Modern English Education

http://dx.doi.org/10.18095/meeso.2019.20.2.54

Efficacy of Task Type and Involvement Load on EFL Learners' Vocabulary Knowledge*

Bei-Bei Xu**

Korea University

ARTICLE INFO

Received 19 March 2019
Revised 30 April 2019
Accepted 17 May 2019

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

KEYWORD

L2 vocabulary learning/
task type/
involvement load/
L2 어휘학습/
과업유형/
개입부하

ABSTRACT

Xu, Bei-Bei. (2019). Efficacy of task type and involvement load on EFL learners' vocabulary knowledge. *Modern English Education, 20*(2), 54-65.

This study attempted to investigate how task type and involvement load affected EFL learners' receptive and productive lexical knowledge across time. A total of 74 participants were randomly assigned to one of four groups and each group learned 6 target words through reading with one of four form-focused tasks inducing different involvement load. The lexical gains of both receptive and productive knowledge were measured by immediate and delayed posttests. It was found that task type had a significant effect on the acquisition of L2 receptive knowledge across time. In terms of productive vocabulary knowledge, the effect of task type was only significant in the immediate test but not significant in the delayed test. The involvement load had no significant effect on the acquisition of L2 receptive and productive vocabulary knowledge across time. This result did not support the Involvement Load Hypothesis, indicating that the theory did not hold true across L2 vocabulary learning conditions and the three components of task-induced involvement, i.e., *need, search* and *evaluation,* might give no equal weight to their contributions to L2 vocabulary acquisition.

I. INTRODUCTION

Being regarded as the cornerstone of all language learning activities, vocabulary plays a vital role in second language acquisition and development (Nation, 2001; Pulido, 2009; Schmitt, 2008). Some scholars agree that extensive reading and incidental learning lead to a valuable accumulation of vocabulary (Nagy, Herman, & Anderson, 1985; Pitts, White, & Krashen, 1989; Wesche & Paribakht, 1996). Other researchers, however, have proved that incidental vocabulary learning from reading is discouragingly low-efficient for L2 learners (Horst, 2005; Horst, Cobb, & Meara, 1998; Hulstijn, 1992; Laufer, 2003; Waring & Takaki, 2003). One reason is that L2 learners have to pass a threshold for incidental vocabulary learning (Coady, 1997; Nation, 2001, 2006; Schmitt, 2008). Laufer's (2000) research showed that the average vocabulary size of EFL learners from different countries fell well short of the vocabulary threshold. It is not surprising that most L2 learners see vocabulary accruement as a substantial diffi-

* This work was supported by the China Scholarship Council (No. 201508260003).
** Author: Bei-Bei Xu (Korea University, PhD Student)

Bei-Bei Xu
Department of English Education, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea
Tel: (02) 3290-2350 / Email: hsubeibei@gmail.com

culty in learning a foreign language (Laufer & Hulstijn, 2001; Schmitt, 2008). Another explanation may be the low level of individual engagement with the new word in reading-only activities (Schmitt, 2008). Craik and Lockhart's (1972) cognitive theory of depth of processing was widely accepted and manipulated to precisely conceptualize 'engagement' and the notion assumed that the more deeply a word was processed, the more effectively it would be learned (Ellis, 1994; Laufer & Hulstijn, 2001; Pulido, 2009).

A number of researchers have thus argued that vocabulary enhancement techniques and form-focused vocabulary tasks, which induce a certain depth of processing, can help promote effective and efficient lexical learning for L2 learners (Folse, 2006; Hu & Nassaji, 2012; Huang, Wilson, & Eslami, 2012; Hulstijn & Laufer, 2001; Keating, 2008; Laufer, 2003, 2005; Laufer & Girsai, 2008; Min, 2008; Nassaji, 2003). For instance, some vocabulary enhancement tasks such as reading plus dictionary look-up, reading with word-focused vocabulary exercises were empirically proved to outperform reading-only activities in promoting L2 learners' lexical knowledge (K. S. Cho & Krashen, 1994; Knight, 1994; Paribakht & Wesche, 1997). In light of these findings, researchers generally reached a consensus that form-focused instruction was indispensable for L2 vocabulary acquisition (Keating, 2008; Laufer, 2005; Schmitt, 2008). Nonetheless, some word-focused tasks were demonstrated to be more beneficial than others. For example, Joe (1995, 1998) found that using the target words in original sentences was superior to using them in unoriginal sentences in terms of vocabulary knowledge gains. Hill and Laufer (2003) proved that dictionary look-up tasks contributed more to word retention than reading with glosses.

