Task Types and Vocabulary Growth in a Foreign Language: The Involvement Load Hypothesis on Trial

Feng Teng, Nanning University, Nanning, People's Republic of China
Shu Zhang, Nanning University, Nanning, People's Republic of China

Abstract

This study investigated the vocabulary growth of low-proficiency, tertiary-level students learning English as a Foreign Language (EFL). Graded readers, and activities based on the task-induced Involvement Load Hypothesis (ILH) proposed by Hulstijn & Laufer (2001), were used. Through a preliminary test of the first 1,000 English words (Nation, 1993), 180 students were chosen from different majors. The first group of 60 students read graded readers. The second group of 60 students was given graded readers in addition to having the teacher's read the material aloud. The third group read the graded readers and wrote compositions that incorporated the target words. Based on the ILH, it was predicted the third group's vocabulary would improve the most after reading the graded readers and writing compositions. The study lasted for six months, during which all three tasks were conducted two times. A vocabulary test designed to assess the form recall was used as the research instrument. Statistical analysis of the data showed that, in line with the hypothesis, the third group's vocabulary increased the most immediately after tasks. However, the benefits of repeated tasks did not hold for the participants, and even the task with the highest degree of involvement suffered a significant decrease in recalling the form of target words after a two month period. This study offered rich opportunities for English teachers to experience the graded readers-approach in three different ways to help enlarge EFL students' vocabularies.

Keywords: Graded Readers, task, vocabulary learning, Involvement Load Hypothesis, vocabulary gain

Introduction

A great many efforts to teach the English language are devoted to the outcomes of learning vocabulary. The ability to comprehend English words is the central building block for reading, and the lack of a proficient vocabulary is regarded as

a barrier for reading comprehension (Corson, 1995; Elleman, Lindo, Morphy, & Compton, 2009). As has been shown in several studies, reading comprehension involves an interconnection of writer, reader, and context, and also requires multiple exposures to the target language rather than simply the knowledge of individual words (e.g., Beck & Mckeown, 1991; Nagy & Scott, 2000; Stahl, 2003; Stahl & Nagy, 2006). In the Chinese or other Asian context of learning English as a Foreign Language (EFL), constant exposure to the target language is often lacking. Despite this, deliberate attention to individual words receives the greatest attention in teaching English. Under such circumstances, teachers often focus on explaining the meaning of the target words in their native language, with the remainder of the vocabulary learning dependent on students' efforts. This phenomenon often results in low vocabulary learning outcomes even after reading considerable material written in the target language.

However, we cannot deny the benefits of reading in word learning. A number of scientific investigations have directly shown that reading improves language skills such as vocabulary (Cunningham & Stanovich, 1998; Martin-Chang & Gould, 2008; McLeod & McDade, 2011). Graded readers, written with high-frequency words and simplified language structures, are the most popular reading materials for EFL learners. The beneficial relationship between the potential of learning vocabulary and graded readers was demonstrated in Nation & Wang's (1999) study. The coverage, density, and repetition of vocabulary in the graded readers were measured in their study and reading graded readers was shown to be an effective method to improve vocabulary.

The current study used graded readers in the investigation of three lexical tasks with different Involvement Load Indexes (ILI) based on the Involvement Load Hypothesis (Hulstijn and Laufer, 2001). It attempts to expand our existing knowledge of task-induced involvement by testing its predictive power on learning vocabulary by Chinese low-proficiency EFL learners and by assessing its impact on form recall.

Review of the Literature

Graded readers

Graded readers refer to books that are written specifically to include a high frequency of the same words and simple grammatical structures (Nation & Wang, 1999). Hill &Thomas (1988, p.44) used the term "grading scheme" to define graded readers. Graded readers provide an opportunity for EFL learners with low proficiency to read printed English materials readily.

