



This work is licensed
under a Creative
Commons Attribution
4.0 International
License.

Incidental learning of multiword expressions from bilingual captioned videos for language minority students

Mark Feng Teng

*Faculty of Languages & Translation, Macao Polytechnic University, Macau SAR, China
markteng@mpu.edu.mo*

This study aimed to examine incidental learning of multiword expressions (MWEs) from bilingual captioned videos among language minority students. The participants consisted of 403 language minority university students enrolled in the Preparatory Year Programme (PYP) in the Southern part of China. The study evaluated participants' vocabulary knowledge (VK) by examining both its breadth and depth, alongside their working memory (WM) capacity in tasks involving phonological and complex processing. Additionally, a test on meaning recognition of MWEs was administered. The results revealed that: (1) bilingual captioned videos facilitated the incidental learning of MWEs; (2) both VK and WM capacity were significant predictors of incidental learning of MWEs from bilingual captioning and; (3) VK mediated the relationship between WM and incidental learning of MWEs at the level of meaning recognition. These findings underscore the benefits of incorporating bilingual captioning in learning MWEs for language minority students. However, it is crucial to consider the students' prior VK and WM capacity.

Keywords: vocabulary knowledge, incidental vocabulary learning, bilingual captions, working memory, multiword expressions

Introduction

China is a nation characterized by its multi-ethnic, multilingual, and multicultural composition. The Chinese government acknowledges the important role that its ethnic minorities play in enriching the country's diverse population (Feng & Sunuodula, 2009). With 56 ethnic groups and a variety of languages, Mandarin serves as the national common language spoken by the Han Chinese. However, beyond the Han Chinese, the largest ethnic group, other ethnic groups maintain their own languages, with varying levels of proficiency in Mandarin Chinese, let alone English, a foreign language in China. When ethnic

minority students, e.g., Miao. Li, Yao, Zhuang, and Miao enter university, they begin their academic journey by enrolling in the Preparatory Year Programme (PYP) run by the universities themselves. This one-year curriculum serves as a foundational bridge between high school and university-level studies, addressing both academic and social challenges that students may face during this critical period of adjustment. The PYP is designed to equip students with essential skills, such as advanced Chinese and English language proficiency, critical thinking, and study strategies, while also familiarizing them with the expectations and rigor of higher education. This program specifically helps language minority students who are identified as lacking sufficient proficiency in both Chinese and English. Minority students in China face unique challenges due to limited access to educational resources and a lack of exposure to language input in their native languages as well as in English (Teng & Zhang, 2022, 2024).

In the present study, the focus was on language minority students enrolled in the PYP. This research aimed to investigate how bilingual captioned videos could facilitate incidental learning of multiword expressions for this specific group of learners. Additionally, we explored the mediation effects of participants' prior VK on WM and their vocabulary learning outcome in word meaning recognition. By examining the effectiveness of bilingual captioned videos as a learning tool, this research contributes to research by Teng and Mizunmoto (2023) through addressing the challenges in learning English multiword expressions (MWEs) faced by ethnic minority students in the context of limited educational resources and language input. Through this research, we strive to contribute to the understanding of ethnic minority students' incidental learning of MWEs through bilingual captioning, which is an important dimension of vocabulary learning.

Literature review

Understanding incidental vocabulary learning and multiword expressions

Learning vocabulary is a crucial aspect of language acquisition. VK enables learners to comprehend and produce meaningful speech and written texts (Nation, 2022). Learning MWEs is crucial in vocabulary acquisition and language proficiency development (Majuddin et al., 2021). MWEs, also known as collocations or lexical chunks, are fixed combinations of words that commonly occur together in natural language use (Schmitt, 2010). Recognizing and understanding multiword expressions is essential because they play a significant role in native-like language production and comprehension (Nation, 2022). Providing ample exposure to authentic language materials, such as video clips that contain rich examples of multiword expressions, can enhance learners' familiarity with these lexical chunks (Majuddin et al., 2021).

In the realm of second language (L2) vocabulary research, there has been a growing focus on incidental vocabulary learning while considering modality of input (Sydorenko, 2010). This approach involves acquiring VK unconsciously through engaging in meaning-focused tasks, including, but not limited

to, reading, listening, and viewing (Teng, 2024a; Webb, 2020). To achieve effective comprehension of spoken texts, learners typically require knowledge of approximately 3,000 word families (van Zeeland & Schmitt, 2013). Furthermore, a broader vocabulary of 8,000 to 9,000 words becomes necessary for understanding written texts (Nation, 2006). While explicit vocabulary instruction has shown effectiveness in second language vocabulary learning (Schmitt, 2010), the challenge lies in the limited class time available for learners to acquire such a substantial number of words (Webb, 2020). This challenge becomes more pronounced for language minority students, who face additional barriers such as a lack of language input, limited English curriculum hours, and a scarcity of opportunities to develop vocabulary incidentally (Teng & Zhang, 2022, 2024).

Given these circumstances, it becomes imperative to provide learners with opportunities for incidental vocabulary learning. Supportive environments can help learners enrich their word knowledge, particularly concerning multiword expressions. Incidental vocabulary learning offers a complementary approach to explicit instruction, allowing learners to encounter and internalize vocabulary in authentic contexts. This approach not only enhances learners' lexical repertoire but also promotes their ability to comprehend and employ language naturally.

Incidental vocabulary learning from captioned viewing

Based on the Cognitive Theory of Multimedia Learning (Mayer, 2009), incorporating both verbal and visual representations in multimedia materials can optimize the cognitive capacity for processing and understanding information. This theory posits that individuals have separate channels for processing auditory/verbal information and visual/pictorial information. When these channels are simultaneously engaged, learners can benefit from the integration of these two modalities, leading to enhanced comprehension and knowledge retention (Mayer, 2009). In the context of vocabulary learning, the inclusion of captions or subtitles in audiovisual materials aligns with the principles of the Cognitive Theory of Multimedia Learning (Teng, 2021). By presenting both spoken language and written text simultaneously, learners are provided with multiple sources of information that reinforce each other. Verbal representations in the form of spoken language are processed through the auditory channel, while visual representations in the form of captions or subtitles are processed through the visual channel. This dual processing allows learners to allocate their attention to both auditory and visual cues, facilitating the encoding of information for vocabulary learning.

