The Reading Matrix: An International Online Journal
Volume 22, Number 2, September 2022

Comparing the Multidimensional Effects of Oral Versus Written Exercises on the Retention of L2 Vocabulary

Anna C-S Chang
Hsing Wu University

Brandt T-Y Chen
Giraffe English Language Institute

ABSTRACT

This study compared the multidimensional effects of written and oral vocabulary exercises on EFL learners' retention of single-word items (SWI) and multi-word items (MWI). Two teaching experiments (times) were conducted with 54 college students, who studied four themed-based texts, two at a time. A total of 60 unfamiliar vocabulary items, 30 in each experiment, were tested in four dimensions one week after the treatment. Hierarchical linear modeling (HLM) was performed to analyze the data. The results showed that all four fixed variables (times, vocabulary exercises, vocabulary dimensions, and types of vocabulary items) had significant effects on students' retention of SWI and MWI. Overall, students performed better at Time 1 for doing the written exercises; students scored highest for L2 meaning recall, followed by aural form recall, aural meaning recall, and L2 form recall. Students consistently performed better on the MWI than on SWI. Regardless of whether students scored higher on written exercises, a short questionnaire indicated that they perceived oral exercises interesting and helpful. Class discussion revealed that students enjoyed learning MWIs that comprised of familiar words, and various strategies were applied to acquire the MWIs. Pedagogical implications of the results were discussed.

INTRODUCTION

This study aimed at comparing the effects of two types of vocabulary exercises: written exercises versus repeated oral reading exercises. The former has been well-documented; however, the latter is under-researched (Durrant & Schmitt, 2010). Recent research has shown a great interest in the relationship between repetition and vocabulary learning (see Uchihara, Webb & Yanagisawa, 2019 for a review). The general finding for the effect of repetition is the more the better, but the optimal repeating times may vary from different vocabulary dimensions. Given its importance, very little research has explored whether repeated oral reading can produce effects comparable to the effects of written exercises. Repeated oral reading is assumed to offer at least four advantages on vocabulary learning. Firstly, vocabulary learning involves learning its spoken form; if learners orally practice the newly learned items, it may facilitate their knowledge in the linkage between written form and spoken form (. Secondly, oral reading requires learners to read aloud the learned materials, which may improve learners' retention on what is learned. Thirdly,

repetition provides L2 learners with the opportunities for multiple encounters of the same words or same patterns of phrases. Finally, repeated oral reading brings the language to life through voice, which adds more variety to vocabulary exercises, and some studies have shown that L2 learners enjoy this activity very much (Chang, 2019; Shimono, 2019). For these above advantages, it is worthwhile to explore its effects.

LITERATURE REVIEW

Requiring L2 learners to learn lexical items through doing vocabulary exercises after reading a text is a very common practice to increase their vocabulary retention; however, the effects may vary greatly according to which types of exercises L2 learners engage and on which vocabulary dimensions the L2 teachers want to focus. Exercise types are many, but most of them are written practice, such as looking up the meaning of an unknown word, L1-L2 meaning translation, matching a word with a correct definition, gap-filling exercises, or sentence-making. These exercises focused mainly on the form-meaning linkage, the most important aspect that L2 learners consider. Vocabulary knowledge, however, involves many other dimensions: comprehending and producing the spoken word forms (spelling), collocations, use, etc. (see Nation, 2001), but vocabulary exercises including spoken practice have been scarce. This study hence intends to compare the effects between written exercises and repeated oral reading on L2 learners' vocabulary retention, and also how the effects of the two types of vocabulary exercises are influenced by vocabulary dimensions and vocabulary types.

L2 vocabulary exercises on vocabulary retention for different dimensions and types

In this section, two types of vocabulary exercises will be discussed, followed by its effects on different vocabulary dimensions and vocabulary types. The findings of some previous studies into these three areas are summarized in Table 1.

Written exercises

Many studies, to name only a few, have provided empirical evidence for the effects of doing written vocabulary exercises (Folse, 2006; Laufer & Hulstijn, 2001; Laufer, 2003, Lu, 2013); however, results have not been consistent. For example, Folse (2006) investigated the effects of three types of written exercises on ESL university students: one and three fill-in-the-blank exercises, and one original-sentence writing exercise. The study found that students who practiced the target words under the three fill-in-the-blank exercise conditions retained more vocabulary knowledge than they did under the other two exercise conditions. Another study was conducted by Lu (2013) with 122 senior high school students, who read a passage, then did four forms of vocabulary exercises (single blank filling, triple blank filling, blank filling of a summary, and summary writing). The triple blank filling exercises were found to be most effective in the immediate post-tests, but the average test scores dropped substantially in the two-week delayed post-test, which resulted in no significant difference between the four forms of vocabulary exercises. The short survey of students' perceived effectiveness, however, indicated that the composition tasks were the least conducive to vocabulary learning whereas blank-filling tasks were the most effective. Both Lu and Folse's studies indicated that multiple gap-filling exercise is the most effective for vocabulary learning. Sentence- and composition- writing, which require more attention and effort, did not lead to higher vocabulary learning rates. The effects of

the exercises might be affected by the time allotted. When time was not controlled, Webb (2005) found reading target words in sentences more effective than writing the sentences with the target words. Laufer (2019) hence suggests that sufficient time should be allotted for students to finish the activities.

Lu and Folse's studies mentioned above seemed to suggest that number of word retrievals required influences the effectiveness of exercises more than the depth of word processing, but their findings seem to contradict some previous studies on involvement load (Laufer, 2003, with Rozovski-Roitblatv, 2011, 2015; Keating, 2008; Zou, 2017) . For example, Laufer (2003) compared the meaning retrievals under four conditions: reading only and reading plus one of the three written tasks (blank filling, sentence-making, or composing) and found that students who did reading plus a written task scored higher than those doing a reading plus a blank-filling task. Laufer's findings were supported by Keating (2008), who found the sentence-writing task the most effective and the reading comprehension task the least effective, and Zou (2017), who showed writing exercises using target words much more effective than cloze-exercises.

From the above, it is apparent that there is no conclusive finding regarding which type of exercise is better than the others. Students' perceptions of their preferred vocabulary exercises and actual effectiveness might be influenced by their language proficiency, ease of task levels, and their ability to complete the tasks (Laufer, 2019).