Although certain studies have pointed out that some graded readers have poor English syntax and lack content (Cobb, 2008; Davison, 1986; Wallace, 1988), many studies have demonstrated the positive effects of using simplified materials to teach English (e.g., Bamford, 1984; Claridge, 2005; Elley, 1991; Hill & Thomas, 1988). Bamford (1984) illustrated the characteristics of graded readers and their contribution to EFL teaching. His research helped validate the grading system and readability of grader readers. Elley (1991) invited elementary school students to participate in an experiment of "book floods." His research showed that the group of students using highly-interesting illustrated story books outperformed the group using a structured, audio-lingual program to learn the target words. This research demonstrated that tight control over syntax and vocabulary in story books leads to improved effects in learning new words incidentally. In a survey review of graded readers, Hill and Thomas (1988) stated that the graded readers are an excellent resource for learning and teaching English. In addition to this, Webb, Newton and Chang (2013) testified that learners could learn collocations from reading graded readers incidentally. Although there are still limits to graded readers, if designed appropriately, an effective reading program could bring results by making extensive use of these books (Teng, 2014a).

The Involvement Load Hypothesis (ILH)

Craik & Lockhart (1972) proposed the theory of depth and levels of processing in their research on human memory. They pointed out that a deeper level of processing leads to more detailed, meaningful, durable, and stronger memories. Hulstijn & Laufer (2001) further proposed the motivational-cognitive constructs of task-induced involvement, based on Craik & Lockhart's (1972) theory of depth

and levels of processing. To put it succinctly, word learning and retention are dependent on the amount of effort or involvement that a task induces.

The three involvement load constructs are need, search and evaluation. The component of need is a motivational but non-cognitive dimension of involvement which is differentiated as strong need (++) and moderate need (+) based on intrinsic or extrinsic factors. A need is strong when it is self-motivated (e.g., a learner wishes to learn for what he/she lacks) and moderate when it is imposed by extrinsic factors (e.g., tasks are imposed by teachers). Search and evaluation are the two cognitive dimensions of involvement. Search indicates the process to finding the meaning of an unknown word, (e.g., finding the definition by using a dictionary or ascertaining the explanation of unknown words by asking the teacher for help). Evaluation refers to the comparison of a new word with other words and measurement of its suitability in a given context. Evaluation is strong (++) when a task requires the learners to combine new words and known words in an original context (e.g., to create a sentence or write a composition) and moderate (+) when a task only requires the learners to recognize differences between words provided in a given context (e.g., make a decision about which meaning of the new word best fits the given context).

According to the Involvement Load Hypothesis, the three components do not always appear simultaneously during a reading task. Hustijn & Laufer (2001) proposed an Involvement Load Index (ILI) to define the depth of processing, in which the absence of a component is marked 0, the moderate presence of a component is marked 1, and the strong presence of a component is marked 2. According to the ILH, tasks with higher indexes are deemed more effective for learning and retaining a word than those tasks with lower indexes.

Empirical evidence for ILH

Empirical support for ILH is found in several studies that compare reading tasks vs. reading supplemented with focus-on-forms activities. The findings are as follows:

In Keating's (2008) study, word learning and retention for 79 Spanish speaking, beginning English learners were found to be dependent on a task's involvement load. Significant differences were also found in the three tasks: Reading comprehension (task one, ILI = 1), reading comprehension supplemented with tar-

get word suppliance (task two, ILI = 2), reading comprehension plus sentence writing (task three, ILI = 3). His study revealed that passive knowledge of the target words in task three was improved by 64.3%, task two by 47 %, and task one by 16.8%. Active knowledge of the target words in task three was improved by 42.1%, task two by 22.7%, and task one by 7.3%. To put it simply, tasks with higher involvement load were found to be more effective for word learning and retention than those with lower involvement load; similar results could also be found in other studies (Eckerth & Tavakoli, 2012; Hill & Laufer, 2003; Webb, 2005).

Huang, Eslami, & Willson's (2012) study investigated the effects of output tasks on incidental vocabulary learning. They analyzed 12 studies on the Involvement Load Hypothesis, measuring five mediator variables: Design quality, types of output task, time on task, genres of text, and text-target word ratios. Five conclusions resulted: a) English learners who complete an output task bring more satisfactory results than learners who are only tasked with reading comprehension; b) tasks with higher Involvement Load Indexes (ILI) yield more vocabulary gains than tasks with lower ILIs; c) studies with better scientific research designs produced more reliable results than studies with poorer research designs; d) positive effects were related to the time on task; e) learners who read a text with a text-target word ratio of less than 2% outperformed those learners who read texts with a ratio of 2%-5%.