For the purpose of explaining why certain tasks produced greater gains in vocabulary acquisition than others, Laufer and Hulstijn (2001) proposed the Involvement Load hypothesis (ILH), which contended that the acquisition and retention of a new word were conditional upon task-induced involvement, i.e., the higher the involvement load, the more the vocabulary knowledge would be acquired. The notion of task-induced involvement makes it possible to operationalize the general cognitive concept of elaboration and depth of processing from the perspective of form-focused vocabulary tasks. The proposal of ILH has triggered much research on form-focused vocabulary task effectiveness in terms of different involvement load indices (e.g., Bao, 2015; Bruton, 2007; Hulstijn & Laufer, 2001; Keating, 2008; Y. J. Kim, 2008; Laufer, 2003, 2005; Zou, 2017), and it still remains inconclusive whether task-induced involvement has better explanatory power for L2 lexical retention and retrieval. The task-induced involvement is a three-dimensional construct, with *need, search* and *evaluation* as its components. The ILH also hypothesizes that the presence of one component is always better than its absence and the same component always leads to the same amount of learning. The research conducted on this issue was far from sufficiency and most previous studies were concerned with *evaluation*. Another constructing component *search*, therefore, was examined in this study. The operational research questions were as follows:

1) How do task type and involvement load affect L2 learners' immediate and delayed vocabulary acquisition?
2) Is there any interaction effect between task type and involvement load?
3) How does the constructing component *search* of task-induced involvement contribute to the degrees of L2 vocabulary acquisition?

II. LITERATURE REVIEW

Laufer and Hulstijn (2001) proposed ILH to capture the cognitive notion of elaboration (Craik & Tulving, 1975) and depth of processing (Craik & Lockhart, 1972) by conceptualizing the motivational-cognitive dimensions of task-induced involvement. According to ILH, the construct of task-induced involvement includes three factors, namely *need, search* and *evaluation*, which are quantifiable to show different prominence degrees and determine the total involvement load of a given task (Laufer & Hulstijn, 2001). As the motivational, non-cognitive element, *need* is the intrinsic or extrinsic drive to finish a task, with its presence/absence and prominence degrees indexed as no *need* (need− = load index 0), moderate *need* (need+ = load index 1) and strong *need* (need++ = load index 2). As the cognitive dimension, *search* and *evaluation* involve attentions related to the meaning and form of vocabulary knowledge. *Search* is the effort to find the meaning or the form of unknown words, dichotomously indexed either as no *search* (search− = load index 0) or the presence of *search* (search+ = load index 1). *Evaluation* is comparing a given word with other words and deciding its proper use in a specific context, indexed as no evaluation (evaluation− = load index 0), moderated evaluation (evaluation+ = load index 1), or strong evaluation (evaluation++ = load index 2). The ILH assumes that higher involvement indices will result in better learning in L2 vocabulary knowledge.

There emerged a number of empirical studies to investigate the theory of ILH by examining the efficacy of form-focused tasks inducing different involvement load indices on L2 vocabulary learning and retention, but the results were inconclusive. On the one hand, some studies provide confirmatory evidence for the predicative power of ILH. Hulstijn and Laufer (2001) compared three tasks with different involvement load indices: reading with an index of 1 (need+, search−, evaluation−), reading and gap-filling with an index of 2 (need+, search−, evaluation+), composition writing with an index of 3 (need+, search−, evaluation++), and they found that composition writing which induced the highest involvement load was more conducive to both the initial learning and retention of unknown words than the other two with different lower

loads. Incorporating proficiency level as a moderator, Y. J. Kim (2008) partially replicated the experiment done by the ILH proposers and extended the predicative power of ILH to lower-proficient learners. Keating (2008) found sentence writing (need+, search−, evaluation++), which induced the same involvement elements and the same total load index of 3 as composition writing, also led to better gains in lexical retention than cloze activities. Y. J. Kim's (2008) experiment comparing sentence writing and composition writing confirmed the ILH assumption that the same involvement load would result in the same acquisition regardless of task type.

On the other hand, there are also some studies supplying unsupportive results. Although most studies favored the efficacy of output writing tasks over other tasks such as cloze activities (e.g., Bao, 2015; Eckerth & Tavakoli, 2012; Hulstijn & Laufer, 2001; Keating, 2008; Y. J. Kim, 2008), Folse (2006) suggested that sentence writing task which induced strong evaluation was equally effective as gap-filling task which induced moderate evaluation. This finding was further evidenced by Laufer's (2003) research which indicated that sentence writing was no more effective than sentence completion plus dictionary use (need+, search+, evaluation+), although both of them induced the same involvement load index of 3. Contradictory with Y. J. Kim's (2008) research, this study suggested that different factors might contribute differently to task-induced involvement. Meanwhile, cloze activities (involvement load index = 2) was not statistically proved to be better than the reading-only task (involvement load index = 1), especially in terms of L2 word retention (Hulstijn & Laufer, 2001; Y. J. Kim, 2008). The above studies suggested that the same involvement load did not lead to similar word learning, nor did the higher involvement loads necessarily lead to better word learning (Bao, 2015).