A significant amount of research has assessed the advantages of captions – on-screen text matching the audio language – compared to input without captions. Recent studies indicate that audiovisual content with captions can facilitate incidental vocabulary acquisition (Fievez et al., 2023; Montero Perez et al., 2014). This is because captions provide opportunities for adoption and adaptation, through which learners could extract usable language while viewing (Vanderplank, 2016). Wang and Pellicer-Sánchez (2022) carried out an

eye-tracking study to examine the effectiveness of bilingual subtitles versus captions, subtitles, and no subtitles in facilitating incidental vocabulary learning. The findings demonstrated that bilingual subtitles provided an advantage over captions for recognizing meanings and surpassed first language (L1) subtitles in aiding meaning recall. Additionally, the study revealed that the extent of attention given to the English target words was a predictor of learning gains. Other studies (Markham, 1999; Neuman & Koskinen, 1992; for a meta-analysis, see Montero Perez et al., 2013) also highlighted the role of captioned viewing in vocabulary learning. Teng (2019) supported that incidental uptake of MWEs, like collocations, from captioned audiovisual materials is possible. He included a comparison of full captions and keyword captions, and the target items were all MWEs (verb–noun collocations). Captions may help learners notice MWEs, leading to the uptake of lexical composition (or form) of MWEs. Teng and Cui (2024) compared the learning gains of single words and collocations while examining how VK and WM affect the vocabulary learning outcomes under different captioning conditions: full captions, keyword captions, and no captions. Full captions were the most effective condition for improving the learning of both single words and collocations.

Armed with the above research, captioned viewing has been shown to have positive effects on vocabulary learning, including MWEs. The simultaneous presentation of spoken language and written text in captions enhances learners' comprehension and provides additional support for word recognition, meaning clarification, and spelling. Learners can associate the auditory information with the corresponding written form, reinforcing the connection between the spoken word and its written representation (Teng, 2021). Moreover, the visual representation of words in captions helps learners develop orthographic awareness and expand their visual recognition of vocabulary items.

The role of working memory (WM)

WM is described as a system with limited capacity that is essential for the concurrent storage and processing of perceptual inputs necessary for complex cognitive activities (Baddeley, 2007). The WM model includes a central executive, which functions as the system's control hub, along with three subsidiary components: the phonological loop, the visuospatial sketchpad, and the episodic buffer.

The phonological loop handles the short-term storage and manipulation of verbal or auditory data. It comprises the phonological store, which briefly retains auditory information, and the articulatory rehearsal process, which keeps this information active through subvocal repetition. Meanwhile, the visuospatial sketchpad is for the temporary storage and manipulation of visual and spatial data, enabling individuals to mentally visualize and manipulate images, objects, and spatial arrangements. A more recent addition to the model is the episodic buffer, which serves as a temporary storage area that integrates information from the phonological loop, the visuospatial sketchpad, and long-term memory. It provides a workspace for the integration of information and

facilitates the transfer of information to long-term memory. Based on this WM model, incorporating audio-visual input in the form of simultaneous verbal and pictorial information can have potential benefits for knowledge processing. By presenting information in both verbal and visual formats, multimedia materials provide learners with multiple modalities through which they can encode and process information.

Second language researchers have extensively examined the central executive and the phonological loop components of WM. The central executive acts as the control center, managing and coordinating the three subsystems. It integrates information from different sources and allocates attentional resources effectively (Baddeley, 2007). To evaluate the central executive, tasks like reading span and operation span are frequently used. In the reading span task, participants must remember the final words of sentences that progressively increase in length. The operation span task involves solving simple equations and recalling words presented afterward. The phonological loop is vital for the temporary storage of auditory information, which is essential for tasks such as word recognition, speaking fluency, and reading comprehension (Ellis & Sinclair, 1996; French, 2006). This component is often assessed through tasks like forward digit-span or nonword span, where participants are required to recall sequences of numbers or nonwords that grow progressively longer.

Research has consistently shown a link between WM capacity and vocabulary acquisition. For example, Martin and Ellis (2012) conducted a study with 50 monolingual native English speakers to explore how phonological short-term memory (PSTM) and WM affect explicit learning of vocabulary and grammar. Their results indicated that both PSTM and WM separately impacted vocabulary learning in a second language. Similarly, Montero Perez (2020) examined the connection between WM existing VK, and vocabulary acquisition using audiovisual input with 63 advanced French learners. WM capacity was assessed using tasks like forward digit span, backward digit span (to measure PSTM), and operation span (to evaluate complex WM). The study found that learners with higher complex WM capacity showed greater incidental vocabulary gains in tasks involving form and meaning recognition. However, PSTM did not significantly affect vocabulary learning in this context. In contrast, Teng and Zhang (2023) found that both complex and phonological WM significantly contributed to vocabulary acquisition and retention under three input conditions: Definition + Word information + Video, Definition + Word information, and Definition only. Teng (2025) focused on young learners' incidental vocabulary learning from three types of captioning: glossed captions, full captions, and keyword captions. The study revealed that phonological WM was crucial for incidental learning related to form recognition and meaning recall, whereas complex WM did not play a significant role. Different results, however, were depicted in Teng and Cui (2024) when examining the incidental learning of single words and collocations, highlighting that both phonological and complex WM were significant factors in acquiring new language items. Teng (2024b) investigates how captioned videos in three genres – comedy, education, and documentary – support incidental vocabulary learning, focusing on the roles of

L2 proficiency, WM, and language aptitude. Comedy videos showed the highest vocabulary gains, followed by education, while documentaries had minimal impact, likely due to their complex content. WM and aptitude indirectly influence vocabulary learning when considering the mediating role of proficiency.