Laufer & Rozovski-Roitblat (2011) compared the effects of task-induced involvement and frequency of word occurrence on long-term retention of words. In their study, learners attended two tasks: reading a text with occasional Focus on Form, such as referring to a dictionary resulted in lower ILIs and reading a text followed by Focus on Forms or word-focused activities resulted in higher ILIs. Learners encountered 60 test items during a 13-week course, and were subsequently tested on their passive recall and passive recognition. Their findings revealed that, although the second task of reading plus Focus on Forms is not superior to the first task of reading plus occasional Focus on Form when the words were met 2-3 times, the second task was significantly better than the first task with 4–5 and 6–7 word encounters, both for recall and recognition. In addition, in their questionnaire, all students recommended including more output activities with a high degree of involvement. Their results was in line with Min (2008), where learners who completed a task of reading comprehension supplemented with

word-focused activities significantly outperformed in target word learning and retention than those who only had a reading of thematically related texts with no follow up activities.

Rationale for current study

Research on the Involvement Load Hypothesis conducted so far has mainly been focused on advanced learners, who might have a more superior ability in school-based learning. With a background of success in learning a foreign language, and they may have more motivation or investment in completing the tasks. It remains to be seen whether EFL learners in Asian contexts with a low-proficiency level would also benefit from tasks with a high involvement load. It is quite possible that there are proficiency and lexical thresholds that learners must attain in order to fully benefit from extensive reading (Pichette, 2005) and the lexical tasks required in the research to date. Second, research conducted to date has not investigated the effect of task repetition.

Research questions

This study attempted to provide a more complete picture of vocabulary gains from task types through reading graded readers. The following research questions were addressed:

1) Do low-proficiency students assigned to each group have significant vocabulary growth (as measured by form recall) after taking part in the three tasks for the first time?
2) Do tasks with different ILIs lead to different gains in learning vocabulary (form recall), and does the third task, with the highest ILI, yield the greatest improvement (immediate post first-time task)?
3) Do the benefits hold when the tasks are repeated?
4) Do learners exhibit long-term retention of vocabulary (two months after the experiment), and does the third task, which has the highest ILI, yield the best results?

Method

Participants

There were 260 students who originally took part in the study. They were aged 18-21 and had no experience studying abroad. They were from a variety of majors and studied English as an elective course. They all agreed to take part in this reading experiment and to take Nation's (1993) 1,000 word level vocabulary test before finally being admitted to the study program.

Table 1. Results of Pre-task test

Lower than 10	11-20	21-30	31-40	M	S.D.
1	38	182	39	25.62	6.28

There were 182 students whose scores ranged from 21-30 points and these were chosen as participants in this study. After the test, two students stated they were not satisfied with the arrangement and elected to drop out. Thus, the final number of participants was 180, with 120 males and 60 females. They were assigned randomly to three groups of 60 learners each.

Target words

Forty items were carefully selected from the graded readers used by the learners. These words were: *disappear, growl, ache, frightened, pass, castle, return, notice, sound, comfortable, deeply, gloom, crash, foggy, miserable, portrait, extraordinary, reach, invent, imagination, airless, turn, measure, obey, light, realize, damage, thoughtful, tear, recognize, criminal, murder, condition, yell, tiredness, whistle, desperately, explode, adventure, complain*. These words were selected based on two criteria: First, the frequency of word occurrence (measured by one of the computer programs on the Compleat Lexical Tutor website (Cobb, n.d.). Second, the target words appeared at least eight times in each book. According to previous research (Teng, 2014b; Waring & Takaki, 2003), learners tended to incidentally learn words with a frequency of at least eight.