In a nutshell, the ILH has yet to be completely confirmed by empirical studies in terms of form-focused task effect on L2 vocabulary acquisition in the English as a Second Language (ESL) context. As a matter of fact, if we expand the examination of ILH from ESL context to EFL context, the research results could become rather complicated by differing from each other so much: some fully support the hypothesis (e.g., H. S. Kim & Y. H. Na, 2010), some provide partial evidence (e.g., J. U. Hyun & J. Y. Kim, 2016), and some provide no evidence (e.g., Bao, 2015; M. H. Kang, 2015). Although proficiency level is not regarded as a significant moderator (Keating, 2008; Y. J. Kim, 2008) in ESL context, H. S. Kim and Y. H. Na (2010) noted that task efficacy differed significantly from low-proficiency students to high-proficiency students.

As for the reasons giving rise to these conflicting results, there are two kinds of explanations. One explanation approaches the problem from the external perspective, for example, the differences in research design. Webb (2005) found that task design, task time and type of vocabulary measurement would partly determine rejecting and accepting ILH. With task time controlled, Folse (2006) found that three gap-filling activities were more effective and efficient than one sentence-writing activity. Bao (2015) noted that the type of vocabulary knowledge measured might lead to different findings, for example, sentence writing tasks (need+, search−, evaluation++) was not more effective than translation tasks (need+, search−, evaluation+) in terms of productive vocabulary knowledge measured. In line with the above-mentioned limitations, this study considered both the receptive vocabulary knowledge and the productive vocabulary knowledge with time controlled. Following Min (2008), receptive vocabulary knowledge was defined as the "ability to provide a correct translation equivalent of a target words" (p. 81), whereas productive vocabulary knowledge referred to using a target word appropriately in a sentence.

The other explanation perceives the inconsistency of the research findings from the internal perspective, i.e. the deficiency of ILH itself. For instance, many researchers suggested that the number of word exposures (e.g., Folse, 2006; Laufer & Rozovski-Roitblat, 2011) and different weights of need, search and evaluation might be important factors contributing to word retention (e.g., Bao, 2015; J. U. Hyun & J. Y. Kim, 2016). With regard to the three elements of task-induced involvement, most studies focused on the element of *evaluation.* Y. J. Kim (2008) and Zou (2017) noted strong evaluation might be far more important than moderate evaluation for L2 vocabulary acquisition and retention. The contribution of *need* and *search* to task-induced involvement was less studied in the existing literature, which would be addressed in the present analysis.

The present research was mainly concerned with the effectiveness of output tasks on receptive and productive vocabulary knowledge. In addition, most previous researches were interested in varying the *evaluation* indexes with *need* and *search* held constant (e.g., Bao, 2015; J. U. Hyun & J. Y. Kim, 2016). Given the fact that each involvement load component might give different weights in contributing to L2 vocabulary acquisition, whether the previous findings concerning *evaluation* could be generalized to other components (i.e., *need* and *search*) remained to be confirmed. In line with the limitations of the previous literature, this study aimed to investigate how different task types inducing various involvement loads affect the acquisition of L2 vocabulary knowledge of middle school students (Grade 8) in an English as a Foreign Language (EFL) context. Meanwhile, the constructing component *search* was manipulated as presence or absence to test its specific effect on L2 vocabulary acquisition.

III. METHOD

1. Research Design

This study adopted a 4x2 factorial design with two between-subjects factors, L2 vocabulary acquisition being the dependent variable, which was measured as immediate recep-

tive knowledge, immediate productive knowledge, delayed receptive knowledge and delayed productive knowledge. Task type and involvement load constituted the between-subjects variables. Following Hulstijn and Laufer's (2001) tradition of using reading as stimuli, this study designed four tasks based on passage reading, i.e., three sentence translation tasks and one sentence writing task. As one of output tasks widely used in EFL vocabulary teaching and learning, translation tasks were less investigated than writing tasks with only a few examples (e.g., Bao, 2015; Laufer & Girsai, 2008). The three translation tasks were divided into two type: two were English to Chinese (E-C) translations (Task 1 and Task 2) and Task 3 was Chinese to English (C-E) translation. Task 4 required the participants to do sentence-writing based on the learning activities. The four treatments induced two kinds of task-induced involvement: low (index = 3) and high (index = 4). The control group of reading-only was excluded since most of the existing studies have already proved the efficacy of form-focused tasks. All the tasks were illustrated in detail in Table 1.

TABLE 1
Task Types and Involvement Load

	Task 1	Task 2	Task 3	Task 4
Participants	Group 1 (n = 17)	Group 2 (n = 17)	Group 3 (n = 20)	Group 4 (n = 20)
Treatment	Reading (+ marginal glosses) E-C translation	Reading (+ dictionary look-up) E-C translation	Reading (+ dictionary look-up) C-E translation	Reading (+ marginal glosses) Sentence writing
Involvement load	need+ search− evaluation++ (3)	need+ search+ evaluation++ (4)	need+ search+ evaluation++ (4)	need+ search− evaluation++ (3)

Note. (3) and (4) mean involvement load indices.