Overall, the findings reviewed above highlight the role of WM in vocabulary learning across different learning conditions. Both phonological and complex WM have been found to play distinct roles in this process. However, the existing research has not yet fully explored how WM specifically influences incidental vocabulary learning outcomes in the context of bilingual captioning. Therefore, there is a need to further investigate the intricate relationship between WM and vocabulary acquisition within the framework of bilingual captioning, particularly for minority students.

The role of vocabulary knowledge

VK is a fundamental component of language proficiency and serves as a strong predictor of language skills such as reading (Anderson & Freebody, 1981) and listening comprehension (Stæhr, 2009). Prior VK has also been shown to significantly impact incidental vocabulary acquisition. Horst et al. (1998) investigated the link between VK, assessed using the Vocabulary Levels Test (VLT; Nation, 1983), and vocabulary gains from reading and listening materials. Their findings indicated a notable correlation between existing VK and learning outcomes, with the participants possessing higher vocabulary levels achieving more substantial and consistent gains. Similarly, Webb and Chang (2015a) found significant relationships between prior knowledge, evaluated through pretests of target vocabulary and VLT scores, and vocabulary gains from extensive reading. This suggests that learners with a more extensive vocabulary base are more likely to experience significant improvements in vocabulary acquisition during extensive reading activities. These studies underscore the importance of prior VK in enhancing vocabulary learning outcomes. Learners with a strong foundational vocabulary are better positioned to understand and acquire new words, highlighting the necessity of incorporating vocabulary instruction and the development of robust VK into language learning programs.

Numerous studies have examined how audiovisual input affects vocabulary learning, particularly focusing on the role of existing VK. Peters and Webb (2018) explored how prior VK influences vocabulary acquisition when watching a full-length TV program. They found a positive correlation between existing VK and the ability to recognize and recall meanings. Similarly, Montero Perez (2020) investigated the impact of prior VK on incidental vocabulary learning from videos with captions. This study also showed that learners with a stronger vocabulary foundation were more successful in acquiring new words from video content. In contrast, Fievez et al. (2023) studied how vocabulary size affects learners' ability to recall and recognize the meanings of target words. Their results suggested that vocabulary size had a limited effect on learning outcomes, implying that other factors might be more influential in vocabulary acquisition. Teng and Mizumoto (2023) found that language minority students

in Australia benefit from their VK during incidental learning through captioned videos. They reported that the depth of VK significantly predicts incidental learning of both form and meaning, beyond the contribution of vocabulary breadth. Rodgers (2013) conducted research on the relationship between vocabulary size and vocabulary learning gains from watching a TV series. Surprisingly, this study found no significant link between vocabulary size and vocabulary acquisition, indicating that vocabulary size alone might not be a reliable predictor of learning success through audiovisual media. These studies primarily focused on L2 learners and their vocabulary acquisition. However, there is a lack of research on how VK in both the first and target languages influences vocabulary uptake through audiovisual input with bilingual captions among language minority students. Investigating this area could provide valuable insights into the cross-linguistic transfer of VK and its impact on learning outcomes in bilingual educational settings.

Rationale for the present study

In sum, the existing literature suggests that WM significantly influences vocabulary gains through watching videos, though it remains unclear how different types of WM (e.g., PSTM and complex WM) affect vocabulary uptakes in terms of MWEs. In addition, the role of prior VK in incidental vocabulary learning through bilingual captioned videos needs further exploration. The present study contributes to previous studies in exploring the mediation effects of VK on WM and incidental learning of MWEs among language minority students. The present study aims to address the following research questions:

1. Does viewing bilingual captioned videos lead to enhanced performance in the incidental learning of MWEs for language minority students?
2. What are the relative contributions of WM and prior VK to the incidental learning of MWEs from bilingual captioned videos?
3. Does WM have direct or indirect effects on the incidental learning of MWEs through the mediation of English and Chinese receptive VK?

Methods

Participants

This study involved a total of 403 language minority students (Mage = 18.92, SD = 1.02). The participants included 200 women and 203 men. They were from the PYP administered at three universities in southern China. The sample, representing the largest concentration of language minority students in the southern part of China, consisted of 45% Li, 30% Yao, 20% Zhuang, and 10% Miao. Li, Yao, Zhuang, and Miao are main ethnic minority groups in the southern part of China. Identified by a needs assessment as an “at risk” target population, the minority students were at least 2 to 3 years below grade level as measured by a standard national university entering examination. Most of these students had received sporadic education in rural areas, according to

family accounts. All the participants reported that their first language was their local language: Miao, Li, Yao, or Zhuang. They received formal instruction for Mandarin Chinese after entering primary school and English after secondary school.

Upon entrance into the PYP, each student was given the standard English proficiency test administered by the education bureau. As a criterion-referenced test, this test assesses four areas of English proficiency: reading comprehension, vocabulary and structures, listening, and writing. Scores from this test indicated that 245 students in our sample were at the basic level, 134 were at the intermediate level, and 24 were at the advanced level. When profiling with the Common European Framework of Reference for Languages (CEFR), 245 students were at the A2 level, 134 were B1 level, and 24 were B2 level.

All students were enrolled in various configurations (depending on their subject needs) of a transitional PYP. One important policy in China is to lower entry requirements for ethnic minority students at all school levels. The provision of PYP by universities and colleges is to help those students get prepared for the subsequent learning in the degree program. PYP classes normally last one or two years. The main purpose is to enable ethnic minority students to develop further in terms of moral education, intellectual development, and language proficiency, to lay a good foundation for their higher education learning. In the English subject, Mandarin Chinese is used as an instructional medium when needed. Students are mainstreamed to English as soon as sufficiency skills allow them to follow English instruction.