Materials

In order to enlarge the student's lexicons, the students had to be exposed to target words that were beyond their current lexical level. In SLA terms, that is i+1, where *i* represents language at the students' current level of competence (Krashen, 1982, 1985). However, some students' lexicons were so limited that it was very difficult to find appropriate reading materials to increase their vocabulary, as the limits of the students' word recognition and decoding abilities had to be considered. This study scrupulously investigated the appropriateness of the materials for this reading program. The materials for the tasks were selected from the Bookworms series published by Oxford University Press. Considering the range of students' scores from 21-30, Nation (1993) proposed that books in the third level are appropriate. Thus, four books in the third level were used in the experiment: *Frankenstein, The Call of the Wild, The Secret Garden,* and *The Prisoner of Zenda* (See Table 2).

Table 2. Text Size and Vocabulary Size of the Textbooks

Book	Tokens	Type	Families
Frankenstein	17192	3044	647
The Call of the Wild	20295	3526	222
The Secret Garden	19436	3341	202
The Prisoner of Zenda	19524	3289	188

Procedures

Three tasks were used, each with different involvement load indexes (ILI). The details of the tasks and the involvement load index scores are shown in Table 3.

Table 3. Three Tasks and their ILIs

Tasks	Input	Conditions for target words	Procedures	Involvement Load			
				Need	Search	Evaluation	Index
Task 1 RC	Written texts	Natural	Read the text	+	–	–	1
Task 2 RC plus read aloud	Written texts with explanation of target words	Marginally glossed in L2	1, Read the text 2, teacher read aloud, the students listen 3, Read aloud the target words	+	–	+	2
Task 3 RC plus writing a composition with the target words	Written texts with explanation of the target words	Marginally glossed in L2, with examples of usage	1, Read the text 2, Read the target words and the usage of the words 3, write a composition with the target words	+	–	++	3

RC=reading for comprehension

As is shown in the Table 3, task 3 has the highest ILI (3). These tasks were imposed by the teachers; the need was therefore moderate (+). The learners were not required to locate the meaning of unknown words by dictionary, thus search is not present in any of the three tasks. Task 1 is simply a reading task (there is no evaluation), and task 3 requires writing a composition, which requires strong motivation (++), thus has a higher ILI score than task 2 (+).

Data were collected during class hours. The study lasted for 24 weeks (see Table 4).

Table 4. Procedures Summarized

	Group 1	Group 2
Session 1: week 1-12	–	Teacher's read-aloud
	First-time reading of four books. Designed vocabulary test administered before the tasks and after the 12th week.	
Session 2: week 12-24	–	Teacher's read-aloud
	Second-time reading of four books. Designed vocabulary test administered after the 24th week. Questionnaire.	
Session 3: months post tasks	No further chance to read the four books. Designed vocabulary test was administered two months later.	

Students read one simplified book every three weeks. The classroom time allocated for each week was four hours. During that time, the students in the first group read the simplified books while the second researcher supervised the process. The students in the second group spent two hours reading the simplified books and for another two hours, their teacher read the selected texts aloud and the students followed along in their books. The students' reading process was supervised by the second author. For the third group, the students read the books, learned the usage of target words provided by the teacher, and then wrote a composition after reading each book. Both teachers reviewed the students' writing.

Participants finished the requirements of the three tasks in 12 weeks. With respect to task frequency, all participants conducted the same task-based study again in another 12 weeks. The vocabulary test designed to measure learners' vocabulary growth was administered pre-tasks, 12 weeks after first-time tasks, 24 weeks after second-time tasks, and two months post-experiment. During the two months post experiment, the participants did not read the four books. The administered time of the tests was not announced to the learners so that they would not commit the target words to memory to do well on the tests. Participants also answered a questionnaire after the second half of the course (where readings and tasks were repeated).