2. Participants

The participants in this research were two parallel classes of Grade 8 students (n = 74) in a middle school in China. They all started to learn English when they were in Grade 3 in their primary schools, that is to say, they have already received five years of English education following the same curriculum. Before the experiment, all the participants received a pretest to ensure their vocabulary size was on the same level between groups. The vocabulary size test (see Appendix 1) consisting of 20 vocabulary items was adapted from Nation and Beglar's (2007) first 1000 and second 1000 words to measure participants vocabulary size. The pretest showed that their vocabulary size ranged from 200 words to 1000 words with an average of 587 words. The National English Curriculum for Compulsory Education (2011 version) requires 1,500-1,600 words for Grade 9 students. On this basis, the participants' vocabulary size was still limited. Each class was divided into 2 groups randomly and altogether the 4 groups completed the 4 tasks inducing two types of involvement load respectively. Therefore, there were four groups in the experiment as shown in Table 1. Table 2 showed the descriptive statistics for participants' vocabulary size across the four groups, indicating the normal distribution of scores with the Shapiro-Wilk test nonsignificant. The homogeneity of variances was examined by Levene's test ($F(3, 70) = .56$, $p > .05$). Thus, a one-way ANOVA was conducted to show there was no significant difference among the four groups in terms of vocabulary size ($F(3, 70) = .255$, $p = .857 > .05$).

TABLE 2
Descriptive Statistics for Participants' Vocabulary Size Across Groups

Group	*n*	*M*	*SD*	*Skew.2SE*	*Kurt.2SE*	Shapiro-Wilk test	
						W	*p*
1	17	5.82	.95	.39	.39	.90	.07
2	17	6.06	1.09	.20	.20	.92	.16
3	20	5.75	1.21	.34	.34	.92	.10
4	20	5.85	1.14	.32	.32	.92	.10

3. Materials

1) Reading Materials and Target Words

The reading passage (see Appendix 2) was selected from the final examination paper for Grade 8 students. The text containing 124 words and 6 target words was selected according to the following criteria. First, the participants' teacher identified the words that they might not know and then 20 students from a different class at the same grade were asked whether they knew the target words or not. The target words contained 4 nouns, 1 verb and 1 adjective: *type, cashier, downtown, decision, similar and choose.*

2) Tasks

The experimental tasks differed in treatments and degree of involvement load (see Table 1). The tasks involved in the study were divided into 4 types, depending on the induced task-involvement load with the different vocabulary enhancement measures and form-focused exercises (see Appendix 3 for examples). Task 1 involved reading comprehension with marginal glosses and word-focused exercises demanding participant to translate English sentences into Chinese with the target words in bold. Task 2 required participants to complete reading comprehension with dictionary look-up activities and then the same translating exercise as in Task 1. Task 3 and Task 4 focused on the productive output of the target words. Participants in task 3 were required to finish the reading comprehension with dictionary look-up and Chinese to English sentence translation with the given words. The reading comprehension part in Task 4 was the same as that in Task 1 but the word-focused exercise was to write original sentences with the target words. Based on ILH, the component *search* was manipulated to be absent or present to test whether it contributed to vocabulary acquisition. All the four tasks were required to finish in twenty minutes.

3) Vocabulary Test

The study employed Min's (2008) modified Vocabulary Knowledge Scale (VKS) to measure learners' receptive and productive vocabulary knowledge in immediate and delayed lexical retention. Min's VKS was adapted from Paribakht and Wesche's (1993), using a 4-point scale with self-report and performance items incorporated together. The first two categories are related to the participants' self-perceived knowledge of the target words. The third category is used to test the participant's receptive vocabulary knowledge and the fourth category for the productive vocabulary knowledge. In this study, the measurement of the EFL learners' vocabulary were related to the last two categories in Min's modified VKS. For the convenience and avoidance of misunderstanding, the instrument was translated into Chinese (see Appendix 4 for an example).

4. Procedures

The experiment was carried out during class hours on two separate days over a 2-week period. The pretest (vocabulary size test), the treatment and the immediate vocabulary knowledge test were administered on the same day in the first week. The delayed vocabulary knowledge test was carried out two weeks later. During the experiment, the EFL teacher divided each class into two groups with the same number of participants randomly and asked them to finish one of the four tasks. The sentence writing group were required not to copy the sentence in the reading passage. All the four tasks were required to finish in twenty minutes. After the treatment, the participant was given a 5-minute break and then the teacher carried out the immediate vocabulary test without announcing in advance. Two weeks later, the researcher went to the school again to deliver a delayed posttest to the same participants without announcement.

