THE EFFECTIVENESS OF VOCABULARY LEARNING TASKS: AN EXAMINATION OF PREDICTIVE POWER OF TECHNIQUE FEATURE ANALYSIS

Linh Tran
University of Finance and Marketing; khanhlinhtran1104@gmail.com

ABSTRACT

This research aimed to validate the Technique Feature Analysis (Nation & Webb, 2011), a theoretical framework that includes criteria for vocabulary learning to occur and evaluates the effectiveness of a vocabulary learning task. The study investigated the correlation between the predictability of the framework and the task's effect on vocabulary learning and retention. To this end, EFL learners were assigned three vocabulary tasks of learning 15 target words. The tasks were designed based on the Technique Feature Analysis framework and had different presence of each vocabulary learning criteria. Word forms and words' meaning recognition knowledge of the target words were measured immediately and one week later. The results revealed that supporting evidence for Technique Feature Analysis was found in terms of the retention of meaning recognition knowledge whereas no significant difference was detected in the immediate form and meaning recognition knowledge and delayed form recognition knowledge.

Keywords: vocabulary learning task; L2 vocabulary learning; Technique Feature Analysis

1. INTRODUCTION

Vocabulary knowledge is an indispensable part of language learning and a prerequisite for language proficiency (Lin and Morrison, 2010). For learning to take place, inputs must be made comprehensible (Oller & Krashen, 1988) and contains quantities of vocabulary repetition (Vidal, 2011).

Given the limited time in a language classroom, especially in Vietnam where an English class is devoted to system and skill lessons for five to six hours a week, L2 learners do not have sufficient time for reading and are not exposed to the number of lexical repetitions needed for learning new words. In this case, a word remains unnoticed and if noticed, its meaning cannot be retained (Laufer, 2019). If vocabulary learning from input does not expand or deepen lexical knowledge, learners can benefit from word-focused exercises (Laufer, 2019). The questions are how to evaluate the effectiveness of vocabulary learning tasks and what makes some tasks more effective than others. Accordingly, three frameworks are proposed to predict the effectiveness of the tasks, namely the Involvement Load Hypothesis (hereafter ILH; Laufer & Hulstijn, 2001), Type of Processing –Resource Allocation (TOPRA) model (Barcroft, 2002) and Technique Feature Analysis (hereafter TFA; Nation & Webb, 2011). While there are several studies confirming the predictive power of the ILH and TOPRA model (Keating, 2008; Kida &

Barcroft, 2018) the TFA framework has not received much attention and the need of verifying its predictive power is what urged the need of this empirical research.

2. LITERATURE REVIEW

2.1 Definition of vocabulary exercise

Richards et al. (1985) defined a task as an activity or action that requires language knowledge to carry out. According to Laufer (2019, p.352), a vocabulary exercise is "a response to a language prompt, a response that requires the understanding of the words on which the exercise focuses, with or without producing them." The exercises can facilitate communicative and non-communicative purpose. For example, the former can be an act of reading and looking up unknown word in the dictionary to answer comprehension questions. Exercises such as gap-filling and sentence writing are non-communicative because words are practised in isolation and without context.

2.2 Factors affecting vocabulary learning

There are many factors affecting vocabulary acquisition: interlexical, intralexical and contextual factors. Interlexical factors concern the interaction between the new word and the words the learner knows in L1 and L2. Intralexical factors stem from the word itself which can be pronounceability, orthography, morphology and semantic features (abstractness, register, idiomaticity, polysemy) (Laufer, 1997). Contextual factors provide the context in which word knowledge can be learned from and the two key conditions that facilitate learning are repetition and attention (Nation & Webb, 2017). The Technique Feature Analysis deals with contextual conditions; therefore, this composition would provide literature review on factors facilitating vocabulary learning.

As can be seen from Figure 1, Nation and Webb (2017) developed a framework of vocabulary learning condition that involves repetition and quality of attention (e.g. noticing, retrieval, varied encounters or varied use and elaboration).

Figure 1

Framework of vocabulary learning condition

Repetition	= number of encounters (first encounter repetition)	
Quality of attention	Incidental	Deliberate
Noticing		
Retrieval (receptive or productive)		
Varied encounters (receptive) or varied use		
(productive)		
Elaboration (receptive or productive)		

2.2.1 Repetition

As asserted by Nation (2013), one of the vital facilitators of vocabulary learning, incidentally or deliberately, is repetition. Past research has shown that vocabulary gains increase among both L1 learners (Shu et al., 1995) and L2 learners (Webb, 2007) as a result of repeated lexical encounters. Linguistic researchers then began the quest for the extent of vocabulary uptake resulting from repetition. The findings can be categorized into the frequency and spacing of repetition.