Assessment

Vocabulary Levels Test

To measure students' readiness to read simplified materials, Nation's (1993) test of 1,000 words was used. The first 1,000 words are essential in reading simplified materials and were used to assign appropriate graded readers. This test consisted of 40 items. Test takers got one point for each correct item. When a student's score is less than 10, he or she should be reading the first level of graded readers; students with scores of 10-20 are assigned the second level of graded readers; students with scores of 21-30 are assigned third level, and students with scores above 30 received level four readers (Nation, 1993, p.197). The following is an example of a test item:

We cut time into minutes, hours and days ____

(Write T if it is true, write F if it is false, write X if you don't understand)

Newly-designed vocabulary test

A new vocabulary test was created by the authors and administered four times: before the tasks, immediately post first-time tasks and second-time tasks, as well

as two months post experiment (see Appendix I). The test was identical each time except for the order in which the test items were presented. The main purpose of this test was to evaluate the effects of task type on vocabulary growth (as measured by form recall). According to some recent studies (Eckerth & Tavakoli, 2012; Peters, 2014), form recall is the smallest dimension of lexical knowledge that could be acquired by EFL students. Hence, more research is needed for measuring form recall. Participants had to supply the target words by using the given Chinese translation and the English definition. The first letter of the English word was given to prevent learners from creating an alternative, semantically correct answer (e.g., fade instead of disappear).

Example: 消失(something that cannot be seen suddenly) d_____

A correct answer received one point, an incorrect answer zero points. Answers containing minor spelling mistakes such as 'diappear' (instead of 'disappear') were scored as incorrect. The maximum score for this vocabulary test is 40 points.

Results

Descriptive statistics

Table 5 presents the descriptive statistics and scores for the newly-designed vocabulary test. The test was administered at four different times: pre-study, after the first 12-week session, after the second 12-week session (same readings and tasks repeated), and two months post-study. Figure 1 presents the same results graphically.

Table 5. Scores for the Vocabulary Test Administered at Four Different Times

Groups	Tasks	Pre-study		Post first session		Post second session		Two months post study	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.
Group 1(60)	RC	20.88	1.82	31.15	1.79	31.18	1.73	25.65	2.08
Group 2(60)	RC+RA	20.53	1.66	33.78	1.29	33.81	1.24	26.76	2.57
Group 3(60)	RC+RW	20.55	1.73	35.21	1.13	35.28	1.17	27.65	1.85

RC=Reading for Comprehension; RA=Reading Aloud; W=Writing Max=40

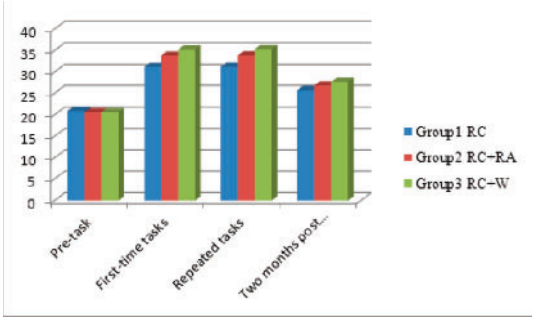

Figure 1. Vocabulary score: Form recall (Max=40)

As shown in Table 5, participants' scores on the vocabulary test revealed that they entered this reading experiment with fair vocabulary knowledge. After conducting the three tasks for the first time, the vocabulary outcomes in form recall for the three experimental groups were 31.15, 33.78, and 35.21 respectively, which demonstrated a substantial gain compared to their previous lexical knowledge. However, learners did not present an improvement after conducting the three tasks for the second time. In addition, learners showed decreased mean scores two months post experiment, as might be expected.

Table 6 summarizes responses to the questionnaire.

Table 6. Questionnaire (n=120)

1. Did you have any chances to learn the unknown target words outside this classroom study?	Yes No	0% 100%
2. Did simply reading help you remember new words?	Helped me a lot Helped me a little Did not help me at all	20% 45% 35%
3. Did teacher's read aloud help you remember new words?	Helped me a lot Helped me a little Did not help me at all	20% 45% 35%
4. Did sentence writing help you remember new words?	Helped me a lot Helped me a little Did not help me at all	20% 45% 35%
5. Did repeated tasks help you remember new words?	Helped me a lot Helped me a little Did not help me at all	20% 45% 35%
6. What other methods, in your opinion, are good for learning new words? (Please only choose one)	1. Be more engaged in more writing activities with target words 2. Have more post-reading exercises 3. Read-aloud and memorize the new words 4. Extensive reading 5. Simply memorizing new words	20% 45% 35%

