An Experimental Study on the Effects of Different Reading Tasks

on L2 Vocabulary Acquisition

Jianping Xu

School of Foreign languages, Jiangsu University

Zhenjiang 212013, China

E-mail: xujp@ujs.edu.cn

Abstract

This empirical study was undertaken to test the Involvement Load Hypothesis (Laufer and Hulstijn, 2001) by examining the impact of three tasks on vocabulary acquisition. It was designed to test and develop the involvement load hypothesis by examining the impact of different reading tasks on the L2 vocabulary acquisition. The results show that reading tasks could facilitate L2 vocabulary acquisition. The hypothesis is basically supported, but it is expected that it will be further improved and needs some modifications. Furthermore, the results also indicate that using new words in contextualized communication is an efficient means to extend and consolidate learners' vocabulary acquisition.

Keywords: Vocabulary acquisition, Task involvement load, Tasks, Reading

1. Introduction

The study of vocabulary is at the heart of language teaching in terms of organization of syllabuses, the evaluation of learner performance, and the provision of acquisition resources (Candlin, 1988). Furthermore, vocabulary acquisition is crucial to students' traditional language skills: reading, writing, and listening. Without enough vocabulary, listening, reading comprehension, and writing are inefficient. Besides, "without grammar very little can be conveyed; without vocabulary nothing can be conveyed" (Wilson, 1986). So vocabulary is essential to language acquisition.

With enhancement of the status of vocabulary in language learning, research into vocabulary acquisition becomes a focus of research at present. Instructors and learners have always tried to find out ways in which instructional programs might best foster the acquisition of vocabulary. This study set out to examine the effect of reading-based tasks on vocabulary acquisition. Nearly 152 freshmen non-English majors from Jiangsu University participated in the study. Based on the results of vocabulary tests, this study aimed to find answers to the four questions surveyed in this study.

2. Literature Review

2.1 Vocabulary acquisition

There are different pairs of modes on vocabulary learning. In this thesis, we will use the term 'incidental vocabulary acquisition' discussed in Eysenck (1982) as one of our theoretical foundation. Incidental vocabulary learning in our research means that learners are required to finish a task involving the processing of some unfamiliar words without being told in advance that they will be tested afterwards on their recall of the meanings of those novel words. It is different from implicit vocabulary learning which holds that the meaning of a new word is acquired totally unconsciously as a result of abstraction from repeated exposure in a range of activated contexts. Implicit learning can be incidental only, but incidental vocabulary learning can include both implicit and explicit learning since "linking word form to word meaning is an explicit learning which holds that there is some benefit to vocabulary acquisition from the learner noticing novel vocabulary, selectively attending to it, and using a variety of strategies to try to infer its meaning from the context" (Ellis, 1994: 219). We also cannot say vocabulary learning here is an indirect learning since we have vocabulary exercises in our reading tasks including guessing words from context and using target words to make sentences which belong to vocabulary learning. The controlled experiments in the present study aim at investigating the effects of varying reading tasks on learners' vocabulary retention. Therefore, the term incidental learning is used as an opposing concept of intentional learning. The subjects in this study are required to read the passages with an intention to understand them and answer some comprehension questions but not with an intention to learn the target words. It is in this sense that learning of the target words is incidental.

Although the learners acquire vocabulary incidentally through reading, they also need to process the unfamiliar words in order to understand the contents of the passages. What do we know about the processes that facilitates vocabulary learning? Then another theoretical foundation of the current study is the depth of processing model which is launched by Craik and Lockhart (1972). However, some researchers (Baddeley, 1978; Eysenck, 1978, 1977) have challenged their levels of processing theory. The main points focus on the following two questions: (1) What exactly constitutes a

level of processing, and (2) How do we know that one level is deeper than another. In 2001, Laufer and Hulstijn present the Involvement load hypothesis which firstly adopts the measurable and operational factors (need, search, evaluation) to define the involvement loads which are used to judge the different degree of processing the unfamiliar vocabulary through reading. We have this empirical study designed exactly on the theoretical basis of the Involvement Load Hypothesis and use the measurable criteria of three components to define three different reading tasks.

2.2 The Involvement Load Hypothesis

Laufer & Hulstijn (2001) proposed the Involvement Load Hypothesis which was a motivational-cognitive construct of involvement, consisting of three basic components: need, search, and evaluation. Retention of unfamiliar words was claimed to be conditional upon the amount of involvement while processing these words. Involvement was operationalised by tasks designed to vary in the degree of need, search and evaluation.

The need component was the motivational, non-cognitive dimension of involvement. It was concerned with the need to achieve. This notion here was not interpreted in its negative sense, based on fear of failure, but in its positive sense based on a drive to comply with the task requirements which could be either externally imposed or self-imposed. Need was moderate when it was imposed by an external agent, e.g. the need to use a word in a sentence which the teacher has asked the learner to produce and need was strong when imposed on the learner by him-or herself. In the case of need, moderate and strong subsume different degrees of drive.