2.2.2 Frequency of repetition

Despite a large number of studies on the effect of repetition on the vocabulary learning process, the frequency of encounters for it to be effective remains inconclusive as studies on incidental and deliberate learning come up with different values. Saragi et al. (1978) concluded that for incidental learning, a word needs to be repeated at least ten times so that participants can recognise its meaning. Rott (1999) investigated the effect of repetition (0, 2, 4, 6 times) in reading text on both L1 (English) and L2 (German) vocabulary learning. The results show learning occurs at the second exposure and more encounters correlates to higher learning gains. Recent findings from Nation (2014) and Laufer (2017) suggested 12 encounters as a "moderately safe goal" (Nation, 2014, as cited in Malone, 2018).

2.2.3 Spacing of repetition

As important as the number of repetitions, researchers also concern how these repetitions are distributed over time. Drawing on the definition of massed and distributed practice in learning by Stevick (1996), Goossens et al. (2012) state that repetitions can be spaced (i.e. learners meet a word several times with intervals between each encounter) and massed (i.e. a word is encountered repeatedly in a period of time). Research in vocabulary acquisition has shown that spaced repetition is superior over massed repetition in terms of retention (Kornell, 2009; Nakata, 2015; Nakata & Webb, 2016). Baddeley (1990) found that the form and meaning links can be strengthened every time a word is retrieved. Given that spacing benefits vocabulary gains, the question is that how the length of intervals within spaced repetition can be distributed to maximise learning. Many studies have investigated two types of relative spacing - equal and expanding spacing. In equal spacing, each interval is spaced equally (e.g. 2-2-2 minutes) while expanding spacing refers to the idea that the spacing of interval is gradually increased (e.g. 1-3-5 minutes). Expanding spacing is often believed to be the most effective in promoting long-term retention (Baddeley, 1997; Ellis, 1995, as cited in Roediger & Karpicke, 2010). However, studies comparing the vocabulary gains between expanding and equal spacing show no difference (Pyc & Rawson, 2007). Nakata (2015) points out that "feedback, task difficulty, absolute spacing, and the retention interval (i.e., the interval between the treatment and post-test)" are conditions for relative spacing to occur and yet are not considered in the two studies above. Nakata constructed four conditions in his research and found that expanding spacing has slightly greater positive effect on vocabulary learning. Therefore, it can be said that there is no distinctive discrepancy between the two types of relative spacing as regards to vocabulary learning.

2.2.4 Quality of attention

As mentioned above, learning depends on how often learners encounter a word. According to Nation (2015), vocabulary gains would be significantly enhanced if a word receives a deeper level of attention in every meeting. Krashen (1985) suggested that comprehensible input plays a vital role in language acquisition. However, given that not all of the input can be utilized as intake, Schmidt (1990) claimed that it is impossible for learning to occur without attention to input and there are some studies share the similar view (Robinson, 1995; Schmidt, 1990). Comparing the effect of repetition and quality of attention on different types of word knowledge, Laufer and Rozovski-Roitblat (2014) and Webb (2008) discovered that the later has the stronger effect.

2.2.5 Noticing

Noticing occurs when learners draw their attention to the target words, which according to Rutherford and Sharwood Smith (1985), facilitates language learning. During this process, the linguistic feature will be stored in short-term memory and later in long-term memory (Robinson, 1995). Schmidt (1990, 1994) proposed the noticing hypothesis which states that word forms and its meanings in the input must be consciously noticed by learners for the conversion of input to intake. The richer the input, the more likely it is for intake to occur. In fact, learners are more likely to learn what they pay attention to – that is if word form is attentive to, this aspect is more likely to be gained than its meaning (Barcroft, 2009).

2.2.6 Retrieval

Retrieval is the second condition contributing to the quality of attention and it is only possible after the initial encounter because a word cannot be recalled if it is not stored in memory previously. Retrieval can be productive and receptive. When learners see or hear a word form and have to retrieve its meaning, as when reading or listening, it is known as receptive retrieval. On the other hand, productive retrieval entails expressing word meaning and retrieving its form in speaking and writing (Nation, 2013). Repetition is known to be facilitative to vocabulary learning. An extensive investigation by Baddeley (1990) suggests it is more important to have repetition opportunity for retrieval to strengthen the form-meaning link.

