

# The Involvement Load Hypothesis or the Technical Feature Analysis: Evidence from Four EFL Vocabulary Learning Tasks

Mengjuan Qin

College of Continuing Education, Nanjing Media College, Nanjing, 210000, China

---

**Abstract:** L2 vocabulary acquisition is of much importance to L2 learners and has long been the hot topic in the field of applied linguistics. Stimulated by the Technical Feature Analysis (TFA) and the Involvement Load Hypothesis (ILH) in particular, a large body of research has attempted to find the relative efficacy of various tasks in facilitating L2 vocabulary acquisition. However, there is a paucity of research regarding the relative effectiveness of the ILH and the TFA in accounting for the task type effects on L2 vocabulary acquisition. Therefore, the present study attempts to compare the relative effectiveness of the ILH and the TFA. The two main research questions are stated as follows: 1. How does task type (reading comprehension with glosses, gap filling, rewording and composition) affect EFL vocabulary acquisition? 2. Which better explains the relative effectiveness of tasks (reading comprehension with glosses, gap filling, rewording and composition) in facilitating EFL vocabulary acquisition, the Involvement Load Hypothesis (ILH) or the Technical Feature Analysis (TFA)? This study employed one-way ANCOVA (analysis of covariance) design with the between-subject factor being task type and the covariate being EFL vocabulary knowledge. The dependent variable was EFL vocabulary acquisition, which included two dimensions, i.e., word form recognition and passive word meaning recall. The EFL vocabulary knowledge posttests included the immediate and delayed posttests. After completing the EFL vocabulary knowledge (vocabulary size) test, 117 non-English major freshmen were randomly assigned to one of four learning tasks, i.e., reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA), gap filling (an index of 2 in the ILH versus an index of 7 in the TFA), rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA). The total task time required for the different groups was held constant. Then, on completion of the respective tasks, the participants were required to take the immediate posttests of word form recognition and passive meaning recall. One week after the immediate posttest, all the four groups took the delayed posttests. All the data collected were valid.

**Keywords:** Task type, The Involvement Load Hypothesis, The Technical Feature Analysis, L2 vocabulary acquisition.

---

## 1. Introduction

A large amount of research on Second Language Acquisition (SLA) was based on the Involvement Load Hypothesis (ILH) which investigated the effects of different tasks on EFL vocabulary acquisition and confirmed the effectiveness of the ILH in accounting for task type effects on EFL vocabulary acquisition (Bao 2015; Hu & Nassaji 2016; Laufer & Hulstijn 2001; Wu 2010; Miao 2014). However, there were still some researchers who saw the limitations of the effectiveness of the ILH in accounting for task type effects on EFL vocabulary acquisition (Folse 2006; Hu & Nassaji 2012; Kong & Wang 2014; Yu 2015). On the basis of the above findings, some researchers found that, the task-induced involvement load was not necessarily the only factor determining the effectiveness of the task. This culminated into further research on other factors such as vocabulary acquisition strategies, context cues (Gu & Song 2010; Bao 2015) and so on. In order to find a more comprehensive model accounting for the relative effectiveness of task types on EFL vocabulary acquisition, a paucity of scholars researched the Technical Feature Analysis (TFA) so as to compare the effectiveness of the ILH and the TFA, both qualitatively and quantitatively. (Nation & Web 2011; Hu & Nassaji 2016; Gohar et al. 2018).

The present study was set up to examine the effects of four different tasks on vocabulary acquisition and to compare the relative effectiveness of the ILH and the TFA in accounting

for the task type effects on EFL vocabulary acquisition. More specifically, its aim is to determine which of the two frameworks could better explain the relative effectiveness of tasks in facilitating EFL vocabulary acquisition. Four vocabulary learning tasks that differ in their indexes in the ILH and the TFA were chosen, including reading comprehension with glosses, gap filling, rewording and composition.

## 2. Methodology

### 2.1. Research Questions

This research studies the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition. The following questions are to be answered in this research:

1. How does task type affect EFL vocabulary acquisition?
  - 1.1 How does task type affect EFL vocabulary form recognition and meaning recall in the immediate posttest?
  - 1.2 How does task type affect EFL vocabulary form recognition and meaning recall in the delayed posttest?
2. Which better explains the relative effectiveness of tasks in facilitating EFL vocabulary acquisition, the ILH or the TFA?
  - 2.1 Which better explains the relative effectiveness of tasks in facilitating EFL vocabulary form recognition and meaning recall in the immediate posttest, the ILH or the TFA?
  - 2.2 Which better explains the relative effectiveness of tasks in facilitating EFL vocabulary form recognition and meaning

recall in the delayed posttest, the ILH or the TFA?

## 2.2. Experimental Design

This study employed one-way ANCOVA (analysis of covariance) design with the between-subject factor being task type and the covariate being EFL vocabulary knowledge. This study included four reading tasks, reading comprehension with glosses, gap filling, rewording and composition. The EFL vocabulary knowledge was operationalized by a vocabulary size test. The dependent variable was EFL vocabulary acquisition, which included two dimensions, i.e., word form recognition and passive word meaning recall. The EFL vocabulary knowledge posttests included the immediate and delayed posttests.

## 2.3. Participants

In this study, 136 freshmen were selected to participate in the experiment. They are non-English majors from Nanjing Xiao Zhuang University. They all have Chinese as their mother tongue and have learned English for a long time. The average of their ages is 19 years, ranging from 17 to 21 years old.

The students were randomly assigned to one of four reading tasks, i.e., reading comprehension with glosses, gap filling, rewording and composition after completing a vocabulary size test. They were then asked to finish the word form recognition and passive meaning recall exercises. One week later, they were tested again on the word form recognition and passive word meaning recall. 19 students were absent from the experiment due to reasons beyond the researcher's control. Consequently, 117 participants' scores were valid and analyzed. Specifically, 32 participants were in the group of reading comprehension with glosses, 29 participants in the group of gap filling, 27 participants in the group of rewording and 29 participants in the group of composition.

## 2.4. Materials

### 2.4.1. Vocabulary Size Test

The vocabulary size test was of word meaning recognition, and the test items were selected from the vocabulary level test developed by Schmitt et al. (2001). 14 target words and 2 distractors were both added to this test to find whether or not these words were unknown to the participants. The number of test items was 54 and the total score of the test was 162. For each test item, there were 6 words on the left and 3 meanings on the right. The participants were required to match the words on the left with the meanings on the right. The revised vocabulary size test was attached to this study and here is an example. In this example, 'tip' means 'end or highest point', 'motor' means 'this moves a car' and 'copy' means 'thing made to be like another'.

### 2.4.2. EFL Reading Tasks and Target Words

The reading text of this study, 294 words long and entitled A Costume Party, was adapted from Nist & Mohr (2002). The participants are familiar with the subject of this text, whose language is considerably understandable for participants.

Target words were chosen based on the pilot study and were operationalized as unknown words to all the participants. The pilot study was carried out among 20 first-year postgraduate students majoring in applied linguistics. It was assumed that the vocabulary proficiency of postgraduate students majoring in applied linguistic was higher than that of the participants who were freshmen of non-English major. A

conclusion could be made that the participants were unfamiliar with these words considering the fact that these words were unfamiliar to the postgraduate students. Also, the researcher used a pilot study as an opportunity to solicit advice relating to the research material, the time allocation and so on. There were 14 target words plus 2 distractors. The target words included 5 nouns (discretion, travesty, facsimile, metamorphosis, retrospect), 5 adjectives (esoteric, grotesque, notorious, provocative, perfunctory) and 4 verbs (circumvent, embellish, mesmerize, ve]nerate). The distractors were 2 nouns (assassination, mockery). The reason for adding two distractors was to distract the attention of the participants to the target words. The roughly equal number of nouns, verbs and adjectives was chosen so that the effects of part of speech on vocabulary learning was controlled. All the words appeared in the text only once and were glossed (i.e., part of speech and Chinese meaning). All the related information was cited from the Longman Dictionary of Contemporary English (2004). For example, "venerate" was glossed as follows: venerate: vt.; Past tense and past participle: -d: respect. For more information.