5. Scoring and Data Analysis

The scoring criteria for measuring participants' vocabulary knowledge followed Min (2008). One point was awarded to Category 3 and Category 4 if the answer was right. Category 3 was used to test the receptive knowledge of a target word and if the Chinese meaning of the target word was right, one point was given. Category 4 was used to test the learners' productive knowledge and if participants used the target word correctly in making the sentence, one point was given. The grammatical misuses and improper meaning of the target words led to zero point. One experienced EFL teacher graded the tests and the researcher checked the grading. To address the research questions, one MANOVA was performed first to find the effect of task type and involvement load; then a set of ANOVAs were conducted to detect the respective difference. All the statistical analyses were conducted via SPSS 25.

IV. RESULTS

1. Descriptive Statistics

Table 3 showed the descriptive statistics of task type effect. The task-induced involvement was controlled for two types: lower involvement load with an index of 3 (Task 1 and Task 4) and higher involvement load with an index of 4 (Task 2 and Task 3). We could see that learning gains in receptive vocabulary knowledge were always greater than the productive counterparts in both immediate and delayed posttests, although the gap between the receptive and productive knowledge in the delayed test was becoming smaller. It was also noteworthy that different task types contributed differently to L2 vocabulary acquisition. Task 4 which involved sentence writing resulted in highest scores in L2 vocabulary learning (receptive: M = 5.00; productive: M = 2.60) and retention (receptive: M = 2.90; productive: M = 1.75), followed by Task 3 which was Chinese to English translation. Concerning the effect of task-induced involvement, it was noted that the descriptive statistics provided contradictory results. In terms of both receptive and productive lexical knowledge, Task 3 with a higher involvement load (index = 4) produced better results than Task 1with a lower involvement load (index = 3), which was consistent with ILH predication. Nonetheless, Task 3 did not produce a higher mean score than Task 4 even though it carried a higher involvement load. This result was contradictory with ILH, which claimed higher involvement load produced better learning. Table 4 showed that if the involvement loads were grouped into low and high levels, the tasks with a lower involvement load resulted in higher gain scores in both initial learning and retention than the tasks with a higher involvement load, lending no support for ILH.

TABLE 3
Descriptive Statistics of Task Type Effect

Task	*n*	Immediate posttest				Delayed posttest			
		Receptive		Productive		Receptive		Productive	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1 (3)	17	3.18	1.59	1.24	1.30	1.82	1.70	1.18	1.51
2 (4)	17	2.35	1.66	1.18	1.24	1.18	1.02	1.00	1.12
3 (4)	20	4.80	2.04	2.35	1.87	2.10	1.71	1.70	1.66
4 (3)	20	5.00	1.97	2.60	1.70	2.90	1.83	1.75	1.33

Note. (3) and (4) mean involvement load indices.

TABLE 4
Descriptive Statistics of Involvement Load Effect

Involvement load	*n*	Immediate posttest				Delayed posttest			
		Receptive		Productive		Receptive		Productive	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Low (index = 3)	37	4.16	2.00	1.97	1.66	2.41	1.83	1.49	1.43
High (index = 4)	37	3.68	2.22	1.81	1.68	1.68	1.49	1.38	1.46

2. Effects of Task Type and Involvement Load

Since L2 vocabulary acquisition was measured by both immediate and delayed receptive and productive vocabulary knowledge, a MANOVA was run to test the effect of task type, the effect of involvement load and their interaction. The results of MANOVA for the immediate and delayed L2 vocabulary acquisition were shown in Table 5. The results indicated that task type ($F(4, 67) = 8.540$, $p < .05$, $\eta^2 = .192$) had a significant effect on L2 vocabulary acquisition and retention. However, there was no significant effect of involvement load ($F(4, 67) = 2.576$, $p > .05$), nor did the interaction between task type and involvement load reach a significant level ($F(4, 67) = .149$, $p > .05$).

TABLE 5
Effects of Task Types and Involvement Load

Effect	*Wilks'Λ*	*df*	*F*	*p*	η^2
Task type	.662	4,67	8.540	.000*	.192
Involvement load	.867	4,67	2.576	.055	
Task × Load	.991	4,67	.152	.963	

*$p < .05$

Based on the above result that task types significantly influenced L2 vocabulary learning, two one-way ANOVAs were conducted to demonstrate the difference of the task types in immediate posttest and delayed posttest (see Table 6 and Table 7). Levene's test of homogeneity of variances showed the significance level was all above .05. The one-way ANOVAs conducted on the immediate posttest revealed a significant task type effect on receptive ($F(3, 70) = 8.836$, $p < .05$, $\eta^2 = .275$) and productive ($F(3, 70) = 84.084$, $p < .05$, $\eta^2 = .149$) vocabulary knowledge. Post hoc Tukey's analysis showed the mean scores of Task 3 and Task 4 were significantly better than the previous two tasks in terms of the immediate receptive knowledge test, whereas Task 4 resulted in significantly higher scores than the other three tasks in the immediate productive test.