Results showed that no participants had been exposed to the unknown target words outside the classroom (Question 1). Second, it showed the perceived value of each task, which is in line with the predictive power of each task (Question 2, 3, and 4). Third, it showed that more than a half of the participants were not willing to be engaged in repeating the tasks (Question 5). Fourth, concerning further suggestions for learning new words, students tended to prefer tasks that involved using the new words in post-reading and writing activities. Few participants perceived the value of simply memorizing new words (Question 6).

Inferential statistics

To understand whether each task provided a predictive power in form recall, a paired-sample *t*-test was applied in analyzing whether there was a statistically significant improvement between each administered time (See Table 7).

Table 7. Comparison of Group t-scores between Each Administration

Groups	Tasks	*t*-test (Pre-task and Post first-time tasks)		*t*-test (between first-time task and Post second-time tasks)		*t*-test (between Post second-time tasks and Two months later)	
		t-score	*p=*	*t-score*	*p=*	*t-score*	*p=*
Group1(60)	RC	-16.88	.00*	-.62	.53	27.45	.00*
Group2(60)	RC+RA	-34.48	.00*	-.70	.48	45.39	.00*
Group3(60)	RC+W	-37.31	.00*	-2.05	.14	58.69	.00*

**P*< 0.05 RC=*Reading for Comprehension; RA= Reading Aloud; W= Writing*

As shown in Table 7, each group showed a statistically significant improvement after completing the tasks for the first-time (negative *t*-scores indicate the second set of scores were superior to the first set; $p<0.05$). Apparently, however, benefits from completing the tasks for a second-time were insignificant ($p>0.05$). Finally, as might be expected, all groups' scores decreased two months later.

To probe further how the facilitative effects that each task provided differed from one another, Multivariate ANOVA was conducted. The results are presented below.

Table 8. Comparison of the Three Groups' Test Scores at Different Test Time

	Pre-task		First-time tasks		Second-time tasks		Two months post experiment	
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
Among three tasks	.772	.46	124.12	.00*	136.35	.00*	156.74	.00*

*$p<0.05$

As shown in Table 8, the F-value demonstrated that the three groups of students had almost the same pre-task levels of lexical knowledge. In addition, Multivariate ANOVA for vocabulary form recall showed main effects for task type.

To probe further which task had the most facilitative power in promoting form recall, post-hoc Tukey's analysis was applied. Results indicated the mean scores of Task 3 and Task 2 were significantly larger than that of Task 1($p<0.05$), and the mean score of Task 3 was significantly larger than the mean score of Task 2($p<0.05$). This result was consistent in three administered test times (post first session, post second session, two months post study). To put it simply, the RC+W (reading plus writing task) group ranked the highest, the group with RC+RA (reading plus reading aloud task) group ranked second, and the group that simply read the graded readers ranked the lowest.

Discussion

Responses to the research questions

In summary, the data revealed, in response to the four research questions, the following:

1) The students who were assigned to each group had significant vocabulary growth in learning form recall after taking part in the three tasks for the first time
2) Tasks with differential ILIs led to differential gains in learning form recall. The third task, which had the highest ILIs, yielded the greatest improvement in learning form recall immediately post first-time task
3) Participants did not show improved scores after taking the tasks for second time (i.e., mere repetition led to no further improvement).
4) Students in the third task showed the greatest result in retention of target words two months after the experiment (though all students in all tasks showed an overall decrease in retention).

Implications for teaching and learning

Task type was found to affect students' vocabulary growth in form recall significantly. The general conclusion concerning the effects of task type is that all the tasks facilitated students' proficiency in form recall, and that the facilitating power of each task was significantly different. The findings also suggested that the graded readers, which are written with a limited vocabulary, could be used as helpful reading resources for the students to improve their lexicons, especially for those learners with a vocabulary lower than 1,000 word families.

