, 33–69.
- Wen, Z., Borges Mota, M., & McNeill, A. (2015). *Working memory in second language acquisition and processing*. Bristol, UK: Multilingual Matters.
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language, Learning and Technology*, 14, 65–86.

Mark Feng Teng, Ph.D., is Associate Professor at Macao Polytechnic University. He was the recipient of the 2017 Best Paper Award from the Hong Kong Association for Applied Linguistics (HAAL), 2023 Best Paper Award in social sciences from Education Ministry in China. His research portfolio mainly focuses on computer-assisted vocabulary learning, and L2 writing from the perspective of metacognition. His publications have appeared in international journals, including *Applied Linguistics*, *TESOL Quarterly*, *Language Teaching Research*, *System*, *Applied Linguistics Review*, *Computer Assisted Language Learning*, *Computers & Education*, *Foreign Language Annals*, and *IRAL*, among others. His recent monographs were published by Routledge, Springer, and Bloomsbury. He also edited and co-edited special issues for international journals, including *Journal of Writing Research*, *Studies in Second Language Learning and Teaching*, and *TESOL Journal*.