Materials

There are a few factors to be considered in selecting input materials. First, the video had to be appealing enough for motivational purposes. As learners' language proficiency level is one concern, the video should sustain the learners' interest enough for them to watch the video from start to finish. Another important factor that influenced the choice of input material was the lexical coverage and speed of dialogue. The video content should be easy to follow in terms of lexical coverage and speed of dialogue while at the same time containing a certain number of words unlikely to be familiar. The video, *Friends Power Hour*, available on YouTube, was a compilation of interesting episodes of the TV series, *Friends*. The video is one hour, three minutes, and three seconds long. This video was selected as the material because this comedy genre contains a high number of everyday spoken formulaic sequences. VocabProfile (<https://www.lex tutor.ca/>) was adopted to determine the lexical profile. The script contained 6,552 running words. The 1,000-, 2,000-, and 3,000-word families covered 87.18%, 90.24%, and 91.69% of all running words in the script, respectively. The participants were administered the updated Vocabulary Levels Test (VLT) (Webb et al., 2017). Their mean scores indicate a mastery of the 2,000-word level. In a recent study (Durbahn et al., 2020) on the relationship between vocabulary and viewing comprehension, learners need a lexical coverage of around 90% for adequate viewing comprehension. The participants'

VLT scores (see Results section) suggest that they were able to follow the video. We re-edited the subtitled texts through Amara, an online software for captioning. We included both English and Chinese (see Figure 1) because the original captions were machine-transcribed, which were not correct in some way. In a pilot study, a group of 20 ethnic minority students reported that the topic of the video was interesting, and that the difficulty level of the video was appropriate for them. Although this is a famous TV program, the participants reported that they had not watched it because they were from rural areas.



Figure 1. An example of the video with bilingual captions

Target items

Authentic materials offer opportunities for learners to encounter MWEs of diverse kinds. MWE learning in the present study focused on certain parts of speech, such as adjective + noun collocations, verb + noun collocations, verb-particle construction, and a compound, etc. Although we did not attempt to prioritize certain syntactic categories over others, not all MWEs are “equal” in the way that they are processed (Columbus, 2010). This also applies to the audiovisual materials we have used. We first screened for MWEs that were likely to be unfamiliar to the students in the present study. We also consulted the Corpus of Contemporary American English (COCA) to ascertain that the target MWEs were well established. All potential MWEs were frequently used expressions in the corpus, and all MWEs were found to have a mutual information (MI) score, a measure of collocational strength, of at least 3.0. We compiled a list of potential target MWEs from the video, and invited two teachers of the learners to evaluate whether they believed the learners had prior knowledge of these items. Based on the information provided by the teachers, we selected 50 MWEs as target items (see Table 1). Each MWE only occurred once in the video.

Table 1. A list of 50 target multiword expressions

Multiword expressions	Multiword expressions
Wrinkly skin	Break up
Wreck a day	Hang on
Tug at (something)	Move on
A boatload of	Make up
An imaginary childhood friend	Take advantage of (somebody)
Swirl around	Take up
Turn down (something)	Work out
Defend against (something)	Bear with (someone)
Squint at (something)	Visually accessible to (someone)
Play scrabble with (somebody)	No wonder
Screw (something) up	Wrap up
Ramble back to (somewhere)	Fill (somewhere) with (something)
Mighty roar	Catch fire
Ironic idea	Misconstrue (somebody's) comments
Indignant reply	Freak out
Drastic measures	Mix (something) up
Share custody of (somebody)	Put (something) on
Assign heads to (something)	Jazz something up
Awkward moment	Drunken mistakes
Acts of aggression	Arouse (somebody's) interest
Bumpy road	In a good mood
Gain perspectives	A petty man
Take responsibility for	Become extinct
Cleanse the body	Invisible superhero
End up with	Keep (someone) on their toes

Vocabulary learning measure

Of principal interest was knowledge of word meanings of the MWEs. The reason for this is that although captions can render language forms more salient, it may be too challenging for the ethnic minority students to remember the form through a short period of video viewing. This evidence was shown in a pilot study of 20 students mentioned above. None of the students could write down the word form correctly after viewing the videos. So, the focus was turned to meaning. In addition, the bilingual captions in the present study clarify the meanings of the lexical items. The initial idea was to understand the multiple aspects of word meaning, including recognizing and recalling word meanings. Only word meaning recognition was included as the pilot participants expressed difficulties in writing down the word meanings.

To assess the recognition of MWEs, a written multiple-choice test was administered. Each test item included four options in Chinese: the correct answer (key), three distractors, and an “I don’t know” option to discourage guessing.

This test served as both a pretest and an immediate posttest, with the order of items randomized in both instances. Participants were given 20 minutes to complete the test. To ensure the reliability of the test, the internal consistency of all items was examined using Cronbach's alpha. The results indicated good reliability, with coefficients above .80 for all tests: $\alpha = .84$ and .87 for pretest and posttest meaning recognition, respectively. Participants received one point for each correct answer, with a maximum possible score of 50 points.

Instruments

Prior vocabulary knowledge

English Vocabulary Levels Test. The Vocabulary Levels Test (VLT) is perhaps the most popular test of English vocabulary for L2 learners. It was originally developed by Nation (1983) and later adapted by Schmitt et al. (2001) as a measure of the form-meaning connection of a word at four frequency levels (2000, 3000, 5000, 10000) and an academic level. VLT is a form recognition test (Schmitt, 2010) in which learners are expected to match the correct words with each meaning. Webb et al. (2017) updated the VLT by selecting test items from five-word frequency levels (1000, 2000, 3000, 4000, 5000) based on Nation's (2012) British National Corpus and Corpus of Contemporary American English word lists to better reflect current English. Initial validation demonstrated that the updated VLT is a valid tool to measure L2 VK (Webb et al., 2017), and it was used in the present study.

LexCH. We used LexCH to measure learners' receptive Chinese VK. LexCH is a receptive vocabulary size assessment tool for Chinese learners (Qi et al., 2024). It adopts a Yes/No format, in which learners are asked to indicate whether they know a word or not. To avoid inflation of scores due to guessing behavior, nonwords are included in the test. When a learner responded positively to a nonword, one point was deducted from the final scores. Test items covering a wide range of difficulty levels and with strong discriminative power were selected based on Item Response Theory. Supportive evidence for construct validity was provided in Qi et al. (2024).