The test results also revealed that the overall power of each task was consistent with the Involvement Load Hypothesis (Hulstijn & Laufer, 2001). In line with previous research (Hill & Laufer, 2003; Keating, 2008; Laufer, 2003; Webb, 2005), the R+W task, which was assigned the highest level in the motivational-cognitive constructs of task-induced involvement, best facilitated mastery of vocabulary. Tasks that differed in the construct of evaluation led to different results in learning vocabulary, which indicates that evaluation, a construct of task-induced involvement, is crucial to vocabulary learning. Based on our results, the optimal involvement in this study is moderate need, and strong evaluation. The two essential constructs of need and evaluation have different roles and effects on mastering vocabulary. Need is the precondition and guarantee to achieve the learning in a task. Evaluation, which means comparing and assessing the knowledge and use of the target words, leads to the final acquisition of target words.

The results of the current study also support the idea that word learning and retention are contingent on a focus-on-forms component. Of the three reading tasks, the task of reading plus writing activities was superior to reading only task, because word-focused activities were involved. These results were similar to those in Laufer's (2006) study, where practicing new words in two focused word activities (higher ILIs) yielded better results than reading texts or referring to a dictionary (lower ILIs). Taken together, successful vocabulary instruction should include word-focused activities that present a high degree of involvement.

Results also revealed that mere repetition of tasks are not necessarily effective for EFL vocabulary learning. Students tended to be unwilling to be engaged in repeated tasks, which suggested that learners' motivation is also a factor to be considered in designing and implementing tasks. This requires that the effectiveness of tasks were dependent on a high degree of task-induced involvement as well as

learners' high engagement. In addition to this, a balanced integration of input and output (e.g., reading input plus word-focused activities as in task 3) is suggested to be of paramount importance in achieving high performance in vocabulary learning, which aligns with Nation's (2008) four strands approach. In other words, form, input, output, and fluency should be attended to equally.

A final implication to be drawn from the present study is that deliberate rehearsal of newly learned words is necessary. In this study, even the task with the highest degree of involvement suffered a significant decrease in recalling the form after two months period, which indicates that systematic rehearsal of new words is essential for learning vocabulary. As stated in Nakata (2006), "How to distribute rehearsal opportunities affects the effectiveness and efficiency of the learning activities" (p.19). In his study, Nakata used the Low-First Method, an algorithm which was developed from cognitive psychology to realize the most optimal scheduling of rehearsal opportunities. Other studies have also clearly demonstrated the importance of rehearsal in vocabulary instruction (Folse, 2006; Keating, 2008). Armed with the above knowledge, teachers can curb unexpected declines in word retention by deliberate rehearsal of newly learned words.

Conclusion

The current study extends empirical support for the construct of task-induced involvement in EFL vocabulary growth by demonstrating that learners with limited vocabulary level benefit from more involved tasks. In other words, tasks with higher involvement load lead to greater gains. In addition, motivation is an important factor to consider when designing effective vocabulary learning tasks for EFL learners. Deliberate rehearsal of newly learned words is also necessary.

There are limitations in this study. The first limitation lies in the fact that the vocabulary test that was used was the same assessment tool all four times. Thus, the results may be influenced by test-retest effects.

Another limitation of this study was the fact that, although the results between task types are significantly different, statistically speaking, the differences are, in reality, somewhat small. It might be explained that the construct of search is not included in this study. The students were not required to look for the lexical meaning of the unknown target words. If the active role of search was added, it might better facilitate vocabulary outcome. In addition, if a self-imposed strong need was

added, the results might also be different. In conclusion, other possible involvement-heavy task types and combinations still need to be investigated. This means there is much room still for further research on ILH.

Acknowledgements:

We express our gratitude for the research funding provided by Nanning University (grant number: 2014JSGC10). Many thanks also go to the anonymous reviewers, who spent their valuable time reading the manuscript carefully and who generously provided very constructive comments.

References

Bamford, J. (1984). Extensive reading by means of graded readers. *Reading in a Foreign Language, 2*(2), 218-260.

Beck, I. L., & Mckeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 789-814). New York, NY: Longman.