### 2.4.3. Task Types

There were four tasks in this research, including reading comprehension with glosses, gap filling, rewording and writing. The participants of reading comprehension with glosses were asked to finish the related questions and write a short composition within 100 words based on the content of text instead of the target words after reading the text with glosses. To begin, the participants had to answer four questions based on the content of the text instead of the target words after reading the short passage. In order to control the task time, these participants were then asked to write a short composition within 100 words about their own experiences relating to the passage. The glosses in the reading comprehension included the part of speech and the Chinese meaning of the target words. For example: circumvent: vt.; Past tense and past participle: -ed: avoid.

The participants of gap filling were asked to use the correct form of given words to fill in the blanks of the text. For this task, 14 target words and 2 distractors were taken out from the passage. They were all glossed and the participants had to use the correct form of 16 words to fill in the blanks. For example:

On the afternoon of a friend's New Year's Eve costume party, I made only a \_\_\_\_\_ effort to put a costume together.

For this example, the answer should be "perfunctory".

The participants of rewording were asked to paraphrase the sentences using the correct form of the given words in the brackets. For this task, 16 sentences without 14 target words and 2 distractors were shown after the passage. The participants were asked to use the words in the bracket to paraphrase the given sentence. For example:

1. copy	end or highest point	<u>(6)</u>
2. event	this moves a car	<u>(3)</u>
3. motor	thing made to be like another	<u>(1)</u>
4. pity		
5. profit		
6. tip		

The doctor's examination was not careful. (perfunctory)

For this example, the answer should be "The doctor's examination was perfunctory".

The participants of composition were asked to write a

composition. While writing, they were given the clues. Some clues or plots were extracted from the passage and the participants were asked to base their writing on these clues with the target words.

Nation and Web (2011) gave the indexes of the ILH and the TFA in different activities. The four reading tasks in the current research were chosen from these activities, including reading comprehension with glosses, gap filling, rewording and composition. For these four tasks, their involvement load and technical feature indexes are shown in Table 1.

**Table 1.** Comparison of the ILH and TFA on four activities

Activity	The ILH index	The TFA index
Reading comprehension with glosses	1	5
Gap filling	2	7
Rewording	3	6
Composition	3	8

As Table 1 shows, reading comprehension with glosses has a lowest index in both the ILH and the TFA when compared to composition which conversely has a highest index in both the ILH and the TFA. Based on the comparison between gap filling and rewording, the former task has a lower index in the ILH but a higher index in the TFA with the latter task being opposite, making the comparison of the relative effectiveness of the ILH and TFA more vivid and simple. Additionally, the tasks of rewording and composition have the same index in the ILH and different indexes in the TFA, also making the comparison of the relative effectiveness of the ILH and TFA more vivid and simple.

#### 2.4.4. EFL Vocabulary Knowledge Posttests

The vocabulary knowledge posttests included the immediate posttest and the delayed posttest. The immediate posttest and the delayed posttest were both composed of two parts, i.e., word form recognition and word meaning recall. About the word form recognition in the immediate posttest, the participants had to choose the correct spelling of all the 16 words after completing the reading task. For example:

- ( ) A. metermorphosis    B. metormorphosis  
C. mitamorphosis      D. metamorphosis

With regard to the word meaning recall in the immediate posttest, the participants had to recall the correct Chinese meaning of all the 16 words.

The immediate vocabulary knowledge posttests were attached. The delayed posttest was similar to the immediate posttest and merely differed in the sequence of test items. The sequence of exercises in the delayed posttest was obtained from randomly disturbing the sequence of test items of the immediate posttest.

## 2.5. Procedures

### 2.5.1. Pilot Study

The reasons why this pilot study was conducted were obvious. On one hand, some modifications of the instructions were made. For example, one suggestion was given regarding the instruction of the rewording task that it wasn't clear enough due to the fact that there was no example and that the instructions would be much clearer for participants by adding one example. On the other hand, it was helpful to find out whether or not the time allotted for each step was suitable. Another advice was given concerning the task of composition that it may require more time. Based on the pilot study, the time for each reading task was 15 minutes and the time for

word form recognition was both 10 minutes in the immediate posttest and the delayed posttest. The time for word meaning recall was also 10 minutes in the immediate posttest and the delayed posttest.

### 2.5.2. Experimental Procedures

The experimental procedures consisted of three parts, i.e., four tasks, the immediate posttest and the delayed posttest. All the procedures were conducted in the class. The participants are of different majors and were in three different classrooms, with one instructor in each classroom. All the instructors had been trained to be familiar with the details of the experiment, including the procedures, instructions and the time allotted for each procedure.

The first procedure was about four tasks. The participants were randomly assigned to finish four tasks within 15 minutes, including reading comprehension with glosses, gap filling, rewording and composition. All the task papers were collected afterwards.

The second procedure was about the immediate posttest. All the participants were first of all required to finish the immediate posttest of word form recognition within 10 minutes. When the time was up, all the test papers were collected. After that, the participants were required to finish the immediate posttest of word meaning recall within 10 minutes.

The third procedure was about the delayed posttest. One week after the immediate posttest, the participants were first of all required to finish the delayed posttest of word form recognition within 10 minutes. When the time was up, all the test papers were collected. After that, they were required to finish the delayed posttest of word meaning recall within 10 minutes.

## 2.6. Scoring and Data Analysis

Vocabulary knowledge posttests consisted of word form recognition and word meaning recall. For word form recognition, 1 point was given for a correct choice and 0 point for an incorrect choice. As for word meaning recall, 1 point was given for a correct answer, 0.5 point for a partially correct answer. For example, 0.5 point was given for translating 'discretion' as 'discriminative' since 'discretion' is a noun, not an adjective. Two postgraduate students were trained to score and check data. This is due to the fact that there may have been some errors and the scoring of word meaning could be subjective to certain degrees. They scored these tests independently based on the above-mentioned standards. All the divergence was discussed by them to reach a consensus when they were scoring the tests. As a result, the inter-rate agreement of the two tests was 100%.

All participants' scores on each test were submitted to SPSS 25.0 for data analysis. The analysis of covariance (ANCOVA) was conducted to address each of the research questions.

## 3. Results and Discussions

### 3.1. Task Type Effects on EFL Vocabulary Knowledge on the Immediate posttest

An analysis of research data gathered in the immediate word knowledge posttest is presented in this section, and the research questions posed in Chapter 3 are reiterated and addressed. Foremost, this section reports the descriptive statistics of EFL vocabulary knowledge across task type on the immediate word knowledge posttest. In addition, it reports