Working memory

The assessment of WM was through two tasks on learners' phonological and complex WM. Data collection was through the online Inquisit Web (www.millisecond.com). The participants completed the tests separately.

Phonological working memory. PSTM was evaluated using an English non-word repetition task (Gathercole et al., 2001). This task targeted the storage aspect of short-term memory (Baddeley et al., 2019). Participants were presented with 22 sequences of nonwords, each containing two to six one-syllable consonant-vowel-consonant constructs designed to mimic real words. After

hearing each sequence, participants were required to repeat the nonwords in the exact order they were given. The nonwords included five short and five long vowel sounds, selected from phonemes that are generally familiar to English as a foreign language (EFL) students. Each sequence featured nonwords with a distinct vowel sound, and efforts were made to ensure the consonants within each sequence were as varied as possible. However, due to the limited selection of phonemes and lexical constraints, some sequences with five or six items could not avoid repeated consonants. The sequences increased in length as the task progressed, starting with two-item sequences and incrementally adding more items until reaching six-item sequences. Each nonword was played for about 750 milliseconds, with a pause of approximately 1.5 seconds between the end of one sequence and the start of the next. Participants were already acquainted with this task from a practice session, and the entire assessment was completed in about five minutes. In total, the 22 sequences comprised 100 nonwords. Participants earned one point for each correctly recalled nonword in the correct sequence, with a maximum possible score of 100 points.

Complex WM. This was evaluated using the O-span task (Turner & Engle, 1989; Unsworth et al., 2005). This task was designed to assess both the processing and storage facets of WM. Participants were asked to solve mathematical problems while concurrently trying to memorize a series of unrelated words in their second language. It is important to highlight that WM is a cognitive function that operates independently of language proficiency (Mackey et al., 2010). To reduce the impact of participants' existing VK on the results, simple words were selected for the task. During each trial, participants encountered a math problem followed by a word, presented one after the other. They were instructed to solve the math problem and then read the subsequent target word aloud. For instance, when given " $7 + 9 = ?$ ", participants needed to compute the answer while remembering the next word, such as "garden." These operation-word pairs were organized into sets ranging from two to five items, known as set sizes. The task began with two-item sets and progressed to sets of three, four, and five items. After each set, participants had to recall the words in the order they appeared. Taking into account the cognitive demands on young learners, the task included 80 math problems and 80 target words. During the recall phase, participants clicked on the words they believed they had remembered correctly. To discourage intentional memorization of words and guessing the math solutions, an accuracy threshold of 85% was set for the math calculations (Unsworth et al., 2005). These criteria were aligned with the task's objective of evaluating WM's processing and storage capabilities. Participants became accustomed to the task through three practice sessions. In line with Unsworth et al. (2005), one point was awarded for each correctly recalled word in the right order, with a maximum score of 80 points. Information retention in short-term memory typically lasts a few seconds, and participants were given up to 10 minutes to complete the task.

Procedure

The data collection process spanned four sessions. During the first session, participants were given an overview of the research project. The true aim of incidental vocabulary learning was revealed only after the study concluded. Participants also signed a consent form at this time. In the second session, they completed a vocabulary outcome pretest, which was immediately followed by assessments of vocabulary test and WM test. Two weeks later, in the third session, participants watched a video in a laboratory setting. Before viewing, they were told they would need to complete some exercises afterward and were permitted to take notes during the video. The video was watched once. Surprise tests on vocabulary learning were conducted using paper-and-pencil, with instructions provided in Mandarin Chinese. All participants gathered in a lab to finish the study.

Data analysis

To assess improvements in recognizing word meaning after watching captioned television, relative gains were used. This method is advantageous because it accounts for the different potential knowledge increases among participants, unlike absolute gains. Following the approach of Webb and Chang (2015b), relative learning gains were calculated with the formula: $[(\text{posttest score} - \text{pretest score}) / (\text{number of test items} - \text{pretest score}) \times 100]$. However, it also has limitations as following relative gains, we would not take pretest as a control variable in data analysis.

The analysis began with calculating descriptive statistics and verifying assumptions of both univariate and multivariate normality. Next, Pearson product-moment correlation coefficients were used to examine the relationships between the variables measured. Hierarchical regression analysis was then employed to determine the relative impact of previous VK and WM on the incidental learning of multiword expressions. Finally, mediation analyses were conducted to explore the direct and indirect effects of WM on learning multiword expressions, taking into account VK in both English and Chinese.

Results

RQ 1: Does viewing bilingual captioned videos lead to enhanced performance in the incidental learning of MWEs for language minority students?

Table 2 presents the outcomes of the descriptive analyses and normality tests. The standard deviations indicate that the data had a sufficient range of variability. Skewness confirmed that the data were reasonably symmetric, but it seems that the kurtosis values were a little bit high for the English language proficiency test, which is a limitation for data analysis in the present study. These findings suggest that there were noticeable individual differences among the variables studied. Overall, these results affirmatively answer Research

Question 1, indicating that watching captioned television indeed improves incidental learning of multiword expressions.

Table 2. Descriptive statistics

Variables	N	Min.	Max.	M.	Std.	Skewness	Kurtosis
Phonological WM	403	10	70	45.39	8.004	-1.157	6.620
Complex WM	403	13	44	28.41	7.512	-0.420	-0.685
English Vocabulary Levels Test	403	15	105	70.95	10.227	-1.887	9.751
Chinese receptive vocabulary knowledge	403	33	67	55.08	11.487	-0.785	-1.021
Word meaning recognition (Relative scores)	403	9.09	20.00	14.7313	3.73182	0.158	-1.191
Word meaning recognition Pretest	403	5.00	10.00	7.6489	1.69627	0.080	-1.182
Word meaning recognition Posttest	403	10.00	20.00	15.2978	3.39255	0.080	-1.182

RQ 2: What are the relative contributions of WM and prior VK to the incidental learning of MWEs from bilingual captioned videos?