2.2.7 Varied encounter and varied use

The third condition occurs when a word is subsequently met or used in different contexts or forms (Nation & Webb, 2017). Hall (1992) investigated the effect of split information tasks (i.e., pair discussion), instructed tasks and individual work on learning mathematics vocabulary. The results showed that discussion tasks were the most effective. A possible explanation is that the new meeting forces learners to retrieve and rethink previous forms and uses of a word before a new knowledge is reconstructed from the new context. This leads to a strengthening effect as a word is retrieved and enriched by the variation in encounter or use.

2.2.8 Elaboration

Elaboration means encountering different meanings, forms and uses of a word

that later leads to enrichment in different aspects of vocabulary knowledge. Deighton (1960) suggested that learners gradually acquire meaning(s) of a word throughout their numerous experiences in different contexts with that word. Indeed, Di Vesta and Peverly (1984) found that encountering various semantic aspect of words produced superior retention than mere studying their definitions. However, excessive elaboration can have negative effect. Ellis (1995) conducted two treatments: one includes input of semantic explanations in text and the other was provided orally as a result of students' requests. He found that acquisition was more likely to take place with shorter and simpler definitions. This suggested that too much semantic elaboration could overload learners and impede learning through oral input.

2.3 Technique Feature Analysis

The features include: the main aims of said activity, its condition regarding learning, any indications that said learning happens at all, and how the activity is designed (Nation, 2001).

The TFA outlines five categories that dictate word learning efficiency, including motivation, noticing, retrieval, generation and retention. Each component is subdivided into 18 features in the form of questions and the effectiveness of a task is based on how well it scores. Scores are added or deduced (assigned to 0) in accordance with whether a feature is present or absent, and a task can get a maximum score of 18. (Figure 2)

Figure 2

Technique Analysis Hypothesis (Nation & Webb, 2011)

Criteria	Scores	
Motivation		
Is there a clear vocabulary learning goal?	0	1
Does the activity motivate learning?	0	1
Do the learners select the words?	0	1
Noticing		
Does the activity focus attention on the target words?	0	1
Does the activity raise awareness of new vocabulary learning?	0	1
Does the activity involve negotiation?	0	1
Retrieval		
Does the activity involve retrieval of the words?	0	1
Is it productive retrieval?	0	1
Is it recall?	0	1
Are there multiple retrievals of each word?	0	1
Is there spacing between retrievals?	0	1
Generation		
Does the activity involve generative use?	0	1
Is it productive?	0	1
Is there a marked change that involves the use of other words?	0	1
Retention		
Does the activity ensure successful linking of form and meaning?	0	1
Does the activity involve instantiation?	0	1
Does the activity involve imaging?	0	1
Does the activity avoid interference?	0	1
Maximum score		18

2.4 The present study

This study investigates the following questions:

1. The effects of vocabulary tasks during vocabulary learning as measured in an immediate test and a delayed test?

2. To what extent does the Techniques Feature Analysis predict the effectiveness of vocabulary tasks on learning and retention?

3. METHODOLOGY

3.1 Participants

The participants in this research were 38 Vietnamese students (30 females, 8 males, age from 20 to 28). All the participants achieved an IELTS score of 5.5 and above.

The participants are recruited randomly via the researcher's network and social media. As the representativeness of the sample and its size is crucial to a robust research conclusion, random sampling is used because it minimizes the effect of subjectivity; hence, the subjects would be more representative and similar to the population than non-random samples (Dörnyei, 2007).

3.2 Procedure

The study had three main procedures. The participants do the three vocabulary tasks in order. All participants were asked to read the texts and perform the tasks in the allotted time of 30 minutes (10 minutes for each task). After each assigned task, the participants were tested immediately to measure their learning of target words. A delayed post-test was conducted a week later to evaluate their vocabulary retention.

3.3 Design

The study was adopted within-group design, Vocabulary Treatment (three vocabulary-oriented tasks) and Time (immediately and one week after the experiment).

The studies adopted an intentional learning design in which participants were informed about the vocabulary tests after they performed the treatments. The target words in the texts were replaced by pseudowords that resemble English words phonetically and orthographically. The 15 pseudowords were selected from a series of studies by Webb (2007a, 2007b, 2008a, 2008b) for reasons of reliability. Their forms and meanings are as follows: bandet (bank), masco (peep), dangy (daisy), denet (rabbit), hodet (hole), ictay (pressure), denent (slums), copac (terrible), gishom (shocking), ancon (inhabitant), faddam (urbanisation), nasin (friendship), shoten (sibling), intay (interaction), pathen (parents).