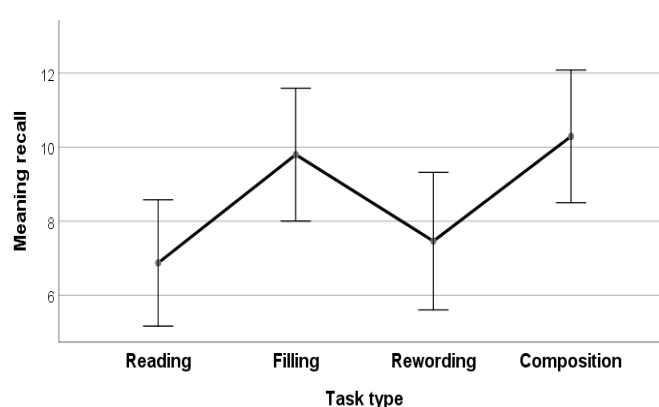
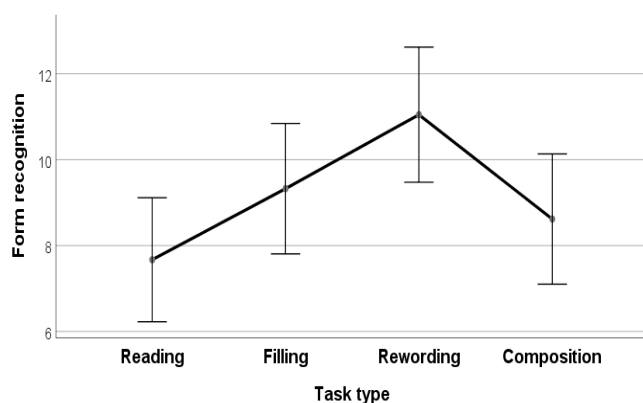
an analysis of covariance (ANCOVA) of task type effects on EFL vocabulary knowledge on the immediate posttest. The analysis of covariance was used to test the relationship between task type and vocabulary acquisition in terms of two word knowledge aspects, i.e., word form recognition and passive word meaning recall. Previous vocabulary knowledge was used as the covariate. The purpose of using the covariate is to partial out the influence of one or more variables on word learning before conducting the analysis of interest. A covariate could best be defined as a variable that has a substantial correlation with the dependent variable and is included in the experiment to adjust the results for differences existing among subjects before the start of the experiment. Task comparisons effect sizes were judged by Cohen's criteria (1988):  $\hat{\eta}_p^2 = 0.01$ , a small effect size;  $\hat{\eta}_p^2 = 0.06$ , a medium effect size;  $\hat{\eta}_p^2 = 0.14$ , a large effect size.

### 3.1.1. Descriptive statistics of EFL vocabulary knowledge across task type on the immediate posttest

Table 2 reports the descriptive statistics of EFL vocabulary knowledge across task type on the immediate posttest, covering the sample size, means, standard deviations and the 95% confidence intervals for word form recognition and passive word meaning recall on the immediate posttest.

**Table 2.** Descriptive statistics of EFL vocabulary knowledge across task type on the immediate posttest

	Task	<i>M</i>	<i>SD</i>	<i>N</i>	95% CI
<b>Form</b>	Reading	7.72	4.04	32	[6.23, 9.11]
	Filling	9.21	5.03	29	[7.81, 10.84]
	Rewording	11.15	3.56	27	[9.47, 12.62]
	Composition	8.59	4.21	29	[7.10, 10.13]
<b>Meaning</b>	Reading	6.95	5.16	32	[5.17, 8.58]
	Filling	9.60	5.13	29	[8.01, 11.59]
	Rewording	7.63	5.73	27	[5.60, 9.32]
	Composition	10.24	4.71	29	[8.50, 12.08]



**Figure 1.** A comparison of adjusted means of EFL vocabulary knowledge across task type on the immediate posttest

In Figure 1, 'Reading' is short for reading comprehension with glosses and 'Filling' is short for gap filling. Based on Figure 1, in the immediate word form recognition posttest, there is no overlap between reading comprehension with glosses and rewording, which illustrates that there is an apparent difference between the means of two groups, suggesting that there might be a significant difference between reading comprehension with glosses and rewording in the immediate word form recognition posttest. However, there is much overlap of 95% CIs between reading comprehension with glosses and gap filling, between reading

In Table 2, 'Reading' is short for reading comprehension with glosses and 'Filling' is short for gap filling. As Table 2 shows, in the immediate word form recognition posttest, rewording ( $M = 11.15$ , 95%CI [9.47, 12.62]) performed better than reading comprehension with glosses ( $M = 7.72$ , 95%CI [6.23, 9.11]), gap filling ( $M = 9.21$ , 95%CI [7.81, 10.84]) and composition ( $M = 8.59$ , 95%CI [7.10, 10.13]). The mean difference between rewording and reading comprehension with glosses is 3.43, the mean difference between rewording and gap filling is 1.94 and the mean difference between rewording and composition is 2.56. Conspicuously, rewording outperformed other groups in the immediate word form recognition posttest.

In the immediate word meaning recall posttest, composition ( $M = 10.24$ , 95%CI [8.50, 12.08]) performed better than reading comprehension with glosses ( $M = 6.95$ , 95%CI [5.17, 8.58]), gap filling ( $M = 9.60$ , 95%CI [8.01, 11.59]) and rewording ( $M = 7.63$ , 95%CI [5.60, 9.32]). The mean difference between composition and reading comprehension with glosses is 3.29. The mean difference between composition and gap filling is 0.64 and the mean difference between composition and rewording is 2.61. Conspicuously, composition outperformed other groups in the immediate word meaning recall posttest.

On the basis of the comparison of two immediate posttests, it could sum up that the performance of reading comprehension with glosses was the worst among all the groups in the immediate posttests. However, the comparisons of other groups in the immediate word form recognition posttest were totally different from those in the immediate word meaning recall posttest.

A comparison of adjusted means of EFL vocabulary knowledge across task type on the immediate posttest is shown in Figure 1.

comprehension with glosses and composition, between gap filling and rewording, between gap filling and composition and between rewording and composition, which illustrates that there is no apparent difference between the means of two groups, suggesting that there might be no significant difference between the two groups in the immediate word form recognition posttest.

Based on Figure 1, in the immediate word meaning recall posttest, there is little overlap of 95% CIs between reading comprehension with glosses and composition, which illustrates that there is an apparent difference between the

means of two groups, suggesting that there might be a significant difference between reading comprehension with glosses and composition in the immediate word meaning recall posttest. However, there is much overlap between reading comprehension with glosses and gap filling, between reading comprehension with glosses and rewording, between gap filling and rewording, between gap filling and composition and between rewording and composition, which illustrates that there is no apparent difference between the means of two groups, suggesting that there might be no significant difference between the two groups in the immediate word meaning recall posttest.

### 3.1.2. An Analysis of Covariance of Task Type Effects on EFL Vocabulary on the immediate posttest

The analysis of covariance (ANCOVA) was used to test the relationship between task type and vocabulary acquisition. According to the results, there was no interaction between task type (independent variable) and vocabulary knowledge (covariate), which showed the homogeneity of regression slope ( $p = .301 > .05$ ). Levene's test indicated the homogeneity of variance on the immediate word form recognition ( $F(3, 113) = 2.020, p = .115 > .05$ ). Pearson correlation analysis showed that there was a linear relationship between previous vocabulary knowledge and immediate word form recognition ( $p = .005 < .05, r = .26$ ). An ANCOVA of EFL form recognition across task type on the

immediate posttest revealed that task type had a significant main effect on the immediate word form recognition ( $p = .019 < .05$ ) with a moderate effect size ( $\eta_p^2 = .085 > .06$ ). It also indicated that vocabulary knowledge had a significant main effect on the immediate word form recognition ( $p = .005 < .05$ ) with a moderate effect size ( $\eta_p^2 = .069 > .06$ ).

Moreover, Levene's test indicated the homogeneity of variance on the immediate word meaning recall ( $F(3, 113) = 0.897, p = .445 > .05$ ). Pearson correlation analysis showed that there was a linear relationship between previous vocabulary knowledge and immediate word meaning recall ( $p = .000 < .001, r = .33$ ). An ANCOVA of EFL meaning recall across task type on the immediate posttest revealed that task type had a significant main effect on the immediate word meaning recall ( $p = .016 < .05$ ) with a moderate effect size ( $\eta_p^2 = .087 > .06$ ). It also indicated that vocabulary knowledge had a significant main effect on the immediate word meaning recall ( $p < .001$ ) with a moderate effect size ( $\eta_p^2 = .126 > .06$ ).

The second research question examined to what extent the vocabulary tasks with similar and different indexes in the ILH and the TFA contributed to vocabulary learning. To explore this question, task comparisons in EFL vocabulary knowledge on the immediate posttest were used to compare the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition.