TABLE 6
ANOVA for Task Type Effect on Immediate Posttest

	Receptive				Productive			
	MS	*F*(3,70)	*p*	η^2	*MS*	*F*(3,70)	*p*	η^2
Tasks	29.987	8.836	.000*	.275	10.085	4.084	.010*	.149

*$p < .05$

TABLE 7
ANOVA for Task Type Effect on Delayed Posttest

	Receptive				Productive			
	MS	*F*(3,70)	*p*	η^2	*MS*	*F*(3,70)	*p*	η^2
Tasks	9.446	3.622	.017*	.134	1.268	.292	.052	.149

*$p < .05$

In the delayed posttest, the significant task type effect was found only in the receptive knowledge test (see Table 7). The task type, however, produced no difference in the delayed productive vocabulary knowledge test, i.e., all of them played little role in the retention of the productive vocabulary knowledge. In the delayed receptive test, Post hoc Tukey's comparison showed the mean score of Task 4 was significantly higher than the mean score of Task 2 but not significantly higher than Task 1 and Task 3, which implied Task 2 produced the least effect on delayed receptive vocabulary knowledge.

V. DISCUSSION

The focus of the present research was the effectiveness of task type and involvement load on L2 vocabulary receptive and productive knowledge. The statistical results showed that task type had significant effect although different tasks contribute differently to L2 vocabulary learning and retention. However, there was no significant effect of involvement load on L2 learners' vocabulary acquisition. In answering the second research question, the study reported no significant effect concerning the interaction between task type and involvement load, suggesting their independent contributions to L2 vocabulary learning. By separating task type from involvement load, this study rejected ILH. Some previous studies, which regarded task type as an inherent part of involvement load, at least partially support the hypothesis.

Based on the statistical results, it was found that Chinese to English translation task (Task 3) and sentence writing task (Task 4) contributed more to both EFL learners' receptive and productive word knowledge. Task 4 (gloss + sentence-writing), with a lower involvement load index than Task 2 and Task 3, was statistically proved to be a more effective vocabulary learning activity. Task 2 (dictionary look-up + E-C), though it induced a higher involvement load in this study, made the least contribution to vocabulary learning. As a general pattern in both the immediate and delayed posttests, the relative efficacy of all the four tasks could be summarized as: Task 4 (sentence-writing, search−, index = 3) > Task 3 (Chinese-English translation, search+, index = 4) > Task 1 (English-Chinese translation, search−, index = 3) > Task 2 (English-Chinese translation, search+, index = 4). With the value of *need* and *evaluation* being constant across the four tasks, the presence of constructing element *search* in Task 2 did not affect its efficacy in lexical learning and retention.

One plausible explanation for the stronger effect of Task 3 and Task 4 might be that EFL leaners' attention was directed more towards word forms. Both Chinese to English translation and sentence-writing exercises demanded participants' attention not only to meanings of the new words but also their forms. Task 3 and Task 4 can be ascribed to be more output-oriented tasks than English to Chinese translations where the form of the news words was accepted as the receptive information. This was in line with Toth's (2006) research that the more output-oriented tasks might produce greater benefits for learning since those tasks induced more internal language processing engagement for L2 learners. Izumi (2003) also suggested that more output-oriented tasks might facilitate

L2 vocabulary development by forcing EFL learners to notice L2 forms. Nation and Webb (2011) suggested that productive generation, i.e., using a word in a new context, would result in more learning gains. Task 1 and Task 2 in this study were focused on meaning only while Task 3 and Task 4 required the participants to focus on forms, which involved more processing effort to produce the appropriate use of the target word both in meaning and grammar. The two different processes produced different mental effort involved in the tasks. In English to Chinese translation tasks, the EFL learners had to respond by encoding the meanings of the L2 forms. In Chinese to English translation and sentence-writing context, the EFL learners had to respond by matching the L1 with the L2 and encoding the grammatical function of the L2, which were supposed to induce more mental effort and thus resulted in better vocabulary learning. It might be possible to distinguish these two kinds of processes as passive output tasks (Task 1 and Task 2) and active output tasks (Task 3 and Task 4).

It should also be noted that the advantages of active output tasks over passive output tasks had been lost over time. In the delayed productive recall test, no significant difference had been found among the four tasks. As Hulstijn (2001) noted, a decline in knowledge over time was expected since there was no rehearsal or additional exposure to the target words. Folse (2006) also suggested multiple encounters of an unknown word played a facilitative role in lexical acquisition. No extra encounters or any word rehearsals were conducted during the test intervals. Thus, it was not surprising to see the knowledge decline across time. Besides more efficient vocabulary learning tasks, multiple encounters with target words would be added to help the retention of the vocabulary knowledge.