3.4 Vocabulary tasks

As the study aims to compare the effectiveness of word-focused activities, three tasks were designed with great consistency and control. Accordingly, the tasks must obtain different TFA scores, lead to intentional vocabulary learning and must all have new word learning feature. The researcher came up with three reading-based exercises which involved the learning of five new words in each passage and the TFA scores were 4, 5, and 7 respectively (Table 1). Each target word appeared once in the passages.

The reading passages include an extract from Alice's Adventures in Wonderland (Lewis, 1920) and two reading texts about Urbanization and Family taken from Vocabulary

for IELTS (Cullen, 2008). Considerations were made on the running words of the text as vocabulary knowledge is widely acknowledged to have a primary effect on reading performance and several studies have shown their strong correlation (Laufer, 1992; Nation & Hu, 2000; Hirsh & Nation, 1992). Three reading texts were selected and graded in order to ensure that the texts are comprehensible to the participants. Laufer (1992) suggests that readers need to know at least 95% of the words in a text to have any chance of guessing the meaning of unknown words from context. Hu and Nation (2000) suggest a 98% coverage is needed for adequate comprehension. However, the study used a 600-word-text, considering the short length of the passages in this study, ranging from 170 to 250 words for each text, thus a 95 % coverage is an acceptable threshold.

Consonant with the fact that the subjects must know at least 3,000 word families, 95% of the texts should include this high-frequency vocabulary level. The researcher excluded all target words in three texts and conducted a lexical frequency profile analysis via Vocabprofilers on Tom Cobb's Website (http://lextutor.ca/vp/). Vocabprofile (Cobb, 2002) calculates how many percent of words in a text falling into the first 10,000 most frequent words. Results show that the original texts did not meet the required coverage therefore, so the texts were adapted to meet the 95% coverage within the 3,000 word-family level by paraphrasing and replacing less frequency words with high frequency words.

Task 1: Read and write with target words

The participants read an extract from *Alice's Adventure in Wonderland* and performed two tasks: filling in the blanks and matching the target words with pictures depicting their meanings. Pictures were taken from the animated Alice in Wonderland to preserve contextual consistency. The target words are bandet (bank), masco (peep), dangy (daisy), denet (rabbit), hodet (hole).

Task 2: Read plus fill in

Task 2 involves a reading text with blanks replacing the target words. The target words were glossed under the passage. Participants selected the suitable words and filled in the blanks. The target words are denent (slums), copac (terrible), gishom (shocking), ancon (inhabitant), faddam (urbanisation).

Task 3: Read and select word meaning

The participants read a full text. The target words and their definitions were provided disorderly in a separate table. Based on the context of the reading passage, they had to match the target words with the equivalent definitions. The target words are ictay (pressure), nasin (friendship), shoten (sibling), intay (interaction), pathen (parents).

Table 1

Scoring of vocabulary tasks

Criteria	**Task 1**	**Task 2**	**Task 3**
Motivation			
Is there a clear vocabulary learning goal?	1	1	1
Dose the activity motivate learning?	1	1	1
Do the learners select the words?	0	0	0
Noticing			

Does the activity focus attention on the target words?	1	1	1
Does the activity raise awareness of new vocabulary learning?	1	1	1
Does the activity involve negotiation?	0	0	0
Retrieval			
Does the activity involve retrieval of the words?	1	0	0
Is it productive retrieval?	0	0	0
Is it recall?	0	0	0
Are there multiple retrievals of each word?	0	0	0
Is there spacing between retrievals?	0	0	0
Generation			
Does the activity involve generative use?	0	0	0
Is it productive?	0	0	0
Is there a marked change that involves the use of other words?	0	0	0
Retention			
Does the activity ensure successful linking of form and meaning?	1	0	0
Does the activity involve instantiation?	0	0	0
Does the activity involve imaging?	1	0	0
Does the activity avoid interference?	1	1	0
Total score	7	5	4

3.5 Tests

The research adopts two tests: immediate tests and delayed tests which measuring recognition of form and meaning in multiple-choice formats. The form recognition tests preceded the meaning recognition tests to prevent the subjects from familiarising themselves with the word form. As the immediate tests and post-tests were identical, the test questions were rearranged to prevent any chances of memorisation.

The first test measured receptive knowledge of form; the participants had to circle the correctly spelled target words, which appeared with three distracters. The distracters were created to resemble the target words both phonetically and orthographically. In Figure 3, the target words are *dangy* and *bandet*.

Figure 3

Immediate recognition of form test

Circle the correct words in the text

1.	2.
a. dengie	a. bandat
b.dengy	b. bandet
c. dangie	c. bendet
d. dangy	d. bendat

The second test measured receptive knowledge of meaning; the participants had to circle the correct meaning of target words. The distracters were taken The aim of this test was to determine if the participants can guess the meaning from context and the tasks.