**Table 3.** Task comparisons in EFL vocabulary knowledge on the immediate posttest

Task	Mean difference	SE	p	95% CI
Reading Filling	-1.66	1.06	.724	[-4.50, 1.19]
Rewording	-3.38	1.08	.013	[-6.27, -0.48]
<b>Form</b> Composition	-.95	1.06	1.000	[-3.79, 1.89]
Fulfilling Rewording	-1.72	1.11	.731	[-4.69, 1.25]
Composition	0.71	1.08	1.000	[-2.20, 3.62]
Rewording Composition	2.43	1.10	.178	[-0.53, 5.39]
Reading Filling	-2.93	1.250	.126	[-6.28, 0.43]
<b>Meaning</b> Rewording	-.59	1.272	1.000	[-4.01, 2.83]
Composition	-3.42	1.248	.043	[-6.77, -0.07]
Fulfilling Rewording	2.34	1.305	.457	[-1.17, 5.84]
Composition	-.49	1.279	1.000	[-3.93, 2.94]
Rewording Composition	-2.83	1.303	.192	[-6.33, 0.67]

In Table 3, 'Reading' is short for reading comprehension with glosses and 'Filling' is short for gap filling. According to Table 3, there were significant differences between rewording and reading comprehension with glosses in the immediate word form recognition posttest ( $p = .013 < .05$ ) and between composition and reading comprehension with glosses in the immediate word meaning recall posttest ( $p = .043 < .05$ ). Based on the mean differences of the groups, rewording (an index of 3 in the ILH versus an index of 6 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word form recognition posttest and composition (an index of 3 in the ILH versus an index of 8 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word meaning recall posttest. Both the ILH and the TFA predicts that the tasks with higher indexes would contribute to better L2 vocabulary learning. The indexes of rewording and composition in the ILH and the TFA are both higher than those of reading comprehension with glosses. According to the predictions of both the ILH and the TFA, rewording and composition should contribute to better L2 vocabulary learning than reading comprehension with glosses,

which is consistent with the statistical results above. As a result, these findings lent support to both the ILH and the TFA. Moreover, no significant difference in immediate word form recognition or word meaning recall between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) lent support to the ILH but not to the TFA, since the ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in immediate word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

In summary, it could sum up that both the ILH and the TFA could partly explain the task effectiveness in facilitating EFL vocabulary acquisition in the immediate posttest. However, compared to the TFA, the ILH could better explain the task effectiveness in facilitating EFL vocabulary acquisition in the immediate posttest.

### 3.2. Task Type Effects on EFL Vocabulary Knowledge on the Delayed posttest

#### 3.2.1. Descriptive Statistics of EFL vocabulary Knowledge across Task Type on the Delayed Posttest

Table 4 reports the descriptive statistics of EFL vocabulary knowledge across task type on the delayed posttest, covering the sample size, means, standard deviations and the 95% confidence intervals for word form recognition and passive word meaning recall on the delayed posttest.

**Table 4.** Descriptive statistics of EFL vocabulary knowledge across task type on the delayed posttest

Task	<i>M</i>	<i>SD</i>	<i>N</i>	95% CI
Reading	6.50	3.61	32	[5.08, 7.85]
Filling	7.45	4.39	29	[6.08, 8.99]
<b>Form</b> Rewording	6.67	4.24	27	[5.08, 8.10]
Composition	8.38	3.88	29	[6.95, 9.86]
Reading	2.53	3.56	32	[0.78, 4.15]
Filling	5.35	5.64	29	[3.74, 7.28]
<b>Meaning</b> Rewording	4.39	4.81	27	[2.41, 6.08]
Composition	5.14	5.90	29	[3.41, 6.95]

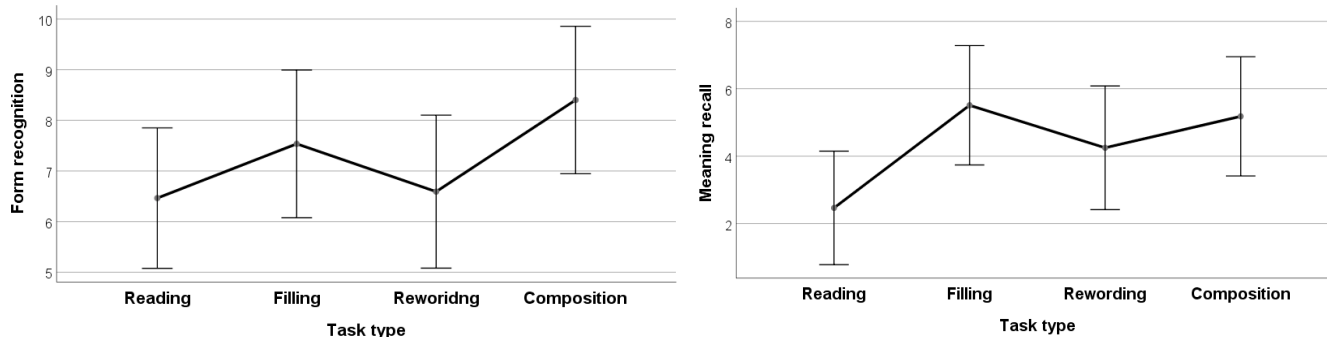
In Table 4, ‘Reading’ is short for reading comprehension with glosses and ‘Filling’ is short for gap filling. As Table 4.3 shows, in the delayed word form recognition posttest, composition ( $M = 8.38$ , 95%CI [6.95, 9.86]) performed better than reading comprehension with glosses ( $M = 6.50$ , 95%CI

[5.08, 7.85], gap filling ( $M = 7.45$ , 95%CI [6.08, 8.99]) and rewording group ( $M = 6.67$ , 95%CI [5.08, 8.10]). The mean difference between composition and reading comprehension with glosses is 1.88. The mean difference between composition and gap filling is 0.93 and the mean difference between composition and rewording is 1.71. Conspicuously, composition outperformed other groups in the delayed word form recognition posttest.

In the delayed word meaning recall posttest, gap filling ( $M = 5.35$ , 95%CI [3.74, 7.28]) performed better than reading comprehension with glosses ( $M = 2.53$ , 95%CI [0.78, 4.15]), rewording ( $M = 4.39$ , 95%CI [2.41, 6.08]) and composition ( $M = 5.14$ , 95%CI [3.41, 6.95]). The mean difference between gap filling and reading comprehension with glosses is 2.82. The mean difference between gap filling and rewording is 0.96 and the mean difference between gap filling and composition is 0.21. Conspicuously, the gap filling task outperformed other tasks in the delayed word meaning recall posttest.

On the basis of the comparison of two delayed posttests, it could be found that the performance of reading comprehension with glosses was the worst among all the groups in the delayed posttests. Composition and gap filling performed better than rewording in the delayed posttests. However, the comparison between composition and gap filling in the delayed word form recognition posttest was different from that in the delayed word meaning recall posttest.

A comparison of adjusted means of EFL vocabulary knowledge across task type on the delayed posttest is shown in Figure 2.



**Figure 2.** A comparison of adjusted means of EFL vocabulary knowledge across task type on the delayed posttest

In Figure 2, ‘Reading’ is short for reading comprehension with glosses and ‘Filling’ is short for gap filling. Based on Figure 2, in the delayed word form recognition posttest, there is much overlap of 95% CIs between reading comprehension with glosses and gap filling, between reading comprehension with glosses and rewording, between reading comprehension with glosses and composition, between gap filling and rewording, between gap filling and composition and between rewording and composition, which illustrates that there is no apparent difference between the means of two groups, suggesting that there might be no significant difference between the two groups in the delayed word form recognition posttest.