Repeated oral reading exercises

Compared to written exercises, the quantity of oral exercises on vocabulary learning is very small. Alali and Schmitt (2012) conducted a study of 35 Kuwaiti female junior high school students' acquisition of single words and idioms to investigate whether doing oral practice or written practice makes a difference in acquiring vocabulary knowledge. Oral review consisted of asking students to read aloud target words and formulaic sequences 10 times in unison in the class for 10 minutes whereas the written review asked students to work in groups for 10 minutes on a written recall task. Thirty idioms, each containing an unknown target single word, were selected from a dictionary. The study showed that the two teaching methods produced similar patterns of learning for single words and idioms; however, written repetition was consistently shown to be more effective than oral review for form and meaning recognition and recall.

Another study was done by Durrant and Schmitt (2010), who examined the effects of learning 20 pairs of words with adjectives and nouns under three different conditions: one exposure in a sentential context only, oral reading two times but increasing speed at the second time, and exposures to different contexts. The two experimental repetition conditions were found to have effects superior to the no repetition condition. The oral repetition of a single sentential context also revealed better learning outcome than exposure to varied contexts. The authors speculated that the better effect for fluency-oriented repetition might be due to the cognitive ease of reading aloud an identical sentence the second time and hence improved collocational memory trace. Another possible reason explained by the authors was that the timed nature of the fluency-based verbatim condition enhanced the participants' attention to the language.

Vocabulary dimensions

The studies by Alali and Schmitt and Durant and Schmitt show that repetition may play some role on vocabulary learning but varying types of repetition also make some differences in learning outcome, and different aspects of vocabulary knowledge may also require different numbers of repetition. Nation and Webb (2011) hence emphasized that the importance of measuring different aspects of vocabulary knowledge as word knowledge is not an all-or-nothing construct. Some learning tasks may contribute more to gains in receptive knowledge while others to gains in productive knowledge. An oft-cited work by Nation (2001) shows that knowing a word involves 18 dimensions, and "using multiple tests to measure receptive and productive knowledge of different aspects may provide a much more accurate evaluation of the relative efficacy of tasks (Webb, 2005, p. 50). The decision on which aspects of knowledge should be measured depends on the extent to which the learning conditions or tasks may contribute to the aspects of vocabulary knowledge. In general, requiring students to produce answers is always more difficult than having them select one from a few options. Table 1 shows that nearly all form and meaning recognition tests are scored higher than form and meaning recall tests; passive recall is also easier than active recall.

Types of vocabulary

Now let us turn to the types of vocabulary. Two types of vocabulary items were examined in this study: multi-word items (MWI) and single-word items (SWI). MWI is used as an umbrella term in this study for strings of language containing more than a single word (Siyanova-Chanturia & Omidian, 2019), as can be seen as idioms, collocations, lexical bundles, phrasal expressions, lexical chunks, formulaic sequences, etc. in the literature. More distinctions among these terms can be seen in Wolter (2019). Many studies have shown that L2 learners' multi-word knowledge is very limited even after many years' learning (e.g., Nguyen & Webb, 2017). Recent research by Park and Chon (2019) revealed that their students' ability to comprehend individual words was statistically significantly higher than for idiomatic expressions. Similarly, Kim (2016) also reported that many L2 learners do not even recognize whether a given group of words is a single expression or individual words when encountering strings of known words. From the two studies above, direct teaching may be necessary to help L2 students to distinguish MWIs from SWIs, and direct teaching can be more effective in helping L2 learners to improve their MWI knowledge.

Studies on the acquisition of SWIs are abundant (see Table 1 for some examples), but research on comparing the extent of learning the two types of vocabulary knowledge is limited. Laufer and Girsai (2008) investigated the effects of three instruction conditions: meaning-focused instruction (MFI), non-contrastive form-focused instruction (FFI), and contrastive analysis and translation (CAT). They found students performed consistently higher on the MWIs than on the SWIs. Alali and Schmitt (2012), however, revealed that the learning of MWIs was slightly lower than that of SWIs. Similarly, Peters (2014) reported that collocations were more difficult to learn than SWIs, especially in the recall of L2 form. From Table 1, we can see that most studies focused on one type, either SWI or MWI; more studies are needed to compare how SWI and MWI are better acquired.

Table 1. Some Relevant Studies on Vocabulary Exercises, Vocabulary Dimensions and Vocabulary Item Types

Studies (year)	**Vocabulary dimensions**	**Task effectiveness**	**Vocabulary types**
Alali & Schmitt (2012)	form and meaning recognition > meaning recall > form recall	Written exercises > oral exercises	SWI ≈ MWI
Webb, Newton, & Chang (2013)	Receptive meaning and form > productive meaning > productive form	Reading while listening to short stories with different frequency of encounter	MWI only (consistently throughout regardless repetitions
Durant & Schmitt (2010)		Verbatim repetition > varied repetition > single exposure	MWI only
Folse (2006)	Passive recall (L1 translation or L2 synonym) + active usage (student-generated sentence) of a word.	Three completions > original sentences > one completion	SWI
Keating (2008)	Passive recall > active recall	Mixed results In Active recall (reading + fill in > sentence writing > reading + glosses) In Passive recall (Sentence writing > reading + fill in > reading + glosses	SWI only
Laufer & Hulstijn (2001)	Passive recall (L2 to L1)	Writing > reading + fill-in > reading	SWI
Laufer (2003)	Passive recall (L2 to L1)	1.Sentence writing > reading 2. Composition > reading 3. Fill in sentence > sentence writing > reading	SWI
Laufer & Girsai, (2008)	Passive recall (L2 to L1) > active recall (L1 to L2) (across all tasks)	contrastive analysis and translation (CAT) > non-contrastive form-focused instruction (FFI) > meaning focused instruction (MFI)	MWIs > SWIs (consistent in both passive recall and active recall)
Laufer & Rozovski-Roitblatv (2011)	passive recognition > passive recall (regardless number of encounter)	reading + word-focused exercises > Reading + dictionary check	SWI
Laufer & Rozovski-Roitblatv (2015)	Passive recognition > active recognition > passive recall > active recall	In passive recognition/recall 1 +Fs > F > R	SWI