The result of the present research did not provide support for ILH in which Hulstijn and Laufer (2001) stated that higher task-induced involvement load would result in better learning and retention. This claim was not verified in this study. In both the receptive and productive word recall posttests, the more engaging tasks were less effective than the tasks with lower involvement load. Task 1 (index = 3) and Task 4 (index = 3) outperformed Task 2 (index = 4) and Task 3 (index = 4) respectively. The higher involvement load did not necessarily lead to better word recall and retention. The lack of significant difference between the low (index = 3) and high (index = 4) groups in this study was in conflict with lots of previous researches that partially or completely support ILH (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Y. J. Kim, 2008), but in line with some studies which suggested other reasons rather than involvement load for the relative effectiveness of the tasks (e.g., Bao, 2015).

One explanation for the finding might lie in the different importance of the three components (*need, search,* and *evaluation*) in ILH. Although Hulstijn & Laufer (2001) treated every component equally important in their ILH, the result of the present study might suggest that the presence and absence of the *search* component should not be treated as an important contributing factor to involvement load. In fact, the present study suggested a completely different result that the presence of *search* inducing a higher involvement load result in a lower mean scores of vocabulary knowledge tests. Task 4 with the component *search* absent was proved to be a more effective activity than Task2 and Task 3 with the element *search* present. The results were in conflict with J. U. Hyun and J. Y. Kim's (2016) study, in which they reported the presence of *search* (search+) made the same contribution as moderate evaluation (evaluation+) in the initial learning of an unknown words. The one possible explanation might be related to the participants who were the first-year middle school students. Their teachers reported that they were not trained to use dictionaries, which might make them waste time in the dictionary look-up activities. The glosses groups might have more time to process the new words and thus had more chances to remember them. Similarly, Folse (2006) found that when task time kept constant, completing three gap-filling tasks was more effective that one sentence-writing task. The findings indicated that when designing the vocabulary learning tasks, frequency of word encounters was also an important factor that should be considered. As Folse (2006) concluded, multiple retrievals of a newly learnt word in superficial exercises were a more contributing factor in facilitating L2 lexical gains than those tasks inducing a higher involvement load. Bao's (2015) study also supported this conclusion that there were other reasons rather than involvement load for the relative effectiveness of the tasks. Hu and Nassaji (2016) compared ILH with Nation and Webb's (2011) Technique Feature Analysis, indicating that more elements should be included to operationalize the notion of depth of processing.

An alternative explanation might be the numerical values assigned to the components of the involvement load index. Different components might carry different weights in determining the task effectiveness, i.e., more weights might be given to some components than others. While Hulstijn and Laufer (2001) treated every component equally when determining the involvement load index, Keating (2008) and Y. J. Kim (2008) in their empirical studies testing ILH suggested that the ILH had not made fine-grained differentiations of the relative load of the three components. Y. J. Kim (2008) indicated that different degrees (moderate and strong) of the three components might contribute differently to the overall involvement load. In her study, she found that strong evaluation might induce more involvement in processing a word than the moderate evaluation and other two factors. Zou's (2017) research confirmed Y. J. Kim's (2008) finding and suggested that another layer of evaluation 'very strong evaluation' should be added to the framework of task-induced involvement load.

VI. CONCLUSION

This study set out to probe the effects of task type and involvement load on L2 vocabulary acquisition across time. It was found that task type had a significant effect on the acquisition of L2 receptive knowledge across time. In terms of productive vocabulary knowledge, the effect of the task type was significant in the immediate test but not significant in the delayed test. The involvement load had no significant effect on the acquisition of L2 receptive and productive vocabulary knowledge across time. This result lends no support to the ILH, indicating that the involvement load hypothesis did not hold true across L2 vocabulary learning conditions and the components of involvement load might give no equal weight to their contributions to L2 vocabulary acquisition. Finally, there was no significant interaction between task type and involvement load, suggesting the separate effects of task type and involvement load. This separability revealed that task type and involvement load were independent in their contributions to L2 vocabulary learning.

It should also be noted that most studies, including the present one, followed Hulstijn and Laufer's (2001) tradition of using readings as stimuli. Whether those findings could be expanded to other stimuli still remains under investigated. Vidal (2011) used listening as stimuli and tentatively suggested that listening might lead to higher retention of incidentally acquired words for more advanced learners. In the context of digital era, other resources could also be applied in EFL classrooms as stimuli for vocabulary learning tasks. For instance, N. Y. Kim (2018) found video-based vocabulary tasks might be more effective than dictionary-based vocabulary tasks. Changing input mode might involve a different cognitive process. For future research, the effect of task type and involvement load should be expanded to other stimuli rather than reading only.

These findings have pedagogical implications. EFL teachers should be aware that not all L2 vocabulary learning tasks serve the same purpose. When language teachers design effective tasks, learners' proficiency level should be considered. For learners with low L2 proficiency levels, dictionary look-up activities might not as efficient as providing glosses directly to the students. The adolescent EFL learners might be not skilled at learning a new word with the help of dictionaries. Nonetheless, for more advanced learners, the situation might be vice versa, since the dictionary look-up activities involve heavier mental load. Translating L1 to L2 is more effective than translation from L2 to L1 in facilitating both active and passive word recall. Sentence-writing might be the most effective vocabulary learning tasks among the four since it entails more mental processing effort in terms of the meaning-form mapping and contextual clueing.