Table 3 presents the correlation results. The results showed that phonological WM is significantly correlated with complex WM ($r = .191, p < .01$), and both phonological WM ($r = .446, p < .01$) and complex WM ($r = .367, p < .01$) are significantly correlated with word meaning recognition. English vocabulary levels test is significantly correlated with Chinese receptive VK ($r = .311, p < .01$). Both English Vocabulary Levels Test ($r = .456, p < .01$) and Chinese receptive VK ($r = .438, p < .01$) are significantly correlated with word meaning recognition.

Table 3. Correlation results

	Phonological WM	Complex WM	English Vocabulary Levels Test	Chinese receptive vocabulary knowledge	Word meaning recognition
Phonological WM	1	0.091	.299**	.289**	.446**
Complex WM	.191**	1	.366**	.412**	.367**
English Vocabulary Levels Test	.299**	.366**	1	.311**	.456**
Chinese receptive vocabulary knowledge	.289**	.412**	.311**	1	.438**

** Correlation is significant at the 0.01 level (2-tailed).

To address the second question, which examined the relative impacts of existing VK and WM on the incidental learning of word meaning recognition,

hierarchical regression analyses were conducted. Initially, we assessed issues of linearity and multicollinearity. A plot of residuals against predicted values verified the linear relationship among the variables. The variance inflation factors (VIFs) ranged from 1.42 to 1.89, confirming that multicollinearity was not a concern. The outcomes of the hierarchical regressions are detailed in Table 4.

Table 4. Hierarchical regression results

	Block and variable	R^2	Adjusted R^2	ΔR^2	ΔF	β	t
Regression 1	Model 1	.307	.303	.307	88.405		
	Phonological WM					.416	9.953
	Complex WM					.329	7.876
	Model 2	.395	.389	.089	29.250		
	Phonological WM					.298	7.096
	Complex WM					.166	3.704
	English VLT					.242	5.491
	Chinese receptive VK					.208	4.635
Regression 2	Model 1	.305	.302	.305	87.893		
	English VLT					.355	8.086
	Chinese receptive VK					.328	7.470
	Model 2	.395	.389	.090	29.656		
	English VLT					.242	5.491
	Chinese receptive VK					.208	4.635
	Phonological WM					.298	7.096
	Complex WM					.166	3.704

In the initial regression analysis, components of WM were included in the first block, while aspects of prior VK were added in the second block. The second regression model reversed this order to better assess the relative impact of each component when controlling for the other. In the first regression, WM accounted for 30.7% of the variance in incidental word recall learning. When prior VK was considered, it contributed an additional 8.9% to the variance in meaning recognition. Conversely, in the second regression, prior VK explained 30.5% of the variance in word recall learning. WM added an extra 9% to the variance in meaning recognition once prior VK was accounted for. These regression analyses demonstrated that both WM (including phonological and complex WM) and prior VK (encompassing English and Chinese receptive VK) significantly influenced the incidental learning of word meaning for MWEs from viewing videos with bilingual captions.

Figure 2. Mediation model results

To further confirm the direct and indirect effects of WM on word meaning recall, standardised direct and indirect effects as well as total effects were calculated in the path model (see Table 5).

Table 5. standardised total effects, direct effect, and indirect effect on meaning recognition

Word meaning recognition	Direct	Indirect	Total effects
Phonological WM	.30	.06 (via English VLT)	.36
		.07 (via Chinese receptive VK)	.37
Complex WM	.17	.06 (via English VLT)	.23
		.08 (via Chinese receptive VK)	.25
English VLT	.24	/	.24
Chinese receptive VK	.21	/	.21

The findings confirmed both direct and indirect influences of WM, as significant indirect effects were also observed ($p < .001$). Notably, phonological WM exhibited stronger direct effects compared to complex WM, while the indirect effects were comparable for both. As shown by the regression weights in Figure 2, this segment of the analysis illustrated that WM impacts word meaning recognition both directly and indirectly through mediating factors, specifically English and Chinese receptive VK.

Discussion

Overall, the current study provided evidence supporting the effectiveness of incidental learning of multiword expressions through bilingual captioned viewing. The findings highlighted the importance of phonological and complex WM, as well as English and Chinese receptive VK, in the incidental learning process for minority students. Moreover, WM was found to directly and indirectly contribute to word meaning recognition, with English and Chinese receptive VK acting as mediating factors.

Incidental vocabulary learning from bilingual captioning

The results of the study indicated that students made progress in their learning of MWEs through watching bilingual captioned videos, consistent with previous research on measuring single words from bilingual captioning (Wang & Pellicer-Sánchez, 2022) and multiword expressions from captioning (Majuddin et al., 2021; Teng, 2019). One possible reason for the benefit of bilingual captioning in learning multiword expressions is related to the characteristics of these expressions themselves. Multiword expressions often have fixed and non-compositional meanings, which means that their meanings cannot be easily inferred from the individual meanings of their constituent words. By providing bilingual captions, learners have access to both the visual representation of the expressions and their corresponding translations. This input aids learners in understanding the overall meaning of the multiword expressions as cohesive units, rather than relying solely on the meanings of individual words (Teng, 2021).

The Cognitive Theory of Multimedia Learning (Mayer, 2009) may explain the positive impact of bilingual captioning on learners' ability to acquire meaning. This theory suggests that the simultaneous presentation of visual and auditory inputs enhances information processing and memory retention. In the context of bilingual captioning, the written bilingual captions serve as a visual representation of expressions, while the accompanying auditory input from the video reinforces understanding. Together, these elements create a stronger connection that helps learners process and encode meaning more effectively. By reducing cognitive load, as noted by Wang and Pellicer-Sánchez (2022), learners are less likely to feel overwhelmed when engaging with subtitles in two languages simultaneously. This balanced integration of inputs ultimately improves memory retention and recognition of the learned expressions.