		In active recognition/ recall 1 +Fs >R > F	
Lu (2013)	Passive recall (L2 to L1) > active recall (L1 to L2) > use	Somewhat mixed 3 BF > SW, or BFS, or BF	SWI
Sonbul and Schmitt (2010)	Meaning recognition > meaning recall > form recall	Read + Instruction > read only	SWI only (consistent in both reading only and reading plus conditions)
Webb (2005)	Receptive > productive (combined scores) very consistent across five dimensions	Mixed results Reading > writing (no time controlled) Writing > reading (time controlled)	SWI
Zou (2017)	Passive recall (L1 translation or L2 synonym) + active usage (student-generated sentence) of a word.	Composition-writing > sentence-writing > cloze-exercises	SWI
Puimège & Peters (2019)	Form recall > meaning recall > form recognition (comparing absolute gains, because form recognition score was very high in the pretest, so the gain was small)	L2 TV viewing	MWI > SWI
Puimège & Peters (2020)	Form recall > meaning recall	L2 TV viewing	MWI only
Peters, 2014	Two form recall tests	eight written, partly decontextualized vocabulary exercises	SWI > MWIs
Teng, F. (2019)	Receptive meaning > receptive form > productive meaning > productive form	full captioning > keyword captioning > no captioning	MWI only (consistent)

The present study

Two teaching experiments (Times) were conducted to investigate two forms of vocabulary practice on L2 learners' retention of SWIs and MWIs in four dimensions. Four theme-based texts were selected for the study. In Time 1, students were taught two theme-based short stories (numbers and animals), followed by doing two types of vocabulary practice. The same procedure was repeated at Time 2 but with different texts with themes on house and transportation. Four research questions were addressed below:

1. What main effects did different times, vocabulary exercises, dimensions, and types have on EFL learners' vocabulary retention?
2. What effect did different types of vocabulary exercises have on retention of vocabulary in different times?
3. To what extent were different dimensions of L2 vocabulary (aural form, aural meaning, written meaning (L2 to L1), use) affected by different forms of vocabulary exercises (repeated oral reading vs written exercises)?
4. How did different forms of vocabulary exercises affect the retention of SWI and MWI?

METHODOLOGY

Participants

A cohort of 97 Chinese EFL college students were recruited to take part in the present study; however, data from 43 participants were removed because they did not fully complete the intervention. The participants were enrolled in their required English reading course. Their overall language proficiency level, informed by their academic advisor was very low, mostly at A2 level (TOEIC scores range from 225-545) according to the European Framework of Reference, so the present study adopted direct teaching followed by two forms of vocabulary exercises after the reading instruction to enhance their retention of SWI and MWI. After the formal reading instruction, students were randomly divided into two subgroups: repeated oral reading and vocabulary written exercises.

Research design and procedure

Two teaching experiments (Time 1 and Time 2) were conducted to examine the effects of oral versus written vocabulary exercises on L2 learners' vocabulary retention of SWI and MWI. Because the participants had little knowledge about MWI, the researcher selected four theme-based stories as study materials. Formal reading instruction was first given followed by the vocabulary exercises. The students were divided into two subgroups: repeated oral reading (6 times) versus written exercises (three forms: gap-filling, meaning-matching, and rearranging scrambled sentences). The repeated oral reading subgroup was given the list of the target items embedded in sentences and orally practiced the target items five times on their own, and then read aloud to their teaching assistants the sixth time. The written exercises subgroup was given three types written exercises: gap-fillings, meaning matching, and rearranging scrambled sentences. At Time 1, students were tested on text 1 and text 2, at Time 2, text 3 and text 4. The research procedures were the same at both times (see below). After the delayed post-test, all

students were asked to fill out a short questionnaire (8 items) regarding their perceptions of the two types of vocabulary exercises, followed by an informal class discussion about the intervention. The results of the questionnaire and discussion were used only to explain the test outcome (see Appendix A for the questionnaire results). The research procedure is summarized below.

Week	Procedure
Time 1	
1	• Pre-screening unknown words through pre-tests
2-5	• Explicit reading instruction on the themes of numbers and animals
6	• Subgroup A doing repeated oral reading, subgroup B doing three kinds of vocabulary written exercises
7	• Administering one-week delayed post-tests
Time 2	• Repeating the steps of Time 1 but using different texts on the themes of house and transportation. • Subgroup B doing repeated oral reading, and subgroup A doing three kinds of vocabulary written exercises • Filling out a short questionnaire and having an informal class discussion about the treatment

Study materials

Four short-story texts were selected from *Idiom Magic* by John Ryan (1994). In this study, four themes on number, animals, house, and transportation were selected. Idioms grouped into thematic categories are easier to learn and memorize than unrelated ones (Boers, 2000; Cooper, 1998), so the target items in each story were all about the same theme. The four texts were analyzed by BNC/COCA, and the vocabulary profile is presented in Table 2. Each text contains 244, 238, 204, and 220 word types respectively. Texts 1 and 2 seem to have slightly lower percentages of words in the first two 1,000-word levels than texts 3 and 4 and the same can be said of the off-list words.

Table 2. Lexical Profile of the Four Theme-texts

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
Text 1 (numbers): Trouble on Cloud Nine			
one	461/85.06	206/84.43	173
two	18/ 3.32	16/ 6.56	16
three	4/0.74	4/1.64	4
Off-the-list	51/ 9.41	11/ 4.51	
total	542	244	200
Text 2 (animals): Helmer and Fanny O'Grady			
one	382/82.15	194/81.51	158
two	28/ 6.02	24/10.08	22
three	2/ 0.43	2/ 0.84	2
Off-the-list	43/ 9.25	9/ 3.78	

total	465	238	190
Text 3 (house): An Off-the-Wall Romance			
one	354/90.31	178/87.25	145
two	18/ 4.59	14/ 6.86	13
three	3/ 0.77	3/ 1.47	3
Off-the-list	11/ 2.81	3/ 1.47	
total	392	204	167
Text 4 (transportation): A Moving Story			
one	385/88.51	195/88.64	166
two	21/ 4.83	13/ 5.91	13
three	1/ 0.23	1/ 0.45	1
Off-the-list	26/ 5.98	9/ 4.09	
total	435	220	182

Target single words and multi-word items

The target unfamiliar items were first pre-screened by the instructor, who selected the possibly unfamiliar items, and then were further confirmed through the pre-tests. Only unfamiliar items were selected. A total of 60 items from the four texts were selected for tests. Each test contained 30 items: 11 single words and 19 multi-word items. Each item appears in the text only once.

Dependent measures

This study involved only a one-week delayed post-test to avoid practice effect; as well, delayed post-tests give a better indication of learning over time (Schmitt, 2010). Four vocabulary dimensions were measured: aural form recall, aural meaning recall, L2 meaning recall, and L2 form recall. Each dimension involves two measures (SWI and MWI), making a total of eight dependent measures. Each measure is described below.

Recall of aural form and aural meaning

Single words: Students heard a target word in a sentence and the target word was repeated once. For example: *Don't give me the runaround ; just tell me if you will come along*. The students had to spell r-u-n-a-r-o-u-n-d and produce the Chinese meaning of *runaround*.