Based on Figure 2, in the delayed word meaning recall posttest, there is little overlap of 95% CIs between reading comprehension with glosses and gap filling, which illustrates that there is an apparent difference between the means of two

groups, suggesting that there might be a significant difference between reading comprehension with glosses and gap filling in the delayed word meaning recall posttest. However, there is much overlap between reading comprehension with glosses and rewording, between reading comprehension with glosses and composition, between gap filling and rewording, between gap filling and composition and between rewording and composition, which illustrates that there is no apparent difference between the means of two groups, suggesting that there might be no significant difference between the two groups in the delayed word meaning recall posttest.

#### 3.2.2. An Analysis of Covariance of Task Type Effects on EFL Vocabulary Knowledge on the Delayed Posttest

The analysis of covariance (ANCOVA) was used to test the relationship between task type and vocabulary acquisition. According to the results, there was no interaction between task type (independent variable) and vocabulary knowledge

(covariate), which showed the homogeneity of regression slope ( $p = .301 > .05$ ). Levene's test indicated the homogeneity of variance on the delayed word form recognition ( $F(3,113) = 1.142, p = .335 > .05$ ). Pearson correlation analysis showed that there was a linear relationship between previous vocabulary knowledge and delayed word form recognition ( $p = .038 < .05, r = .19$ ). An ANCOVA of EFL form recognition across task type on the delayed posttest revealed that task type had no significant main effect on the delayed word form recognition ( $p = .210 > .05$ ) with a small effect size ( $\eta_p^2 = .039 < .06$ ). It also indicated that vocabulary knowledge had a significant main effect on the delayed word form recognition ( $p = .029 < .05$ ) with a small effect size ( $\eta_p^2 = .042 < .06$ ).

Moreover, Levene's test didn't indicate the homogeneity of variance on the delayed word meaning recall ( $F(3,113) = 6.037, p = .001 < .05$ ). Pearson correlation analysis showed that there was a linear relationship between previous

vocabulary knowledge and delayed word meaning recall ( $p = .002 < .05, r = .29$ ). An ANCOVA of EFL meaning recall across task type on the delayed posttest revealed that task type had no significant main effect on the delayed word meaning recall ( $p = .064 > .05$ ) with a moderate effect size ( $\eta_p^2 = .062 > .06$ ). It also indicated that vocabulary knowledge had a significant main effect on the delayed word meaning recall ( $p = .001 < .05$ ) with a moderate effect size ( $\eta_p^2 = .095 > .06$ ). Due to the robustness of ANCOVA (Bao 2015), ANCOVA was still employed here to answer the research questions.

The second research question examined to what extent the vocabulary tasks with similar and different indexes in the ILH and the TFA contributed to vocabulary learning. To explore this question, task comparisons in EFL vocabulary knowledge on the delayed posttest were used to compare the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition.

**Table 5.** Task Comparisons in EFL Vocabulary Knowledge on the Delayed Posttest

Task	Mean difference	SE	p	95% CI
Reading Filling	-1.07	1.02	1.00	[-3.80, 1.66]
Rewording	-.13	1.04	1.00	[-2.91, 2.65]
<b>Form</b> Composition	-1.94	1.02	.353	[-4.67, 0.79]
Fulfilling Rewording	.94	1.06	1.00	[-1.91, 3.80]
Composition	-.87	1.04	1.00	[-3.66, 1.93]
Rewording Composition	-1.81	1.06	.542	[-4.66, 1.04]
Reading Filling	-3.05	1.23	.091	[-6.36, 0.27]
<b>Meaning</b> Rewording	-1.78	1.26	.950	[-5.16, 1.59]
Composition	-2.72	1.23	.178	[-6.03, 0.59]
Fulfilling Rewording	1.26	1.29	1.00	[-2.20, 4.72]
Composition	.33	1.26	1.00	[-3.06, 3.72]
Rewording Composition	-.93	1.29	1.00	[-4.39, 2.52]

In Table 5, 'Reading' is short for reading comprehension with glosses and 'Filling' is short for gap filling. According to Table 5, there were no significant differences among the task comparisons. Specifically, no significant difference between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) in delayed word form recognition and word meaning recall lent support to the ILH but not to the TFA, since the ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in delayed word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

In summary, it could sum up that both the ILH and the TFA could not truly explain task effectiveness in facilitating EFL vocabulary acquisition in the delayed posttest. However, compared to the TFA, the ILH could better explain task effectiveness in facilitating EFL vocabulary acquisition in the delayed posttest.

### 3.3. Discussions

This chapter provides a detailed analysis of key research findings presented in sections 4.1 and 4.2, with reference to each of the research questions. The results of the study are also discussed in relation to previous research studies. The statistical results of two research questions are summarized as

follows:

First, task type had a significant effect on the immediate word form recognition and passive word meaning recall. Specifically, there were significant differences between rewording and reading comprehension with glosses in the immediate word form recognition posttest ( $p = .013 < .05$ ) and between composition and reading comprehension with glosses in the immediate word meaning recall posttest ( $p = .043 < .05$ ). Based on the mean differences of the groups, rewording (an index of 3 in the ILH versus an index of 6 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word form recognition posttest and composition (an index of 3 in the ILH versus an index of 8 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word meaning recall posttest. Both the ILH and the TFA predicts that the tasks with higher indexes would contribute to better L2 vocabulary learning. The indexes of rewording and composition in the ILH and the TFA are both higher than those of reading comprehension with glosses. According to the predictions of both the ILH and the TFA, rewording and composition should contribute to better L2 vocabulary learning than reading comprehension with glosses, which is consistent with the statistical results above. As a result, these findings lent support to both the ILH and the TFA. Moreover, no significant difference in immediate word form recognition or word meaning recall between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) lent support to the ILH but not to the TFA, since the

ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in immediate word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

Second, task type had no significant effect on the delayed word form recognition or passive word meaning recall. Specifically, there were no significant differences among the task comparisons. Specifically, no significant difference between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) in delayed word form recognition and word meaning recall lent support to the ILH but not to the TFA, since the ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in delayed word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

In summary, it could sum up that both the ILH and the TFA could partly explain task effectiveness in facilitating EFL vocabulary acquisition, but not truly. However, compared to the TFA, the ILH could better explain task effectiveness in facilitating EFL vocabulary acquisition in the delayed posttest.

Based on the statistical results shown in the immediate and delayed posttests, the following two parts are illustrated from perspectives of task effects on initial EFL vocabulary learning and task effects on EFL vocabulary retention.

### **3.3.1. Task Effects on Initial EFL Vocabulary Learning**

Task type had a significant effect on EFL word form recognition and passive word meaning recall in the immediate posttest. Specifically, in word form recognition, rewording outperformed reading comprehension with glosses. Composition outperformed reading comprehension with glosses in word meaning recall. Rewording and composition are both output-driven tasks in which more attention is directed to target words and involve the deeply processing of the target words, i.e., word form and meaning. However, reading comprehension with glosses is an input-driven task in which doesn't focus on word form and meaning and are passive. Output-driven tasks could contribute to better L2 vocabulary learning than input-driven tasks and plays an important role in L2 vocabulary learning (Bao 2015; 2017; Webb 2009).