WM and VK in incidental learning of MWEs from bilingual captioning

The present study provides insights into the role of WM in incidental vocabulary learning from bilingual captioned videos, specifically at the meaning recognition level. The findings demonstrate that both phonological WM and complex WM are significantly correlated with vocabulary gains. Regression analysis further revealed that WM accounted for 30.7% of the total variance in posttest scores ($R^2 = .307, p < .01$). Additionally, after accounting for WM, prior VK

contributed an additional 8.9% to 9% of the variance in meaning recognition. The findings regarding WM align with previous research by Teng and Zhang (2023) and Teng and Cui (2024), which highlighted the significant role of both phonological and complex WM in multimedia-oriented vocabulary learning performance. However, the present study only partially agrees with Montero Perez's (2020), where complex WM showed significant prediction of posttest scores, while PSTM did not. It is worth noting that Montero Perez's (2020) study did not include captions, highlighting the need to consider the role of captions in vocabulary learning. However, according to Masoura and Gathercole (2005), the importance of PSTM may vary depending on learners' proficiency levels and familiarity with the phonological structure of the target language. Teng (2024b) only supported the indirect role of WM in vocabulary learning in different captioned video genres. It is suggested that low proficiency learners may require a certain lexical threshold or proficiency level to establish referential connections between pictorial and verbal representations in captioned videos, which are crucial for incidental vocabulary learning to occur (Teng, 2022).

Building on Teng's (2025) research regarding individual differences in phonological WM and incidental vocabulary learning across different captioning formats, such as full and keyword captions, the present study enhances the understanding of the roles of both phonological and complex WM in vocabulary acquisition among minority students in bilingual captioning contexts. An unexpected discovery was the prominent role of complex WM, which aligns with Montero Perez's (2020) findings but contrasts with those of Teng (2025). Complex WM, which is crucial for storing and manipulating information, has been shown to predict vocabulary learning outcomes (Kormos & Sáfár, 2008), and was similarly observed in the present study with minority students engaging with bilingual captioned videos. Complex WM appears to aid learners in effectively allocating their attentional resources to understand input from bilingual captioned videos. This cognitive function supports the storage and manipulation of information, allowing learners to process and integrate various linguistic and visual cues present in the videos. Students with higher levels of complex WM are better positioned to efficiently focus their attention on relevant linguistic details, such as word forms and structures, while simultaneously using the contextual cues provided by the visual elements in the videos.

The present study highlights the significant role of prior VK, including both English and Chinese vocabulary, in the incidental learning of vocabulary from bilingual captioned videos. The results demonstrated that learners' capacity to acquire vocabulary incidentally while watching bilingual captioned television was closely linked to their existing VK. Specifically, learners with more extensive VK were able to acquire more vocabulary items compared to those with less VK. These findings are consistent with previous research (Montero Perez et al., 2014), which also underscored the positive effect of vocabulary size on incidental vocabulary learning through captioned viewing. Furthermore, the depth of VK, as discussed by Teng and Mizumoto (2023), was found to significantly influence learners' incidental vocabulary acquisition from captioned videos. Learners with a robust vocabulary base benefit from having a wide array of

words stored in their long-term memory. As a result, when they encounter unfamiliar words in bilingual captioned videos, they can quickly activate and access relevant VK. This allows them to form connections between familiar and new words encountered in the videos and engage in inference, using their existing lexical knowledge to deduce the meanings of unknown words. These cognitive processes enhance their ability to recognize and understand new words more effectively, facilitating the successful acquisition of MWEs.

The mediating role of prior VK on WM and incidental learning of MWEs

The mediation analysis results offer valuable insights into the interplay between prior VK, WM, and vocabulary acquisition from bilingual captioned videos. Firstly, the study found that prior VK directly impacts vocabulary learning, suggesting that learners with a more extensive vocabulary in both Chinese and English are better positioned to comprehend the input materials, notice new words, and infer meanings from both linguistic and situational contexts. This finding is consistent with previous research (Horst et al., 1998; Peters & Webb, 2018; Teng & Mizumoto, 2023) and supports the “Matthew Effect” in vocabulary acquisition, where learners with a larger vocabulary base have an advantage in further expanding their vocabulary through bilingual captioned videos.

Interestingly, the study also discovered that prior VK serves as a significant mediator for WM. WM influences word meaning recognition through the mediation of both Chinese and English VK. This implies that learners with greater WM capacities can achieve better vocabulary learning outcomes by using their existing linguistic knowledge to leverage contextual information and infer the meanings of new words from the bilingual captioned videos. While WM is an important cognitive factor, differences in learners’ prior VK may mediate its impact on vocabulary learning outcomes. Additionally, the findings suggest that complex WM has smaller direct and indirect effects on vocabulary learning compared to PSTM. This may be due to the availability of bilingual captions in the videos. Learners can use the Chinese subtitles to understand the meanings of English words, thereby reducing the cognitive load involved in the vocabulary learning process. Consequently, the role of complex WM in vocabulary learning may be diminished.

Limitations and implications

This study has several limitations that should be noted. First, the vocabulary tests we used were limited to assessing meaning recognition scores, which may not fully encompass all facets of incidental learning of MWEs. Some of the MWEs may not be natural and may not accurately reflect their usage. Second, the participants were language minority students in China with limited experience using online language learning resources, which may restrict the applicability of the findings to L2 learners with similar proficiency levels and language backgrounds. Bilingual captioning might be less effective for advanced

learners who already possess a high proficiency in processing language input. Third, while we conducted immediate posttests to explore the link between processing and learning gains, the lack of delayed posttests means we cannot draw conclusions about the long-term retention of vocabulary learned through captioned videos. Finally, the benefits of bilingual captions for learning MWEs might be enhanced under conditions of intentional learning. In such scenarios, form-focused strategies like textual enhancement or prereading instruction could be employed to direct learners' attention to both the forms and meanings of MWEs in a more structured and deliberate way.