Figure 4

Immediate recognition of meaning test

Choose the correct meaning of the italic words

1. bandet

a. tree

b. riverbank

c. stone

d. hole

2. dangy

a. yellow flowers

b. red flowers

c. orange flowers

d. white flowers

4. RESULTS

This study adopts the analysis of variance (ANOVA) to determine any differences between the mean scores of each task.

The immediate and post tests were scored dichotomously in which every correct and incorrect answer was assigned a 1 or 0 point respectively. The test scores were analysed using repeated measure with test scores as a within-subject variable.

Wilk's Lambda was used to determine statistical significance ($p<.05$) as recommended by Pallant (2010). Effect size is also assessed to measure the degree of association between the three sets of scores. The effect size is evaluated following guidelines proposed by Cohen (1988, as cited in Pallant, 2010). For repeated measure ANOVA, Cohen (1988) suggested that $\eta p2$ (Partial eta squared) = 0.01 be considered a small effect size, 0.06 represents a moderate effect size and 0.138 a large effect size. This means that if the effect size is no greater than 0.01, the difference is trivial even though there is statistical significance.

The results will be presented and discussed in terms of the remaining research questions:

1. The effects of vocabulary tasks during vocabulary learning as measured in an immediate test and a delayed test?

A one-way repeated measure ANOVA was conducted to compare scores on the immediate tests and delayed post-tests. The means and standard deviation are presented in Table 2.

Table 2

Descriptive statistics of task scores

Condition	N	TFA scores	Form recognition				Meaning recognition			
			Immediate		Delayed		Immediate		Delayed	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Task 1	38	8	4.55	.60	3.21	.96	4.58	.79	4.16	1.05
Task 2	38	6	4.71	.56	3.55	1.06	4.61	.49	3.47	1.03
Task 3	38	4	4.39	.95	3.47	1.18	4.26	1.18	2.74	1.55

The results showed that the participants learned a substantial number of words. Mean scores of the immediate tests remained above 4 out of 5 points scale and there was not a big gap between immediate form and meaning recognition test scores. Task 2 yielded the highest average scores in the Immediate form, Immediate meaning and Form recognition delayed tests with the mean scores of 4.71, 4.61 and 3.55 respectively. Task 3 ranked third and only had the second highest scores on the form recognition delayed test.

Table 2 compares the means of delayed and immediate tests. The participants scored higher on the immediate than on the delayed tests. Task 2 has the highest scores of form retention, followed by Task 3 and Task 1. The retention of word meaning scores the highest in Task 1 and the least in Task 3.

To compare the effectiveness of the three tasks, inferential statistics were conducted. As the research compared the learning effect of three vocabulary tasks, one-way repeated ANOVA was carried out.

Mauchly's test of the Immediate form recognition test indicated that the sphericity assumption is not rejected $\chi^2(2) = 1.195$, $p = .550$; therefore, the results show that there was no statistically significant difference between three conditions on the learning of word form. F (2,74) = 1,978, p = 0,146. Wilks' Lambda = 0.91, F (2, 36) = 1.83, p= 0.174. The effect size is moderate ($\eta p^2 = 0.098$).

As for the Immediate meaning recognition test scores, Mauchly's test showed that the assumption of sphericity has been violated. The results show that there was no overall statistically significant difference between three treatments on the learning of word meaning immediately, F (1.72; 63.51) = 1.78, p = 0,181. Wilks' Lambda = 0.93, F (2, 36) = 1.3, p= 0.283, multivariate spatial eta square =.068, which is a moderate effect size.

In regards to the delayed form recognition tests, Mauchly's test showed no violation of the spherical assumption, $\chi^2(2) = 3.206$, $p = .201$. Results indicated that the vocabulary exercises have little effect on the retention of word form, F (2,74) = 1,978, p = 0,146. Wilks' Lambda = 0.9, F (2, 36) = 1.9, p= 0.164, multivariate spatial eta square =.068, which is a very large effect size.

Mauchly's test on the delayed meaning recognition post-tests showed that the data does not meet the spherical assumption, $\chi^2(2) = 7.898$, $p = .019$. Results of Greenhouse-Geisser ($\varepsilon = 0.835$) indicated statistically significant difference between the three tasks, meaning that the tasks had effects on the retention of word meaning, F (1.67, 61.82) = 19.832, p= .000. Wilks' Lambda = 0.55, F (2, 36) = 14.53, p= .000. Given in the Multivariate Tests output, Partial Eta Squared is $\eta p^2 = .447$, this result suggests a very large effect size.