REFERENCES

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System, 53*, 84-95.

Bruton, A. (2007). Vocabulary learning from dictionary referencing and language feedback in EFL translational writing. *Language Teaching Research, 11*(4), 413-431.

Chinese Ministry of Education. (2011). *National English curriculum for compulsary education* (2011 version). Beijing: Beijing Normal University Press.

Cho, Kyung-Sook, & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading, 37*(8), 662-667.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rational for pedagogy* (pp. 225-237). Cambridge: Cambridge University Press.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*(2), 227-252.

Ellis, N. C. (1994). V*ocabulary acquisition: The implicit ins and outs of explicit cognitive mediation.* London: Academic Press.

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

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics, 41*(2), 87-106.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign language, 11*(2), 207-223.

Hu, H. C. M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *Canadian Modern Language Review, 68*(1), 54-77.

Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus

technique feature analysis. *System, 56*, 28-39.
Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal, 96*(4), 544-557.
Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Palgrave Macmillan.
Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.
Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.
Hyun, Jooe-Un, & Kim, Jie-Young. (2016). Investigating the task involvement load hypothesis in EFL learners' vocabulary learning. *Journal of the Korea English Education Society, 15*(1), 1-22.
Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*(2), 168-196.
Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research, 11*(2), 149-158.
Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377.
Kang, Min-Hwi. (2015). Vocabulary exercises with different involvement load after reading in a university EFL reading class. *Korean Journal of Applied Linguistics, 31*(4), 31-54.
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, Hyung-Sun, & Na, Yoon-Hee. (2010). Vocabulary learning and task-induced involvement. *Korean Journal of Applied Linguistics, 26*(4), 183-211.
Kim, Na-Young. (2018). A study on mobile-assisted vocabulary learning with dictionary-based, corpus-based, and video-based tasks. *Modern English Education, 19*(4), 27-38.
Kim, You-Jin. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285-325.
Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal, 78*(3), 285-299.
Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected papers from AILA'99 Tokyo*, 47-62.
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. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook, 5*(1), 223-250.
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. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.
Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurance and their combination. *Language Teaching Research, 15*(4), 391-411.
Min, H. T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning, 58*(1), 73-115.
Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.
Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly, 37*(4), 645-670.
Nation, I. S. P. (2001). *Learning vocabulary in another language.* Stuttgart: Ernst Klett Sprachen.
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review, 63*(1), 59-82.
Nation, I. S. P., & Beglar, D. (2007). *A vocabulary size test. The Language Teacher, 31*(7), 9-13.
Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle, Cengage Learning.
Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29.
Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174-200). Cambridge: Cambridge University Press.
Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orage study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271-275.
Pulido, D. (2009). How involved are American L2 learners

of Spanish in lexical input processing tasks during reading? *Studies in Second Language Acquisition, 31*(1), 31-58.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning, 56*(2), 319-385.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

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.

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

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.

APPENDIX 1

Vocabulary Size Test

Directions: Please write down the Chinese meaning of the following words.

1. see: They saw it.
2. time: They have a lot of time.
3. period: It was a difficult period.
4. figure: Is this the right figure?
5. poor: We are poor.
6. drive: He drives fast.
7. jump: She tried to jump.
8. shoe: Where is your shoe?
9. standard: Her standards are very high.
10. basis: This was used as the basis.
11. maintain: Can they maintain it?
12. stone: They sat a stone.
13. upset: I am upset.
14. drawer: The drawer was empty.
15. patience: He has no patience.
16. nil: His mark for that question was nil.
17.pub: They went to the pub.
18. circle: Make a circle.
19: microphone: Please use the microphone.
20. pro: He's a pro.

APPENDIX 2

The Reading Passage

Last Monday John went to the downtown. After John spent a long time shopping in the downtown, he got very hungry. Across from the street there was a McDonald's. He didn't usually eat at fast food restaurants and didn't know what to order. When he looked at the menu, it was hard for him to make a decision. Not only there were ten similar types of hamburger dinners, but there were also fish and chicken. He didn't know what to choose. He wanted to ask the cashier for help, but she didn't look very friendly. At last he chose a small chicken meal. "With so many choices, it's not very fast food!" he thought.

APPENDIX 3

Examples of the Tasks

English to Chinese Translation (Task 1 and Task2)
e.g., The teacher chooses this book.
Chines to English Translation (Task 3)
e.g., 老师选了这本书。(choose)
Sentence Writing (Task 4)
e.g., choose:_____.

APPENDIX 4

An Example of the Vocabulary Knowledge Scale

cashier
(1) 我不记得见过这个词。(　　)
(2) 我见过这个词，但不知道它的意思。(　　)
(3) 我知道这个词。它的意思是_____(写出汉语释义)。
(4) 我会用这个词造英文句子。如：_____.