The findings identified offer valuable insights for future research and teaching methods. Firstly, the findings suggest that language minority students can improve their recognition of word meanings in learning MWEs through comprehensible input provided by bilingual captioned videos. These videos create a linguistically rich environment, enabling learners to process and understand word meanings more effectively. As a result, bilingual captioned videos can serve as an effective resource for enhancing MWE learning among language minority students by integrating visual and textual information to aid word comprehension. By engaging with these videos, students receive authentic language input and gain exposure to both the spoken and written forms of the target language. Secondly, the research underscores the crucial role of prior VK in learning vocabulary from captioned videos. It is essential for educators and curriculum developers to emphasize the development of learners' VK as a foundation for successful vocabulary acquisition. By offering diverse instructional activities to expand learners' vocabulary, educators can improve students' ability to understand and learn new words presented in captioned videos. Finally, the study reveals that the influence of WM on vocabulary learning from captioned videos is mediated by learners' existing VK. This indicates that while WM capacity is significant, its effect on learning outcomes may depend on the learner's prior VK. Educators should consider individual differences in WM capacity and adapt their teaching strategies accordingly. Providing support, such as explicit vocabulary instruction, can assist learners with limited WM in effectively processing and retaining vocabulary from bilingual captioned videos.

Acknowledgement

No Artificial Intelligence (AI) tools were used in the first submission. However, ChatGPT was used in the revised version for proofreading the language.

References

- Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie, *Comprehension and teaching: research reviews* (pp. 77–117). Newark, DE: International Reading Association.
- Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

- Baddeley, A. D., Hitch, G. J. & Allen, R. J. (2019). From short-term store to multicomponent working memory: The role of the modal model. *Memory & Cognition*, 47, 575–588. <https://doi.org/10.3758/s13421-018-0878-5>
- Columbus, G. (2010). Processing MWU: Are MWU subtypes psycholinguistically real. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194–210). Continuum
- Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, 88. 102166 <https://doi.org/10.1016/j.system.2019.102166>
- Ellis, N., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, 49(1), 234–250. <https://doi.org/10.1080/71375564>
- Feng, A., & Sunuodula, M. (2009). Analysing language education policy for China's minority groups in its entirety. *International Journal of Bilingual Education and Bilingualism*, 12(6), 685–704. <https://doi.org/10.1080/13670050802684396>
- Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2023). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning*, 36(1-2), 26–51. <https://doi.org/10.1080/09588221.2021.1899244>
- French, L. (2006). *Phonological working memory and second language acquisition*. Lewiston, NY: Edwin Mellen Press.
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology*, 54(1), 1–30. <https://doi.org/10.1080/02724980042000002>
- 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.
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(2), 261–271. <https://doi.org/10.1017/S1366728908003416>
- Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, 60, 501–533. <https://doi.org/10.1111/j.1467-9922.2010.00565.x>
- Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 43(5), 985–1008. <https://doi.org/10.1017/S0272263121000036>
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32, 321–328. <https://doi.org/10.1111/j.1944-9720.1999.tb01344.x>

- Martin, K., & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34(3), 379–413.
<https://doi.org/10.1017/s0272263112000125>
- Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 13, 422–429.
<https://doi.org/10.1080/09658210344000323>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.
- Montero Perez, M. (2020). Incidental vocabulary learning through view video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 42(4), 749–773.
<https://doi.org/10.1017/s0272263119000706>
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118–141.
<http://dx.doi.org/10125/44357>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41, 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, (5), 12–25.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59–81.
<https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press.
- Neuman, S., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27, 95–106.
<https://doi.org/10.2307/747835>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577.
<https://doi.org/10.1017/s0272263117000407>
- Qi, S., Teng, M., & Fu, A. (2024). LexCH: A quick and reliable receptive vocabulary size test for Chinese learners. *Applied Linguistics Review*, 15(2), 643–670. <https://doi.org/10.1515/applirev-2022-0006>
- Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. Unpublished doctoral thesis. Victoria University of Wellington.
- Schmitt, N. (2010). *Researching vocabulary*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1191/026553201668475857>
- Stæhr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(04), 577. <https://doi.org/10.1017/s0272263109990039>
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, (14), 50–73. <http://dx.doi.org/10.125/44214>
- Teng, M. F. (2019). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education*, 142, 103655. <https://doi.org/10.1016/j.compedu.2019.103655>.
- Teng, M. F. (2021). *Language learning through captioned videos: Incidental vocabulary acquisition*. New York: Routledge.
- Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, 105, 102736. <https://doi.org/10.1016/j.system.2022.102736>
- Teng, M. F. (2024a). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, 124, 103381. <https://doi.org/10.1016/j.system.2024.103381>
- Teng, M. F. (2024b). Incidental vocabulary learning from captioned video genres: proficiency, working memory, and aptitude. *Computer Assisted Language Learning*, 1–43. <https://doi.org/10.1080/09588221.2024.2421517>
- Teng, M. F. (2025). Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory. *Computer Assisted Language Learning*, 38(1-2), 206–234. <https://doi.org/10.1080/09588221.2023.2173613>
- Teng, M. F., & Cui, Y. (2024). 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*, 40(3), 973–989. <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/aral.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*, 61(3): 1021–1049. <https://doi.org/10.1515/iral-2021-0130>
- Teng, M. F., & Zhang, L. J. (2022). Development of metacognitive knowledge and morphological awareness: A longitudinal study of ethnic minority multilingual young learners in China. *Journal of Multilingual and Multicultural Development*, 1–17. <https://doi.org/10.1080/01434632.2022.2052301>

- Teng, M. F., & Zhang, L. J. (2024). Ethnic minority multilingual young learners' longitudinal development of metacognitive knowledge and breadth of vocabulary knowledge. *Metacognition and Learning*, 19(1), 123–146. <https://doi.org/10.1007/s11409-023-09360-z>
- Turner, M., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/BF03192720>
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- Vanderplank, R. (2016). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235–250. <https://doi.org/10.1017/s0261444813000207>
- Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*, 72(3), 765–805. <https://doi.org/10.1111/lang.12495>
- Webb, S. (2020). Incidental vocabulary learning. In S. Webb (ed.). *The Routledge handbook of vocabulary studies* (pp.225–239). New York: Routledge.
- Webb, S., & Chang, A. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651–675. <https://doi.org/10.1017/s0272263114000606>
- Webb, S., & Chang, A. (2015b). 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. <https://doi.org/10.1177/1362168814559800>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL-International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>