Table 3

Pairwise comparisons between delayed meaning tests

(I) Delayed meaning test	(J) Delayed meaning test	Mean Difference (I-J)	Sig.
1	2	.684*	.001
	3	1.421*	.000
2	1	-.684*	.001
	3	.737*	.007
3	1	-1.421*	.000
	2	-.737*	.007

Post hoc test using the Bonferroni correction was generated to determine the degree of significant difference between each task. Data from Table 3 revealed that participants retained more words from Task 1 (mean= 4.16; SD = 1.05) compared to Task 2 (mean= 3.47; SD = 1.03; p=.001) and Task 3 (mean= 2.74; SD = 1.55; p<.001). Task 3 was significantly less effective than Task 2 (p= .0.007), indicating that Task 3 resulted in the least retention of word meaning.

To conclude, data analysis found no statistically significant difference in the immediate tests and delayed form recognition test. There were task effects in the delayed meaning recognition test in which each task was significantly different from the other.

Research question 2: To what extent does the TFA framework predict the effectiveness of vocabulary tasks on learning and retention?

The scoring of TFA does not align with the descriptive statistic for immediate recognition tests. Task 2 has lower TFA scores than Task 1; however, its mean scores were the highest. Meanwhile, the subjects scored the lowest in Task 3. Therefore, the TFA does not consistently predict which task is more effective than one another in terms of immediate meaning and form recognition. Results from immediate tests yielded no significant difference among the tasks.

Similar findings can be found in the case of delayed form recognition knowledge in which no statistical significance was found between the tasks. However, the difference lies in the analysis of delayed tests results which indicated the TFA scores correlate with the retention of word meaning as participants fared best in Task 1, scored lower in Task 2 and lowest in Task 3. Therefore, the research found that the framework only predicts the retention of vocabulary meaning.

To sum up, findings revealed that learning effect only differs in delayed meaning recognition tests while results from other tests have no significant difference among the tasks.

5. DISCUSSION

The research aimed to investigate the predictive power of TFA framework by comparing the effect of three vocabulary tasks on learning and retention of 15 target words. Each task involved a reading text and a follow-up exercise. Considering all target words were unknown to the participants, three tasks led to new word learning.

The hypothesis is that the TFA scores correlate with task effectiveness, meaning that learners were expected to score highest in Task 1 and lowest in Task 3. However, analysis found a lack of significant difference in the scores of immediate form and meaning recognition knowledge. One possible explanation is that the tasks were effective as they were based on TFA framework and obtained all or most of the features that aid learning (Motivation, Noticing, Retrieval, Generation and Retention). In specific, all tasks shared the Motivation, Noticing and Retention features in common. As mentioned in the literature review, all of these features are facilitative to vocabulary learning. Another reason may lie in the nature of tests in which the participants have to recognise word form and its meaning. Task 2 required filling in the blanks with glossed words. Task 3, on the other hand, consisted of matching a word's meaning. The tests tapped into the same aspects of word knowledge (i.e. form and meaning); hence, the task-test similarity resulted in less of a distinction in test performance. Participants had already worked with the target words' meaning in Task 2 and Task 3. Therefore, to some extent, the multiple-choice tests may have favoured them in recognising the correct answers.

Mean scores revealed that performance in Task 2 fared the best. The washback effect stemming from the research procedure may be the cause. The subjects were aware of the test format after Task 1 and they may pay more attention to the word form-meaning in Task 2 and Task 3. The researcher anticipated this backwash effect and shuffled the task order. Therefore, it was unlikely that the order impacted task performance and the difference stemmed from the task's nature itself. One reason for higher scores in Task 2 is that although Task 1 and Task 2 involve noticing feature, Task 2 is similar to reading with glosses and blank-filling as definitions of target words are provided and the subjects have to do a subsequent exercise. Many empirical studies have found glossing superior over non-glossing activities in immediate vocabulary learning (Hulstijn, 1992; Miyasako, 2002; Watanabe, 1997). Watanabe (1997) examined how text modifications and task affect vocabulary learning and found that Japanese students remembered less words in text with no cue condition than marginal gloss condition. According to Nation (2013) and Watanabe (1997), exercises with glossing are beneficial in vocabulary learning for providing accurate meaning, reducing the chance of incorrect inferences and drawing attention to unknown words. Therefore, this can be the reason why learners scored higher in Task 2 as it had deeper degree of attention.