Multi-word items: Students heard a target MWI in a sentence and the target item was repeated once. For example: *I haven't lost my job yet, but the handwriting is on the wall*. Students had to spell the repeated idiom and translate the target item into Chinese.

L2 written form recall

In this dimension, students had to produce correct SWIs and MWIs. This dimension is usually more difficult than productive translation from L2 to L1 (see below).

Single words: Students saw a sentence with a blank space, followed by its Chinese meaning in parenthesis. The first and last letters were provided as prompts. For example, Don't give me the r___d (藉口); just tell me if you will come along. Students had to translate 藉口 into English according to the prompts.

Multi-word items: As with single word items, students saw a sentence with a blank space in it, followed by its Chinese meaning in parenthesis. One word in each MWI was provided as a prompt. For example, *I haven't lost my job yet, but* _____(預兆很明顯，wall). Students then had to produce the correct multi-word item using the prompt (wall).

L2 written meaning recall

In this dimension, students had to produce the Chinese meaning for the target items. This dimension is the easiest one among the four.

Single words: Students had to translate the target underlined words into its equivalent Chinese meaning. For example, *Don't give me the runaround, just tell me if you will come along*. Students had to produce the Chinese meaning for the word underlined.

Multi-word items: Students had to translate the target MWI into Chinese. For example: *I haven't lost my job yet, but the handwriting is on the wall*. Students had to translate "the handwriting is on the wall" into its equivalent Chinese.

Scoring and data analysis

The marking was first done by one English teacher and one teaching assistant, and then further checked by the researcher. If there was any inconsistency in the scores given by the raters, the researcher served as arbiter to decide the scores. Spelling mistakes were not awarded points. For example, if "smell a rat" was spelled as "smell a ret," no point was given. The meanings for SWI and MWI had to correspond to the context of the texts. SPSS version 25 for Windows was used to analyze the data. The data were checked and entered in the long format, and Hierarchical linear modeling (HLM) was performed. The dependent variable was the score for each dimension of the target MWIs and SWIs. Because HLM takes only whole numbers, each score was binary (0 for an incorrect answer and 1 for a correct answer). The random factor was the student participants, and the fixed factors included the following: Times (1 vs 2), vocabulary exercises (oral vs written); vocabulary dimensions (aural form recall, aural meaning recall, L2 form recall, and L2 meaning recall), types of vocabulary items (SWI vs MWI). The results of the short questionnaire and class discussion were used to explain the results and were not particularly analyzed apart from some frequency count. The first research question looked at the main effects of the fixed variables, and the second, third, and fourth questions then looked at the interaction effects between vocabulary exercises and other three fixed variables.

RESULTS

The descriptive statistics are presented in Table 3 and featured in Figure 1. As shown, students performed better at Time 1 than at Time 2; the written exercises group scored higher than the oral exercise group; students scored the highest for doing L2 meaning production (EC), followed by aural form recall, aural meaning recall, and L2 form recall (CE). Students also performed better on the MWI than single words.

Table 3. Descriptive Statistics of Vocabulary Retention (times, different vocabulary exercises, dimensions, and types)

		Mean	N	SD
Times	1	.60	6480	.49
	2	.54	6480	.50
Vocabulary exercises	Oral	.54	6720	.50
	Written	.59	6240	.49
Vocabulary dimensions	AF (aural form recall)	.59	3240	.49
	AM (aural meaning recall)	.51	3240	.50
	CE (L2 form recall)	.50	3240	.50
	EC (L2 meaning recall)	.66	3240	.47
Vocabulary types	SWI (single word item)	.53	4752	.50
	MWI (multi-word item)	.59	8208	.49

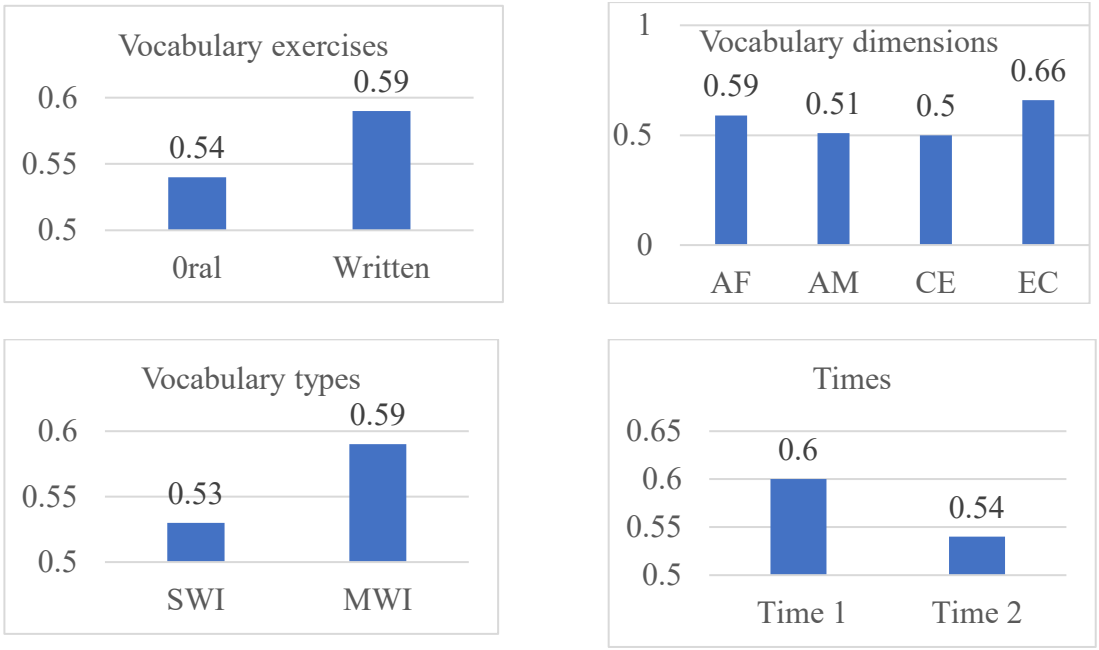

Figure 1. Students' Performance on the Four Fixed Variables in the One-week Delayed Post-test

Note: AF: aural form recall; AM: aural meaning recall; CE: L2 form recall; EC: L2 meaning recall