The findings above revealed that the ILH could partly explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition. According to the ILH (Laufer & Hulstijn 2001), tasks which have the same involvement load would contribute equally to L2 vocabulary learning and tasks which have a higher involvement load should lead to better learning effects. Therefore, in the four reading tasks, participants doing composition (an index of 3 in the ILH) should have the same performance as participants doing rewording (an index of 3 in the ILH), followed by gap filling (an index of 2 in the ILH) and reading comprehension with glosses (an index of 1 in the ILH). In the study, there was a significant difference between rewording and reading comprehension with glosses in the immediate word form

recognition posttest and between composition and reading comprehension with glosses in the immediate word meaning recall posttest. Specifically, rewording (an index of 3 in the ILH) performed better than reading comprehension with glosses (an index of 1 in the ILH) and composition (an index of 3 in the ILH) performed better than reading comprehension with glosses (an index of 1 in the ILH). Additionally, there was no significant difference between rewording and composition in the immediate posttests. All the findings above revealed that the ILH could explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition (Laufer & Hulstijn 2001; Kim 2008). Laufer & Hulstijn (2001) studied the effect of involvement load on the memory of 10 English words by adult EFL learners. They designed an experiment with three tasks (reading comprehension, reading plus gap filling, and composition), which had different involvement load indexes. The results showed that memory was related to the involvement load of the task: the memory in composition task was the best, with that in reading plus filling lower and that in reading comprehension the lowest. Kim (2008) conducted an experiment by partially replicating the research of Laufer & Hulstijn (2001) to investigate whether the different involvement load indexes of tasks had effects on the initial learning and memory of the target words of EFL learners. The results were consistent with the hypothesis of the ILH that the tasks with higher involvement load index should lead to better initial learning and the memorization of new words. Kim (2008) also compared two tasks (composition and making sentence) to find whether the tasks with the same involvement load index would have the same impact on initial learning or subsequent new word memory. The results of this study showed that tasks with the same index of involvement load were equally effective to vocabulary learning.

The findings still revealed that there were some limitations of the ILH's relative effectiveness in accounting for task type effects on EFL vocabulary acquisition (Folse 2006; Huang 2004; Bao 2015; Bao & Li 2017). According to the ILH (Laufer & Hulstijn 2001), tasks which have the higher involvement load would lead to better learning effect. Therefore, in the four reading tasks, participants doing composition (an index of 3 in the ILH) should have the same performance as participants doing rewording (an index of 3 in the ILH), followed by gap filling (an index of 2 in the ILH) and reading comprehension with glosses (an index of 1 in the ILH). Whereas in this study, there were no significant differences between gap filling and reading comprehension with glosses, between rewording and gap filling and between composition and gap filling. All the findings above revealed that there are some limitations of the ILH's relative effectiveness in accounting for task type effects on EFL vocabulary acquisition. Indeed, many studies have proven that tasks with higher involvement load (according to the ILH) may not necessarily produce better learning results (Folse 2006; Huang 2004; Bao 2015; Bao & Li 2017). Folse (2006) employed three vocabulary exercises: one blank filling exercise (an index of 2 in the ILH), three blank filling exercises (an index of 2 in the ILH) and one sentence-writing exercise (an index of 2 in the ILH), to test the prediction. It was found that three blank filling exercises outperformed the other two, rejecting the prediction. He also regarded word exposure as an important factor for EFL vocabulary acquisition. Huang (2004) employed three reading tasks to testify the relative effectiveness of the ILH, including reading

comprehension plus choices, reading comprehension plus gap filling and reading comprehension plus sentence-making. It was found that there existed some limitations of the ILH's relative effectiveness in accounting for task type effects on EFL vocabulary acquisition. Moreover, Bao (2015) found that the definition task (an index of 2 in the ILH) performed better than the writing task (an index of 3 in the ILH). He explained that task effectiveness maybe not related to the involvement load. He also concluded that factors like contextual clueing and frequency of word encounters were more effective in facilitating EFL learners' vocabulary acquisition than the involvement load. Additionally, in the study of Bao & Li (2017), the English learners were randomly assigned to formal, semantic or mixed tasks, and were measured in terms of new word spelling, part of speech and meaning. It was found that there were no differences in spelling and part of speech knowledge among tasks, and there were no differences in semantic knowledge between semantic tasks and mixed tasks, but they were better than formal tasks.

Rewording outperformed reading comprehension with glosses in the immediate word form recognition and composition outperformed reading comprehension with glosses still lent support to the TFA. These findings were consistent with the prediction of the TFA (similar to the ILH prediction) that tasks with higher indexes should lead to better learning effects (Ahour & Dogolsara 2015; Laufer & Hulstijn 2001; Kim 2008). A possible reason could be the orientation of processing tasks. The tasks in Laufer & Hulstijn (2001) were reading comprehension only, reading comprehension plus close and reading comprehension plus writing. Similar tasks designed by Kim (2008) were reading-only task, reading plus filling in task and reading plus writing task. Additionally, Ahour & Dogolsara (2015) designed two meaning-oriented tasks, sentence-writing task and multiple-choice task, and the prediction that tasks with higher involvement load contribute to better learning effects was testified. As stated previously, Rewording and composition are both output-driven tasks in which more attention is directed to target words and involve the deep processing of the target words, i.e., word form and meaning. However, reading comprehension with glosses is an input-driven task in which doesn't focus on word form and meaning and involve the shallow processing of the target words. Output-driven tasks could contribute to better L2 vocabulary learning than input-driven tasks and plays an important role in L2 vocabulary learning (Bao 2015; 2017; Webb 2009).

Compared to the TFA, the ILH could better explain the relative effectiveness of tasks in facilitating EFL vocabulary acquisition. Based on the discussion above, it is concluded that both the ILH and the TFA could partly explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition due to the similarity of the prediction of the ILH and the TFA that tasks with the same indexes would contribute equally to L2 vocabulary learning and tasks with higher indexes would lead to better vocabulary acquisition. In order to further compare the relative effectiveness of the ILH and the TFA, the relationship between gap filling (an index of 2 in the ILH versus an index of 7 in the TFA) and rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) should be taken into serious consideration. Based on the ILH prediction, there should be a significant difference between gap filling and rewording, with

rewording better than gap filling, and there should exist no significant difference between rewording and composition. However, based on the TFA prediction, there should be significant differences between gap filling and rewording and between rewording and composition, with gap filling and composition better than rewording. This study showed no significant differences between gap filling and rewording, supporting neither the ILH nor the TFA. Moreover, there were no significant difference between rewording and composition, which lent support to the ILH, but not to the TFA. These findings supported that compared to the TFA, the ILH could better explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition, which was not consistent with some studies (Hu & Nassaji 2016; Gohar et al. 2018). They have proven that the TFA could better explain the relative effectiveness of task type in facilitating the EFL vocabulary acquisition. Hu and Nassaji (2016) designed four tasks, reading a text with multiple-choice items (an index of 3 in the ILH versus an index of 6 in the TFA), reading a text and choosing definitions (an index of 3 in the ILH versus an index of 6 in the TFA), reading plus filling in the blanks (an index of 2 in the ILH versus an index of 7 in the TFA), and reading and rewording the sentences (an index of 3 in the ILH versus an index of 6 in the TFA). Gohar et al. (2018) did a similar experiment as Hu & Nassaji (2016). There were three tasks in their study: Participants in the first task were allotted five minutes to read the target words alphabetically arranged with their definition (in L1) and an example sentence. Then they had ten minutes to write one sentence for each word. ILH states that this task induces an involvement of 3 (1+0+2). Participants in the second task needed to write a composition (e.g., a letter to someone) using the given target words. Like the previous task, the participants had five minutes to read the target words with their definitions (in L1) and example sentences. ILH imposes a high involvement of 3 (1+0+2). TFA, which however, yields the motivation of 2, noticing of 2, retrieval of 0, generation of 3, and retention of 2 giving a moderate TFA as 9 out of 18 scores. Participants in the third task as the reading comprehension task were required to read the reading passage and answer the provided comprehension questions. Therefore, ILH imposes the low involvement of 1 (1+0+2) and TFA imposes the low index of 3 on learners. The two studies both found that the TFA could better explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition.