Significant difference only occurred in the delayed meaning recognition test; hence, providing evidence for the TFA in the effectiveness of vocabulary retention. Several explanations can be given referring to the level of processing. In Task 1, the subjects had to select the correct word form and infer word meaning from the text to complete two sub-tasks. On pedagogical grounds, Kelly (1989) and Laufer and Sim (1985) object to learning from contextual guessing because context hardly provides sufficient information for correct meaning inference and learners may consequently learn the wrong meaning. However, these studies inspected inferring-from-context method under a reading only condition (i.e. no subsequent vocabulary exercises). Hulstijn (1992) and Koren (1999) concluded treatments that encouraged lexical inferencing led to higher level of retention

than glossing, which can be seen in the results of post tests. Initially, it can be assumed that fill-in-the-blank exercises in Task 1 is a passive use of vocabulary as the subjects only searched word form and filled in the blanks. However, Folse (2006) argued that this exercise involves various thinking processes: learners can try out different words in a slot, probably "by translating many of the words or perhaps by remembering tidbits about some of the words" (p.287). In his opinion, this is indeed deep processing of the word which facilitates retention. Another possible reason can be found in the second sub-task – matching pictures and vocabulary. Studies have showed that words that are strongly associated with images are more likely to be remembered (Underwood, 1989) and recognised (Yanguas, 2009). This has been referred to as the *picture superiority effect* which assumes that pictures are more elaborative and elaboration is a facilitator to vocabulary learning. In a study by Carpenter and Olson (2012), vocabulary learning from picture-word pairing was more effective than words translated in native language.

Finally, the number of word encounters may lead to better meaning retention for Task 1 as learners had to read a text (first meeting), filled in the blank (second meeting) and matched pictures (third meeting). The two sub-tasks also enhanced the connection of lexical form and meaning – that is fill-in-the-blank requires word form search and decision-making on semantic appropriateness while picture matching consists of retrieving meaning for the given word forms. Peters (2012) found that learners retained more words when there were two retrievals in one vocabulary task. As mentioned before, multiple retrievals and encounter opportunities is a strong and facilitative factor of vocabulary acquisition.

6. IMPLICATIONS AND CONCLUSION

This research investigates whether the TFA framework can predict the effect of vocabulary tasks on the learning and retention among Vietnamese learners.

The study showed that there was no difference between a task's effect on the learning of a word's meaning and word form with the retention of word form. One possible explanation is that the three tasks were equally effective as they shared some TFA features in common. Despite having lower TFA scores than Task 1, higher mean scores in Task 2 are due to its resemblance with glossing activity and task-test similarity. Supporting evidence for the framework was found in the delayed meaning recognition tests. This is because Task 1 involved multiple retrievals, deeper processing level and picture superior effect, which are factors that aid vocabulary learning.

There are several pedagogical implications regarding the design of vocabulary exercises.

First, findings from the studies suggest that the design of vocabulary exercises can be based on the TFA framework. There are some components (i.e. motivation, spacing retrievals, marked change in generative use) that require long-term students observation from the teachers in order to design an effective exercise.

Second, it is better to combine exercises for better learning effect. In the study, Task 1 (Reading plus fill-in and picture-matching) facilitates better retention while Task

2 (Reading plus text-fill-in) has greater immediate learning effect. Getting learners to perform the two tasks simultaneously is more likely to result in their better learning and retaining more words.

Certain limitations can be identified in this study. This research is restrained to measuring the recognition of form and meaning knowledge. A more comprehensive approach would be testing the recall of other aspects (i.e. orthography, grammatical functions, and meaning and form).

The small sample size and time are another limitation. As the research was conducted online and the researcher had to give instructions to the participant individually, which limited the number of people approached within the time constraints. Besides, a within group design would provide more accurate data when conducted in longitudinal research. For more cumulative effect and perhaps more differences in the learning gains, future research could have larger sample sizes, provide longer time for post-tests and use more tasks with similar and different TFA features which is then followed by tests of productive word knowledge (i.e., form, meaning, grammatical function).

REFERENCES

Barcroft, J., (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*, 323–363.

Barcroft, J., (2009). Effects of Synonym Generation on Incidental and Intentional L2 Vocabulary Learning During Reading. *TESOL Quarterly*, *43*(1), 79-103.

Carpenter, S., & Olson, K., (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92-101.

Craik, F. I. M., & Lockhart, R. S., (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Behavior*, *11*, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Di Vesta, F. J., & Peverly, S. T. (1984). The effects of encoding variability, processing activity, and rule–examples sequence on the transfer of conceptual rules. *Journal of Educational Psychology, 76*(1), 108–119.

Dornyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL'. *Computer Assisted Language Learning, 8*(2-3), 103-128.