The main effects of times, vocabulary exercises, vocabulary dimensions and types

To answer to the first research question for the main effects of the fixed variables on L2 vocabulary retention, hierarchical linear modeling (HLM) was used to analyze the data. The results of HLM are presented in Table 4. As shown, there were main effects for all the fixed variables. For the teaching results at two times, students retained more L2 vocabulary knowledge at Time 1 than Time 2, B = -0.31, p = .007 and OR = 0.73. Students doing written exercises

performed significantly better than those who did oral reading, $B = 0.24$, $p = .042$, and $OR = 1.27$. For students' performance in the four dimensions, L2 meaning recall (EC) was significantly higher than AF (aural form), L2 form recall (CE), and AM (aural meaning), with ORs 1.48, 2.33, and 2.19 respectively. On the contrary, L2 form recall was the most difficult dimension among the four. As shown, the score of CE was statistically significantly lower than that for AF (OR= 0.63, 95% [0.56-0.71], $p < .001$), and EC (OR= 2.30, 95% [1.92-2.76], $p < .001$. The score of AM, however, was found to be significantly lower than that of AF, as $B = -0.39$, $p < .001$, and $OR = 0.67$. No significant difference was found between CE and AM, $B = -0.06$, $OR = 94$, $p = .414$. For vocabulary types, students performed significantly better on the MWI than SWI ($OR = 1.34$, 95% [1.09-1.64], $p = .006$).

Table 4. Summary of the Results for the Main Effects of HLM

Fixed effects	***B***	***OR* (95%CI)**	***t***	***p***
Intercept	0.24	-	1.46	.145
Times: 2 vs 1	-0.31	0.73 (0.58 - 1)	-2.70**	.007
Exercises: written vs oral	0.24	1.27 (1.01 - 2)	2.05*	.041
Dimensions				
EC vs AF	0.38	1.47 (1.19 - 2)	3.62***	<.001
CE vs AF	-0.45	0.64 (0.57 - 1)	-7.51***	<.001
AM vs AM	-0.39	0.68 (0.56 - 1)	-4.16***	<.001
EC vs CE[a]	0.84	2.30 (1.92 - 2.76)	9.05	<.001
EC vs AM[a]	0.77	2.17 (1.91 - 2.46)	12.04	<.001
CE vs AM[a]	-0.06	0.94 (0.81 - 1.09)	-0.82	.414
Vocabulary types: MWI vs SWI	0.31	1.37 (1.12 - 2)	3.02**	.003

*$p < .05$, **$p < .01$, ***$p < .001$

Note. [a] Reference group; AF: aural form; AM: aural meaning; EC (English to Chinese): L2 meaning recall; CE (Chinese to English): L2 form recall; MWI: multi-word item; SWI: single-word item

The interaction effects between vocabulary exercises and times, vocabulary dimensions, and vocabulary types

Although main effects were found statistically significant for all fixed variables, another main focus of this study was to investigate the interaction effects between vocabulary exercises and times, vocabulary dimensions, and vocabulary types. The results of this analysis provide answers to research questions 2, 3, and 4. The analysis results of HLM are summarized and presented in Table 5. In answer to research question 2, the written group retained significantly more L2 vocabulary knowledge than the oral group, $B = -0.88$, $p = .005$, and OR = 0.41. The simple effect shown in Table 6 revealed that the difference was significant only at Time 1, $B = 0.71$, $p = .001$, and OR = 2.03 (also see Figure 2). The answer to research question 3 was that a significant interaction effect between vocabulary exercises and vocabulary dimensions was found only for EC vs AF, $B = 0.26$ ，$p = .039$. The simple effects (see Table 6) of EC showed that the written group scored significantly higher than the oral group; however, the vocabulary

exercises show no significant effect for the dimensions AF, $B = 0.13$, $p = .431$, and OR = 1.13 No significant interaction effects for the rest of the comparisons (see Figure 3). To answer research question 4, no interaction effect was found for vocabulary exercises and vocabulary types, $B = -0.19$ ，$p = .367$. Students consistently scored higher for MWI in vocabulary exercises.

Table 5. Summary of the Analysis of HLM for the Interaction Effects

Fixed effects	***B***	***OR* (95%CI)**	***t***	***p***
Intercept	0.04	1.04 (0.70 - 2)	0.19	.849
Times: II vs I	0.12	1.12 (0.86 - 1)	0.87	.385
Exercises: written vs oral	0.66	1.94 (1.07 - 4)	2.18*	.029
Dimensions				
EC vs AF	0.26	1.30 (1.04 - 2)	2.28*	.022
CE vs AF	-0.48	0.62 (0.53 - 1)	-6.26***	.000
AM vs AM	-0.50	0.60 (0.48 - 1)	-4.24***	.000
EC vs CE[a]	0.74	2.10 (1.75 - 3)	8.06***	.000
EC vs AM[a]	0.77	2.15 (1.87 - 2)	10.84 ***	.000
CE vs AM[a]	0.03	1.03 (0.86 - 1)	0.29	.773
Vocabulary types: MWI vs Single	0.40	1.50 (1.11 - 2)	2.66**	.008
Exercises * Times: 2 vs 1	-0.88	0.41 (0.22 - 1)	-2.78**	.005
Exercises *Dimensions				
EC vs AF	0.26	1.30 (1.01 - 2)	2.06*	.039
CE vs AF	0.06	1.06 (0.87 - 1)	0.57	.569
AM vs AF	0.24	1.27 (1.00 - 2)	1.93	.054
EC vs CE[a]	0.21	1.23 (0.95 - 2)	1.58	.114
EC vs AM[a]	0.02	1.02 (0.85 - 1)	0.23	.819
CE vs AM[a]	-0.18	0.83 (0.65 - 1)	-1.42	.154
Exercises* vocabulary types:				
MWI vs single	-0.19	0.83 (0.54 - 1)	-0.90	.367

*$p < .05$, **$p < .01$, ***$p < .001$

Note: [a] Reference group; AF: aural form; AM: aural meaning; EC (English to Chinese): L2 meaning recall; CE (Chinese to English): L2 form recall; MWI: multi-word item; SWI: single-word item

Table 6. The Simple Effects of Fixed Variables

Variables	***B***	***OR* (95%CI)**	***t***	***p***
Times				
2	-0.17	0.84 (0.56 - 1.25)	-0.86	.392
1	0.71	2.03 (1.35 - 3.04)	3.43	.001
Dimensions				
EC	0.39	1.48 (1.15 - 1.89)	3.08**	.002
CE	0.18	1.20 (0.88 - 1.64)	1.15	.249
AM	0.37	1.45 (1.09 - 1.92)	2.56*	.011
AF	0.13	1.13 (0.83 - 1.55)	0.79	.431