In summary, both the ILH and the TFA could partly explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition, but not truly. However, compared to the TFA, the ILH could better explain task effectiveness in facilitating EFL vocabulary acquisition. Moreover, according to the statistical results, there existed significant differences between rewording (an index of 3 in the ILH) and reading comprehension with glosses (an index of 1 in the ILH) and between composition (an index of 3 in the ILH) and reading comprehension with glosses (an index of 1 in the ILH). However, there existed no significant differences between gap filling (an index of 2 in the ILH) and reading comprehension with glosses (an index of 1 in the ILH), between gap filling (an index of 2 in the ILH) and rewording (an index of 3 in the ILH), between gap filling (an index of 2 in the ILH) and composition (an index of 3 in the ILH) and between rewording (an index of 3 in the ILH) and composition (an index of 3 in the ILH). This might be because the distinction between low and high involvement load indexes of tasks

might not be large enough to create differences in the learning outcomes and another possibility may also be attributed to the different weights of the three components in the ILH, and this was also proposed by Laufer in 2003 (Hu & Nassaji 2016). In the current study, the three components in the ILH of four tasks are shown as follows: reading comprehension with glosses [1 (need) + 0 (search) + 0 (evaluation)]; gap filling [1 (need) + 0 (search) + 1 (evaluation)]; rewording [1 (need) + 0 (search) + 2 (evaluation)]; composition [1 (need) + 0 (search) + 2 (evaluation)]. It could be found that the need and search components in four tasks were all the same, but the evaluation component was different. The relative effectiveness of the ILH in accounting for task type effects on EFL vocabulary acquisition depends on how large the difference of evaluation component is between tasks.

### 3.3.2. Task Effects on EFL Vocabulary Retention

Task type had no significant effect on the delayed word form recognition or passive word meaning recall. Specifically, no significant difference in word form recognition or passive word meaning recall between rewording and composition supported the ILH but not the TFA. All the other task comparisons demonstrated no significant difference in word form recognition or passive word meaning recall, supporting neither the ILH nor the TFA.

This above results also supported that the ILH could better explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition, which was not consistent with some studies (Hu & Nassaji 2016; Gohar et al. 2018). They have proven that the TFA could better explain the relative effectiveness of task type in facilitating the EFL vocabulary acquisition.

It could be concluded that task type effects on EFL vocabulary acquisition were restrained by the time. There were significant differences between rewording and reading comprehension with glosses and between composition and reading comprehension with glosses in the immediate posttests. However, there were no significant differences among all the task comparisons in the delayed posttests. EFL vocabulary acquisition is a circulatory and long-term process which needs continuous accumulation. The memory of EFL vocabulary could not be permanent merely by adding the word encounter numbers for once (Folse 2006; Eckerth & Tavakoli 2012; Baddeley 1990; Peter et al. 2009; Wei & Wang 2011; Bao & Wang 2013; Huang 2004). Folse (2006) found that the participants who completed three gap filling tasks using target words had better performance on the immediate posttest than those who completed one sentence-making task. Although Eckerth & Tavakoli (2012) also found that there was a significant main effect of frequency in the immediate posttest, the frequency effect disappeared in the delayed posttest. Baddeley (1999) suggested that learners should repeat new words as soon as possible after they first touch them in order to strengthen the connection between the form and meaning of new words, and to promote the retention of memory. The study of Wei & Wang (2011) on the frequency of task also confirmed the point above. Peters et al. (2009) further showed that the delayed acquisition of word meaning was not only the elaborate processing of new words or repeated word exposure, but also the combination of the two. In the study of Bao & Wang (2013), there were significant differences in the immediate posttest and no significant differences in the delayed posttest. In the study of Huang (2004), there were significant differences between reading comprehension plus choices and reading comprehension plus gap filling and

between reading comprehension plus choices and reading comprehension plus sentence-making in the immediate posttest. However, there were significant differences between reading comprehension plus choices and reading comprehension plus sentence-making in the delayed posttest. The above findings showed that task type effects on vocabulary acquisition were restrained by the time.

What's more, the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition was also constrained by the time. In the immediate posttest, the significant differences between rewording and reading comprehension with glosses and between composition and reading comprehension with glosses lent support to both the ILH and TFA. The non-differences between rewording and composition lend support to the ILH and the non-differences among other task comparisons supported neither the ILH nor the TFA. In the delayed posttest, the non-significant differences between rewording and composition also lent support to the ILH, with the non-significant differences among other task comparisons supporting neither the ILH nor the TFA. It could be concluded that the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition was also constrained by the time. The advantage of task induced involvement load was conducive to the immediate EFL vocabulary acquisition, but this advantage was lost without repetition over time. (Wu et al. 2007; Wu 2010). In the study of Wu (2007), in order to further test the ILH proposed by Laufer & Hulstijn (2001), this study recorded the online learning behaviors of Non-English Majors in China when they completed different tasks. The statistical analysis showed that the online learning behavior of the subjects seemed to be more affected by the nature of tasks rather than the amount of involvement load and the amount of involvement load had no significant effect on the initial learning and memory retention of vocabulary. In the study of Wu et al. (2007), they made a further demonstration of the ILH through a questionnaire survey and a scale test of the English majors.

Word exposure is of much importance and significance in EFL vocabulary acquisition. Due to the comparison of the immediate posttest and the delayed posttest, it can be found that the means and 95% CI of four groups in the delayed posttest were all lower than those in the immediate posttest. In the study, one week after the immediate posttest, all the four groups took the delayed posttests without repetition of tasks. As a result, the frequency of word encounters was three in the immediate posttest, but the frequency of word encounters was just two in the delayed posttest. It can be summarized that word exposure is of much importance and significance in EFL vocabulary acquisition. A lot of research found that much new words encounters were needed before a new word was acquired (Nagy et al. 1985; Rott 1999; 2007; Coady 2001). Nagy et al. (1985) found that vocabulary acquisition is a long-term process during which words are gradually mastered through repeated exposure to learners. In his study, he concluded that the probability of acquisition with six or seven encounters was much higher than only one encounter. A similar conclusion was drawn by Rott (1999) who found that repeated encounters of target words helped vocabulary learning with six times contributing the most. It was difficult for EFL learners to acquire the word if the word only appeared once in a text. This statement was further proven in the study conducted by Coady (2001). It has been

found that the probability for the acquisition of a new word at first exposure ranged from 5% to 15%. Noticeably, Knight (1994) found that a single exposure of target words is helpful in vocabulary acquisition. Ellis (2004) also found that repeated encounters might not result in better learning effect due to EFL learners' inability to connect with the target word through every encounter and then failure to maintain form and meaning connection. Although there are some exceptions, a large number of studies have proven the positive effect of repetition. Moreover, the comparison of the results between the immediate posttest and the delayed posttest still conforms to the mechanism of human memory that it's easy for EFL learners to forget new words without timely review even though the new words appear repeatedly in the passage (Spada & Lightbown 1993).

In summary, the ILH could better explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition in the delayed posttest, which was the same with the results in the immediate posttest. It could be concluded that task type effects on EFL vocabulary acquisition and the relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition were both restrained by the time. Moreover, large differences in EFL vocabulary knowledge between the immediate and delayed posttest indicated word exposure is of much importance and significance in EFL vocabulary acquisition.