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

Goossens, Nicole A. M. C., et al., (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*(8), 965-971.

Hall, S., (1992). Using Split Information Tasks to Learn Mathematics Vocabulary. *Guidelines*, *14*(1), 72-77.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. Language learning, *51*(3), 539-558.

Hulstijn, J.H., (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In: Pierre J. L. Arnaud and Henri, B., eds. Vocabulary and Applied Linguistics, London: Macmillan, 1992, 113–125.

Karpicke, J. D., & Roediger, H. L. (2009). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*(1), 116–124. https://doi.org/10.3758/mc.38.1.116

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. https://doi.org/10.1177/1362168808089922

Kida, S., & Barcroft, J. (2018). Semantic and Structural Tasks for the Mapping Component of L2 Vocabulary Learning: Testing the TOPRA Model From A New Angle. *Studies in Second Language Acquisition, 40*, 477–502.

Koren, S. (1999). Vocabulary instruction through hypertext. Are there advantages over conventional teaching method. *TESL-EJ*, *4*(1), 77-84.

Kornell, N., (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*, 1297–1317.

Laufer, B. (2019). Evaluating Exercises for Learning Vocabulary. In: Webb, S., ed. The Routledge Handbook of Vocabulary Studies, Milton: Taylor and Francis Group, 2019, 351- 368.

Laufer, B., & Sim, D.D., (1985). Taking the easy way out: non-use and misuse of clues in EFL reading. *English Teaching Forum, 23*(4), 7-10.

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. https://doi.org/10.1177/1362168814559797

Laufer, B., (1997). The lexical plight in second language reading: words you don't know, words you think you know and words you can't guess. In: Coady, J., Huckin, T., eds. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy.* Cambridge: Cambridge University Press, 1997, 20–34.

Lin, L. H. F., & Morrison, B., (2010). The impact of the medium of instruction in Hong Kong secondary schools on tertiary students' vocabulary. *Journal of English for Academic Purpose*, *9*, 255– 266.

Malone, J., (2018). Incidental vocabulary learning in SLA. *Studies in Second Language Acquisition*, *40*, 651–675. https://doi.org/10.1017/S0272263117000341

Nation, I. S. P., (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16.

Nation, I.S.P., & Hunston, S., (2013). *Learning Vocabulary in Another Language*. 2nd ed. Cambridge Applied Linguistics. Cambridge: Cambridge University Press.

Nation, P., (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Oller, J. W., & Krashen, S. D. (1988). The Input Hypothesis: Issues and Implications. *Language*, *64*(1), 171. https://doi.org/10.2307/414800

Pallant, J., (2010). *SPSS survival manual : a step by step guide to data analysis using SPSS* 4th ed.. Maidenhead: McGraw-Hill.

Peters, E., (2012). The differential effects of two vocabulary instruction methods on EFL word learning: A study into task effectiveness. *IRAL - International Review of Applied Linguistics in Language Teaching*, *50*, 213-238.

Pyc, M. A., & Rawson, K. A., (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory and Cognition*, 35, 1917–1927

Robinson, P., (1995). Attention, memory and the ‘‘noticing’’ hypothesis. *Language Learning, 45*, 283–331.

Rutherford, W., & Smith, S., (1985). Consciousness-raising and universal grammar. *Applied Linguistics*, *6*, 274-82.

Schmidt, R., (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*, 129-158.

Shu, H., Anderson, R.C., & Zhang, Z., (1995). Incidental learning of word meanings while reading: a Chinese and American cross-cultural study. *Reading Research Quarterly*, *30*, 76–95.

Uchihara, T., Webb, S., & Yanagisawa, A., (2019). How useful is repeated exposure to words for vocabulary learning? *OASIS Summary* of Uchihara et al. in *Language Learning*. https://oasis-database.org

Underwood, J., (1989). HyperCard and interactivevideo. *CALICO* 6(3), 7–20.

Vidal, K., (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. doi: 10.1111/j.1467-9922.2010.00593.x

Watanabe, Y., (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*(3), 287-307

Webb, S., (2007). Learning word pairs and glossed sentences: The effects of a single sentence on vocabulary knowledge. *Language Teaching Research, 11*, 63–81.

Webb, S., (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics, 28*(1), *46-65.* https://doi.org/10.1093/applin/aml048

Webb, S., (2007). The effects of synonymy on vocabulary learning. *Reading in a Foreign Language*, *19*, 120–136.

Webb, S., (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*, 79–95.

Webb, S., (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*. *20*(2), 232–245.

Yanguas, I., (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Lang. Learn. Technol, 13*(2), 48–76.