Variables	*B*	*OR* (95%CI)	*t*	*p*
Vocabulary types				
MWI	0.17	1.19 (0.95 - 1.48)	1.50	.135
Single words	0.36	1.44 (0.96 - 2.16)	1.75	.080

*$p < .05$, **$p < .01$, ***$p < .001$

Note: [a] pairwise comparisons

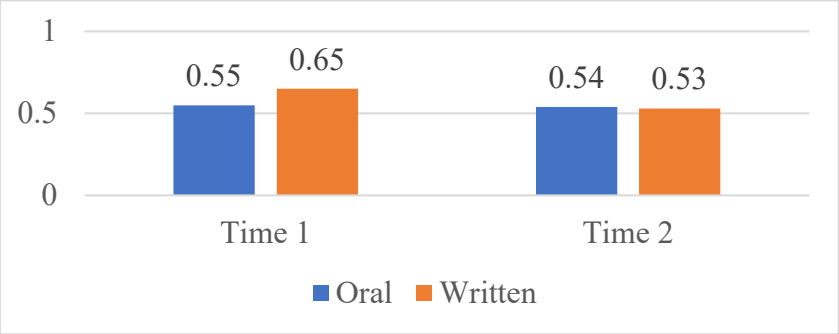

Figure 2. The Effects of Different Types of Vocabulary Exercises on Two Different Times

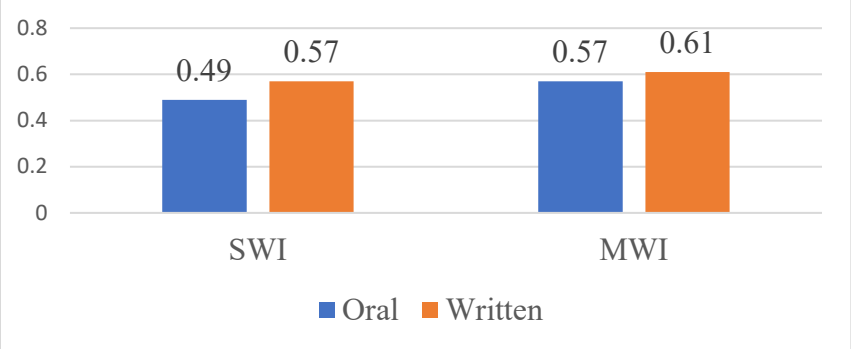

Figure 3. The Effects of Different Types of Vocabulary Exercises on Different Vocabulary Types

Note: SWI: Single word items; MWI: Multi-word items

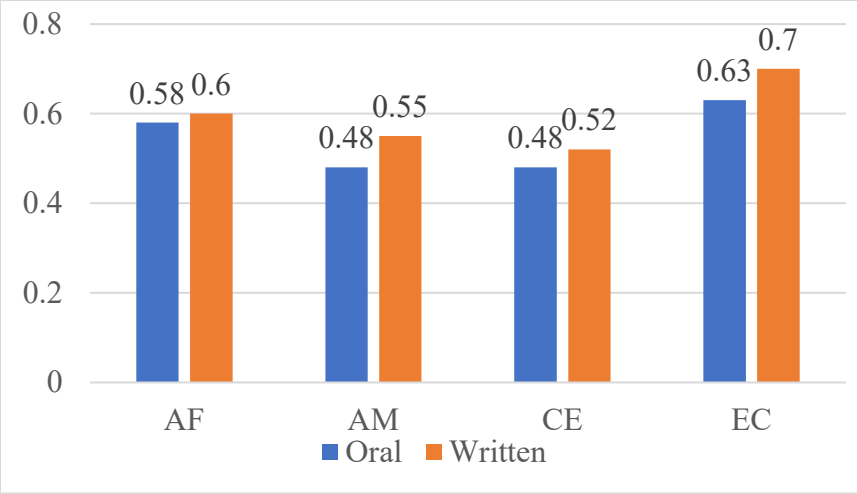

Figure 4. The Effects of Different Types of Vocabulary Exercises on Different Vocabulary Dimensions

Note: AF: aural form recall; AM: aural meaning recall; CE: L2 form recall; EC: L2 meaning recall

DISCUSSION

From the results shown above, we may see that all fixed variables had significant effects on L2 students' vocabulary retention. The significant interaction effects between vocabulary exercises and other fixed variables were found only at the first teaching experiment (Time 1) and L2 meaning recall. In the following, the main effects, together the significant interaction effects will be discussed.

The effects of vocabulary exercises at Times 1 and 2

The overall results revealed students practiced the newly learned items through written exercises led to a higher level of retention rate than oral exercises did. Three possibilities may explain the results. The first main reason could be that the students doing oral exercises might have paid more attention to the accurate linkage between form and pronunciation than the correct meanings of the target items. Secondly, students' final oral reading rates were timed by the teaching assistants, so students might have been pushed to focus more on the speed of the oral production. In order to increase the speed, the oral group students might have simply barked out the words instead of trying to understand the meanings of the newly learned items, as indicated by the oral group spending less time completing the task. Thirdly, the nature of the written exercises seemed to be more related to the delayed post-test because the three exercises tended to have drawn students' attention to focusing on the meaning-form linkage, and the rearranging of scrambled sentences in particular raised students' awareness of the structures of MWIs. The results were somewhat different from those found by Durrant and Schmitt (2010), who found that students did oral reading performed slightly better than those who were exposed to the target items in varied contexts. The target items of the present study and those of Durrant and Schmitt were very different. In their study, the target MWIs were a combination of an adjective and a noun, which seemed to be more suitable for using oral exercises (in my own learning and

teaching experience); however, the target items of the present study are mixed with all kinds of MWIs, such as noun phrases (e.g., *skeleton in the closet*) or verb phrases (e.g., *lay the blame at someone's door*) or preposition phrases (e.g., *on cloud nine*). Peters (2016) found that the Dutch students recalled and recognized better on adjective–noun collocations than on verb–noun ones. From these studies, the effects of vocabulary practice may be influenced by different construction of MWIs. In addition, many other factors (e.g., word length, congruency, participants' vocabulary knowledge) all come to play and affect the outcome. These factors, however, are not the focus of the present study

Let us turn to examine the interaction effects between the written exercises and two teaching times. The results for the effects of written and oral exercises were not consistent: written exercises produced significantly higher outcome than the oral exercises at Time 1 only, and no significant difference was found at Time 2. The inconsistent results might have to do with the novelty of oral exercises at Time 1. As mentioned earlier, some students simply barked out the sentences without putting effort to remembering the meaning of the sentences. Even though the two groups were allowed the same amount of time to complete the tasks, the oral exercises group finished earlier than the written group at Time 1. At Time 2, both groups finished at aproximately the same time. Another likely reason might be simply due to there being different target items at the two times. All target MWIs at Time 1 comprised known words only, but at Time 2, there were three MWIs containing unfamiliar single words, as the overall retention rate for Time 1 was 60% but only 54% at Time 2. The increase in the difficulty of the target items for the students in both groups might account for there being no significant difference between the two groups.