## 4. Conclusion

### 4.1. Major Findings

This study employed one-way ANCOVA (analysis of covariance) design with the between-subject factor being task type and the covariate being EFL vocabulary knowledge. The dependent variable was EFL vocabulary acquisition, which included two dimensions, i.e., word form recognition and passive word meaning recall. The EFL vocabulary knowledge posttests included the immediate and delayed posttests. Two ANCOVAs were conducted to test the task type effects on initial vocabulary learning and retention respectively. The major findings are summarized below:

First, task type had a significant effect on the immediate word form recognition and passive word meaning recall. Specifically, there were significant differences between rewording and reading comprehension with glosses in the immediate word form recognition posttest ( $p = .013 < .05$ ) and between composition and reading comprehension with glosses in the immediate word meaning recall posttest ( $p = .043 < .05$ ). Based on the mean differences of the groups, rewording (an index of 3 in the ILH versus an index of 6 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word form recognition posttest and composition (an index of 3 in the ILH versus an index of 8 in the TFA) outperformed reading comprehension with glosses (an index of 1 in the ILH versus an index of 5 in the TFA) in the immediate word meaning recall posttest. Both the ILH and the TFA predicts that the tasks with higher indexes would contribute to better L2 vocabulary learning. The indexes of rewording and composition in the ILH and the TFA are both higher than those of reading comprehension with glosses. According to the predictions of both the ILH and the TFA, rewording and composition should contribute to better L2 vocabulary learning than reading comprehension with glosses, which is consistent with the statistical results above. As a

result, these findings lent support to both the ILH and the TFA. Moreover, no significant difference in immediate word form recognition or word meaning recall between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) lent support to the ILH but not to the TFA, since the ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in immediate word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

Second, task type had no significant effect on the delayed word form recognition or passive word meaning recall. Specifically, there were no significant differences among the task comparisons. Specifically, no significant difference between rewording (an index of 3 in the ILH versus an index of 6 in the TFA) and composition (an index of 3 in the ILH versus an index of 8 in the TFA) in delayed word form recognition and word meaning recall lent support to the ILH but not to the TFA, since the ILH predicts that tasks with the same involvement load index would contribute equally to L2 vocabulary learning. Similarly, based on the predictions of both the ILH and the TFA, there should be significant differences among all the other task comparisons due to the index differences. It turns out that all the other task comparisons demonstrated no significant differences in delayed word form recognition and word meaning recall, supporting neither the ILH nor the TFA.

In summary, both the ILH and the TFA could partly explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition, but not truly. However, compared to the TFA, the ILH could better explain task effectiveness in facilitating EFL vocabulary acquisition.

### 4.2. Implications

#### 4.2.1. Theoretical Implications

Theoretically, the present study sheds light on theory reconstructing, as both the ILH and TFA need to be reconsidered while exploring the field for a finer-grained theoretical framework.

In this study, both the ILH and the TFA could partly explain the relative effectiveness of task type in facilitating EFL vocabulary acquisition, but not truly. However, compared to the TFA, the ILH could better explain task effectiveness in facilitating EFL vocabulary acquisition. Based on the discussions above, the limitations of the ILH could be explained base on three aspects: For starters, the relative effectiveness of the ILH turns weaker when the distinction between low and high involvement load index of tasks isn't large enough to create differences in the learning outcomes. Another possibility may be attributed to the different weights of the three components in the ILH. It could be found that the need and search components in four tasks were the same, but the evaluation component was different. The relative effectiveness of the ILH and the TFA in accounting for task type effects on EFL vocabulary acquisition depends on how large the difference of evaluation is between tasks. Again, the relative effectiveness of both the ILH and the TFA is also restrained by the time and turns weaker over time.

As a result, both the ILH and TFA need to be reconsidered and a finer-grained theoretical framework needs to be

explored. Moreover, in order to find a more reliable and comprehensive model, the following research should attach much importance and significance to several aspects: firstly, when choosing task type, the researchers should make sure that the distinction of involvement load index of tasks is rational. Secondly, emphasis should be put on the relative effectiveness of the specific component of the ILH and the TFA in facilitating EFL vocabulary acquisition. In this current study, the index of need and search is the same in four tasks so as to demonstrate the effectiveness of evaluation in accounting for task type effects on EFL vocabulary acquisition. More efforts should be made to investigate how variation in the index of need and search may affect EFL vocabulary acquisition. Thirdly, researchers should pay attention to other factors so as to make the study more valid and rational.

#### 4.2.2. Pedagogical Implications

This study gives teachers some suggestions for designing word-focused activities through reading to facilitate EFL vocabulary acquisition. Teachers should make use of different word-focused tasks to enhance EFL vocabulary learning and help EFL learners know the importance of repetition in EFL vocabulary learning.

Specifically, when designing vocabulary exercises, EFL instructors could make use of different word-focused tasks like gap-filling, rewording and composition to enhance EFL vocabulary learning when the need arises since the differences of the effects of four tasks in facilitating EFL vocabulary acquisition were not apparent. When the time is adequate, EFL instructors could make use of composition which needs extra time. When the time is limited, EFL instructors could make use of rewording or composition which doesn't need much time.

Moreover, EFL instructors should put emphasis on repetition from different perspectives such as form recognition, meaning recall and so on since word exposure is of much importance and significance in EFL vocabulary acquisition, which is a circulatory and a long-term process.

#### 4.3. Limitations and suggestions for future research

The most obvious limitation in this research was that the sample size of each group isn't equal in the current study due to the loss of participants. With an equal sample size, more differences would emerge. Therefore, it is necessary for further researchers to conduct similar studies to put emphasis on the equality of the sample size while trying to avoid the loss of participants.

The study was further limited by the types of task. The need and search components in the ILH of all tasks were the same and the distinction of the indexes in the TFA of all tasks was not large as there are all 18 points in the TFA. It is suggested that more studies should be conducted on the tasks with more different indexes in the ILH and the TFA. Moreover, more efforts should be made to investigate how variation in the index of need and search may affect EFL vocabulary acquisition.

Finally, the research findings of the study were limited by the types of tested word knowledge. The current study merely considered word form recognition and passive word meaning recall without other components of word knowledge. It is suggested that more types of tested word knowledge, such as productive word knowledge, receptive word knowledge and so on, should be used in the following studies so as to compare

the effects of tasks on EFL vocabulary acquisition more comprehensively.

## References

- [1] Ahour, T., & Dogolsara, S. A. (2015). The Effect of task type on Iranian advanced EFL learners' vocabulary learning. *Theory and Practice in Language Studies*, 5(3), 657-664.
- [2] Baddeley, A. D. (1990). *Essentials of Human Memory*. Hove: Psychology Press.
- [3] Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, 53, 84-95.
- [4] Bao, G., Li, J. (2017). Effects of output task and glossing on EFL learners' vocabulary acquisition, *Journal of PLA University of Foreign Languages*, (6): 5-14.
- [5] Bao, G., Wang, J. (2013). How prior vocabulary size and time moderate the effect of task on incidental L2 vocabulary acquisition, *Modern Foreign Languages*, (4): 69-76.
- [6] Bao, G., Wu, Y. (2019). An experimental study on task type and English vocabulary acquisition, *Foreign Language Testing and Teaching*, (2): 41-49.
- [7] Ellis, N. (1994). Consciousness in L2 learning: psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, 11(1), 37-56.
- [8] 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.
- [9] Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273-293.
- [10] Kim, Y. J. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285-325.
- [11] Hu, M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with retention of word meanings inferred from context. *Canadian Modern Language Review*, 68, 54-77.
- [12] Hu, H. M., & Nassaji, H. (2012). The relationship between task-induced involvement load and learning words from context. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 5, 69-86.
- [13] Hu, H. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus Technique Feature Analysis. *System*, 56, 28-39.
- [14] Huang. (2004). Testing the Involvement Load Hypothesis: effects of reading-based tasks on vocabulary retention by Chinese EFL learners, *Modern Foreign Languages*, (4): 386-394.
- [15] Gohar, M. J., Rahmadian, M., & Soleimani, H. (2018). Technique Feature Analysis or Involvement Load Hypothesis: Estimating their predictive power in vocabulary learning. *Journal of Psycholinguistic Research*, 47(4), 859-869.
- [16] Gu, Q., Song, M. (2010). Task Presupposition Involvement Load and Vocabulary Acquisition Strategies, *Journal of PLA University of Foreign Languages*, (5): 57-61.
- [17] Nation, P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Heinle.
- [18] 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, 567-587.

- [19] Laufer, B., & Hulstijn, J. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539-558.
- [20] Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- [21] Nagy, W. E., Anderson, R. C., & Herman, P.A., (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- [22] 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.
- [23] Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
- [24] Sternberg, R. J. (1987). Most Vocabulary is Learned from Context. In McKeown, M. G., & Curtis, M. E. (Eds.). *The Nature of Vocabulary Acquisition* (pp. 89-105). Hillsdale, N. J.: Erlbaum.