Claridge, G. (2005). Simplification in graded readers: Measuring the authenticity of graded texts. *Reading in a Foreign Language, 17*(2), 144-158.

Cobb, T. (n.d.). The Compleat Lexical Tutor [Computer Software]. Retrieved from http://www.lextutor.ca/.

Cobb, T. (2008). What the reading rate research does not show: Response to McQuillan & Krashen. *Language Learning & Technology, 12* (1), 109-114. Retrieved from http://llt.msu.edu/vol12num1/cobb/default.html

Corson, D. (1995). *Using English Words.* Dordecht: Kluwer Academic.

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

Cunningham, A. M., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*, 9-15.

Davison, A. (1986). *Readability: The situation today.* (Reading Education Report No. 70). Champaign, IL: Center for the Study of Reading, University of Illinois.

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

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*, 1-44.

Elley, W.B. (1991). Acquiring literacy in a second language: the effect of book-based programs. *Language Learning, 41*, 375-411.

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

Hill, D. R., & Thomas, H. R. (1988). Survey reviews: Graded readers (part 1). *ELT Journal, 42* (1), 44-52.

Hill, D. R., & Thomas, H. R. (1988). Survey reviews: Graded readers (part 2). *ELT Journal, 42* (2), 124-136.

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

Huang, S., Eslami, Z., & Willson, V. (2012).The effects of task involvement load on L2 incidental vocabulary learning: A meta-analysis study. *The Modern Language Journal, 96*, 544-557.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning, 51*, 539-558.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*, 365-86.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition.* Oxford: Pergamon.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications.* New York: Longman

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review, 59*, 567–587.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *The Canadian Modern Language Review, 63*, 149–166.

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

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading, 31*, 273-284.

MeLeod, A. N., & McDade, H. L. (2011). Pre-schoolers' incidental learning of novel words during storybook reading. *Communication Disorders Quarterly, 32*, 256-266.

Min, H. Z. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning, 58*, 73-115.

Nagy, W., & Scott, J. (2000). Vocabulary processing. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269-284). Mahwah, NJ: Lawrence Erlbaum Associates.

Nakata, T. (2006). Implemented optimal spaced learning for English vocabulary learning: Towards improvement of the Low-First Method derived from the reactivation theory. *The JALT CALL Journal, 2*(2), 19-36.

Nation, I. S. P. (1993). Measuring readiness for simplified material: A test of the first 1,000 words of English. In M. L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 193–203). Singapore: RELC.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle.

Nation, I.S.P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language 12*, 355-380.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocation. *Language Teaching Research, 18*(1), 75-94.

Pichette, F. (2005). Time spent on reading and reading comprehension in second language learning. *The Canadian Modern Language Review, 62*, 243-262.

Stahl, S. A. (2003). How words are learned incrementally over multiple exposures. *American Educator, 27*, 18-19.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. London: Lawrence Erlbaum Associates Publishers.

Teng, F. (2014a, June). Vocabulary growth for low-proficiency students through reading graded readers. Paper presented at the 1st TRI-ELE International Conference, Bangkok, Thailand.

Teng, F. (2014b). Incidental vocabulary learning by assessing frequency of word occurrence in a graded reader: Love or Money. *LEARN Journal, 7*(2), 36-50.

Wallace, C. (1988). *Learning to read in a multicultural society: The social context of second language literacy.* New York: Prentice Hall.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*, 33–52.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*, 91-120.

About the Authors

Feng Teng is a lecturer with extensive experience of teaching reading courses in China. His research interests include EFL vocabulary teaching and learning through extensive reading.

Shu Zhang is a language teacher educator in Nanning University. Her research interests include reflective practice and language development.

Appendix 1

Sample test items (four items out of 40)

Directions: Please write down the target words by using the given Chinese translation and the English definition. The first letter of the English word has been given.

1. 消失(something that cannot be seen suddenly) d_____
2. 吼叫(make a low noise in its throat, usually because it is angry) g____
3. 疼痛(feel a steady, fairly strong pain) a____
4. 害怕的(anxious or afraid) f_____