Although on the whole students scored higher on written exercises, some 38.9% of the students considered oral reading more effective than written exercises, another 29.7% held the opposite opinion, and 31.5% straddled the middle. More interesting still, 59.2% of the students preferred doing oral exercises whereas only 16.7% favored written exercises, which might be due to the fact that oral exercises were new to them, and they did not need to understand the meanings of the target items, thereby making it easier for them to compete the tasks. Given the large proportion of students preferring oral exercise, repeated oral reading can be included as one of the vocabulary exercises, but students must be reminded of the importance of understanding form and meaning connection.

The effect of vocabulary exercises on vocabulary dimensions

The main effects of vocabulary dimensions showed that students performed differently on different vocabulary dimensions. The score for EC (L2 meaning recall) was the highest (66%), followed by AF (59%), AM (51%), and CE (L2 form recall) (50%). With an exception of the difference between CE and AM, all other dimensions showed a significant difference between them. It is apparent that L2 form recall, regardless of SWIs or MWIs, was the most challenging task for L2 learners; the results corroborated many previous studies which found that EFL learners seemed to have more difficulties recalling the L2 forms than recalling meanings (see Table 1).

Taken into account the effects of vocabulary exercises on different vocabulary dimensions, students doing written exercises scored significantly higher than those doing oral exercises only on L2 meaning recall. The results were not surprising because the nature of delayed post-test was more relevant to the exercises that the written group practiced; specifically,

the written group was required to understand the meanings of the target items to complete the tasks. The result was also supported by their response in the questionnaire, in which students were asked to rate whether oral or written exercises were more helpful for remembering the written meaning under four options (oral, written, no difference, not sure), and more than a half (55.6%) of the students rated written exercises more effective; however, 66.7% of the students responded that oral exercises were more helpful for remembering aural meaning. Their responses in effect did not reflect their actual performance because the oral group did not perform as expected—better on aural form recall and aural meaning recall than the written group; however, three dimensions of the results were not significantly different. Overall, different vocabulary exercises did not have a strong effect on different vocabulary dimensions. The results seemed to be consistent with those studies on incidental vocabulary learning, in which L2 students consistently performed better on L2 meaning recall (see Table 1). That seems to be the first step for establishing the connection between form-meaning.

Another explanation for the absence of salient differences among other dimensions could be that all tests required production, which might have made the tasks more challenging. If only the L2 meaning recall of the present study was compared with that of Alali and Schmitt's (2010), the written group did better than the oral group in both studies. In terms of retention rates, the present study was marginally higher than those of Alali and Schmitt because their delayed post-test was administered 12 days after the intervention but the post-test for the present study was only 7 days after the intervention. If students could recall approximately 70% of L2 meanings of MWIs, it is likely that they could recognize much a higher percentage of form-meaning knowledge of MWIs.

The effect of vocabulary exercises on vocabulary types

The results revealed that students performed significantly better on the MWIs than SWIs. No significant interaction effect was found between vocabulary exercises and the vocabulary types. Although students had zero knowledge for both SWIs and MWIs, three salient features may explain the different gains. The first feature was the number of known words. All SWIs were unfamiliar to students in all dimensions tested. When they did not know the word, it was difficult for them to guess its meaning, but the MWIs, however, were comprised mostly of known individual words (except three items), which might be have made it easier for them to guess meanings from the other words in the MWI. The second feature was relevance. All SWIs were unrelated to each other, e.g., *lean*, *poke*, and *motion*, but the MWIs were grouped by themes, e.g., *raise the roof*, *open house*, and *hand-writing is on the wall*, all being related to house. The third feature was interest. In the classroom discussion, a large proportion of the students showed a greater interest in learning MWIs than SWIs. Three reasons were offered by the students. Firstly, they could put many known words together and produce a different meaning instead of having to learn another new item. Secondly, some MWIs also have their origins, which interested students very much. For example, *the lion's share* came from Aesop's fable. Finally, it is easier for students to use strategies to remember MWIs. For example, students related *to pull a fast one on someone* to a magician who moves things fast to fool the viewers. The linking between L1 and L2 culture also makes MWIs easier to acquire. For example, *dress to the nines* and *on cloud nine*, number 9 in Chinese normally has positive meanings, implying something good or someone fortunate, and so its meaning can be more easily understood and remembered. The abovementioned features might have facilitated students' retention more for MWIs than for

single words and may have moderated the effects of vocabulary exercises.

The results for students retaining more MWIs than SWIs were consistent with Laufer and Girsai's incidental learning (2008), though their students' gains were very low due to a lack of focused instruction; however, mixed pictures were shown in Peters' studies (2014), which showed that MWIs were more difficult to learn than SWIs. The results of the current study were quite satisfactory because target items were contextualized in theme-based reading texts, followed by vocabulary focused exercises (also see Laufer & Rozovski-Roitblat, 2015).

PEDAGOGICAL IMPLICATIONS AND CONCLUSION

The present study has shown that L2 students could retain more than 50% of SWIs and MWIs through the two types of vocabulary exercises. The retention rates across different vocabulary dimensions can be considered satisfactory, and the two types of vocabulary were also suitable for the participants' level. The results may have some pedagogical implications for classroom vocabulary activities:

1. This study has shown that the written vocabulary exercises have better effects for retaining SWIs and MWIs than oral exercises do, but students' perceptions showed a preference for doing oral exercises over written exercises. This may suggest that the repeated oral reading can be used as a variety for vocabulary practice, which means vocabulary exercises can involve both written and oral exercises.
2. The study showed that low-proficiency students can quickly pick up MWIs that contain high-frequency known words through formal instruction plus doing vocabulary exercises, and retention rates from doing MWI exercises were higher than for SWI exercises. This finding suggests the value of learning the MWIs of the high-frequency level, and through the combinations of known words. As per Shin and Nation (2008), the most frequent 100 node words make up 35% of the total number of collocations of the first1,000 words.
3. Among the four dimensions measured, students' retention rate on the translation from L1 to L2 was the lowest. To enhance students' recall of L2 forms, other forms of vocabulary exercises may be needed. Perhaps writing repetitions can be used for lower-level students, and writing a composition using the target items will suit the more advanced learners.

Although the effects of both forms of vocabulary exercise are satisfactory, some limitations need to be mentioned when interpreting the results. To avoid practice effect, this study did not administer an immediate post-test. This has limited our understanding for the immediate effects of oral exercises. Repeated oral exercises might have a better effect immediate after the learning, so future studies are suggested to include immediate post-tests to confirm this. To boost learning effects, immediate post-tests might enhance retention rates, as Peters (2014) reported that the group given an immediate post-test retained more vocabulary knowledge than the group given a one-week delayed post-test; however, Peters added, "in authentic learning situations, learners are generally not tested immediately after the learning session, and so forgetting is more likely to occur" (p. 91). Another limitation was to prevent students from making wild guesses, this study did not measure students' recognition knowledge across dimensions. If students could recall approximately 70% of L2 meanings, it is likely that they could recognize a much higher percentage of form-meaning knowledge. The study involved only theme-based texts; it is unclear whether theme-based texts are more beneficial than non-theme-

based texts for learning SWIs and MWIs. Previous studies on this issue are rare and inconclusive—positive in Boers (2000) and Cooper (1998), but showing no difference in Zyzik (2011). More research into this area is needed. Another limitation is that the present study involved two separate subgroups doing either oral or written exercises; however, students seemed to prefer oral exercise to written exercises. Future study may include a group doing both written and oral exercises to examine whether mixing two types of vocabulary exercises would produce better effects.

REFERENCES

Alali, F.A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, *3*(2), 153-180.

Boers, F. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics*, 24, 553–71.

Chang, C.- S. (2019). The effects of repeated oral reading practice on EFL learners' oral reading fluency development. *Reading Matrix*, *19*(2), 103-113.

Cooper, T. (1998). Teaching idioms. *Foreign Language Annals*, 31, 255–66.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL quarterly*, *40*(2), 273-293.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, *12*(3), 365–386.

Kim, C. (2016). L2 learners' recognition of unfamiliar idioms composed of familiar words. *Language Awareness*, *25*(1-2), 89-109.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian modern language review*, *59*(4), 567-587.

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, *24*(1), 29-49.

Laufer, B. (2019). Evaluating exercises for learning vocabulary. In *The Routledge handbook of vocabulary studies* (pp. 351-368). Routledge.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied linguistics*, *29*(4), 694-716.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, 15(4), 391-411.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687-711.

Lu, M-H (2013). Effects of four vocabulary exercises on facilitating learning vocabulary meaning, form, and use. *TESOL Quarterly*, *47*(1), 167-176.

Nation, I. S. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.
Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge.
Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298-320.
Park, J., & Chon, Y. V. (2019). EFL Learners' Knowledge of High-frequency Words in the Comprehension of Idioms: A Boost or a Burden? *RELC Journal*, 50(2), 219-234.
Pellicer-Sánchez, A. (2019). Learning single words vs. multiword items. In *The Routledge handbook of vocabulary studies* (pp. 158-173). Routledge.
Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*(1), 75-94.
Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20(1), 113-138.
Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424-438.
Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525-549.
Ryan, J. (1994). *Idiom Magic*. Overseas Radio & Television Inc. Taipei: Taiwan, R.O.C.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
Shimono, T.R. (2019). The effects of repeated oral reading and timed reading on l2 oral reading fluency. *Reading Matrix*, *19*(1), 139-154.
Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, *62*(4), 339-348.
Siyanova-Chanturia, A., & Omidian, T. (2019). Key issues in researching multiword items. In *The Routledge handbook of vocabulary studies* (pp. 529-544). Routledge.
Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *ELT journal*, *64*(3), 253-260.
Teng, F. (2019). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education*, 142. DOI: 10.1016/j.compedu.2019.103655
Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559-599.
Wolter, B. (2019). Key Issues in Teaching Multiword Items. In *The Routledge Handbook of Vocabulary Studies* (pp. 493-510). Routledge.
Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in second language acquisition*, *27*(1), 33-52.
Webb, S., Newton, J., & Chang, C.-S. (2013). Incidental learning of collocation. *Language Learning*. *63*(1), 91-120.

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, *21*(1), 54-75.

Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing. *Language Teaching Research*, *15*(4), 413-433.

Dr. Anna C-S Chang *is Professor in the Department of Applied English at Hsing Wu University, New Taipei, Taiwan, teaching courses on English listening, reading and vocabulary. Her main research interests focus on listening and reading development, and vocabulary learning. She has published extensively with internationally refereed journals.*

Email: annachang@livemail.tw

Brandt T-Y Chen *is an English teacher at Giraffe English Language Institute in Taiwan.*

Email: can90634@gmail.com

Appendix A: Results of the Questionnaire

Responses for Question Items 1 to 4 (n =54)

	1 SD	**2 D**	**3 PD**	**4 N**	**5 PA**	**6 A**	**7 SA**
1. I think oral reading is more effective in retaining vocabulary knowledge than doing vocabulary exercises.							
frequency	1	4	11	17	13	6	2
%	1.9	7.4	20.4	31.5	24.1	11.1	3.7
2. I like doing vocabulary exercises better than oral reading							
frequency	4	10	18	13	6	2	1
%	7.4	18.5	33.3	24.1	11.1	3.7	1.9
3. Doing oral reading made me feel nervous							
frequency	10	4	14	14	4	5	3
%	18.5	7.4	25.9	25.9	7.4	9.3	5.6
4. Doing vocabulary written exercises is boring.							
frequency	4	10	16	13	3	5	3
%	7.4	18.5	29.6	24.1	5.6	9.3	5.6

Note. SD: strongly disagree; D: disagree, PD: partially disagree; N: neutral; PA: partially agree; A: agree; SD: strongly agree

Responses for Question Items 5 to 8 (n =54)

	Oral reading	**Written exercises**	**No difference**	**Not sure**
5. Which way is more helpful for remembering the word forms?				
frequency	14	22	9	9
%	25.9	40.7	16.7	16.7
6. Which way is more helpful for remembering the aural meaning?				
frequency	36	4	7	7
%	66.7	7.4	13.0	13.0
7. Which way is more helpful for remembering the written meaning?				
frequency	9	30	7	8
%	16.7	55.6	13.0	14.8
8. Which practice took more time?				
frequency	10	20	17	7
%	18.5	37.0	31.5	13.0