Search and evaluation were the two cognitive (information processing) dimensions of involvement, contingent upon noticing and deliberately allocating attention to the form-meaning relationship (Schmidt, 2001). Search was the attempt to find the meaning of unknown L2 word or trying to find the L2 word form expressing a concept by consulting a dictionary or another authority (e.g. a teacher).

Evaluation entailed a comparison of a given word with other words, a specific meaning of a word with its other meanings, or combining the word with other words in order to assess whether a word (i.e. a form-meaning pair) did or did not fit its context.

Each of the above three factors could be absent or present when processing a word in a naturally or artificially designed task. The combination of factors with their degrees of prominence constituted the involvement load, i.e., the three components involved in the tasks would be used to count the number of the involvement index which indicated the different degrees of involvement loads. Retention of unfamiliar words was claimed to be conditional upon the amount of involvement while processing these words (Laufer & Hulstijn, 2001).

3. Methodology

3.1 Research questions

The present study attempts to investigate the immediate and delayed effects of reading-based tasks on vocabulary acquisition as follows:

1. What are the overall immediate effects of different reading tasks on vocabulary acquisition?

a. What are the overall immediate effects of different tasks on vocabulary acquisition?

b. What are the immediate tasks effects on acquisition of different word knowledge types?

2. What are the delayed effects of different reading-based tasks on vocabulary acquisition?

a. What are the overall delayed effects of different tasks on vocabulary acquisition?

b. What are the delayed task effects on acquisition of different word knowledge types?

3. Can tasks contribute to vocabulary acquisition through reading by Chinese English learners?

4. With need and search controlled, does evaluation hold significant correlation with acquisition of the target words?

3.2 Subjects

The subjects were 152 freshmen who have been learning English as a second language from Jiangsu University. They were from three intact College English classes, of which two were at the high level and the other one class was at the low level. Placement at these levels was determined by the means of the English proficiency test that was administered upon students' entering the university.

3.3 Instruments

The instruments used in this study can be illustrated as follows

(1). Task 1. The reading material used for the study was a 930-word enjoyable, clearly organized article entitled "Why We Love Who We Love." The text was used in a pilot study with the students at similar levels. The findings from the pilot study showed the text as suitable in terms of content and difficulty level.

(2). Task 2. Three reading tasks were selected with different involvement loads to test their effects on vocabulary acquisition. Each task was randomly assigned to one of the three experimental groups. These tasks consisted of the multi-choice comprehension questions (Task M), blank-filling task (Task B) as well as sentence-making task (Task S).

(3). Task 3. To assess the immediate and delayed effects of the tasks on vocabulary acquisition, two vocabulary tests were administered: an immediate posttest and a delayed posttest. These tests were composed of supply-spelling, matching as well as select-definition.

(4).Task 4. The subjects were required to write a composition using the target words whose meaning had been glossed in the reading passages after each reading. But while writing the composition, they were not required to pay much attention to the grammar.

3.4 Data Collection

The present researcher scored the vocabulary tests after each task correct answer received one point, a semantically approximate explanation or translation received half a point, and a word that was not glossed (either in English or Chinese) or a blank received no points. The maximum grade a student could receive was 30 if all the words were correctly explained. If an answer was controversial in terms of the degree of the semantic approximation, opinions of the researcher's colleagues were sought for the scoring of this item.

Data collection is also from the qualitative study. The instrument involved in this part was group interviews. The interview with each group was conducted in the language lab. They were asked to reflect on the process in which they completed the tasks. And then, they were required to explain their performances in the vocabulary tests, that is, how they came up with the answers in the tests. And meanwhile, the subjects were expected to explain why they responded to the survey questions in a particular way in the questionnaire.

The procedure of the interview was conducted as follows: the interview was conducted in two sessions; one was at the end of the immediate posttest, the other at the end of the delayed posttest. For each session, the researcher interviewed the subjects individually. Chinese was used in the interviews so that the subjects could express their views freely and clearly. The interviews were audio-recorded and transcribed later for further analysis.

3.5 Data Analysis

(1). Scoring. Scoring is based on the matter of counting the correct answers on the reading- based tasks and vocabulary tests. The same scoring system was used for the pretest and posttest,

(2). One-way ANOVA. ANOVA was performed on the immediate posttest, the delayed posttest and responses to the questionnaire respectively.

(3). Paired-samples t-test. The paired-samples t-test was performed on the two vocabulary test scores achieved by each of the three groups.

(4). Qualitative data analysis. After the interview data were transcribed, the main points in the data were analyzed and summarized to help interpret the findings of the statistical analysis. The interviewees recalling process, for example, was analyzed to sort out the information about what word knowledge types the students paid attention to while performing the tasks and why they behaved in a particular way in the tests.

4. Results and Discussions

4.1 Immediate effects

This section consists of comparing the scores on the immediate posttest as a whole among the three groups as well as the scores on the part of the immediate posttest.

4.1.1 The overall immediate effects

To determine whether there was any overall difference among the treatment groups in the immediate posttest, the researcher performed one-way ANOVA by using the immediate posttest scores. Table 4.1 displays the results.

Insert Table 4.1 here!

The table shows that all the three groups manifested high levels of retention, varying from 55.87 to 73.65, which suggests that reading-based tasks did efficiently facilitate lexical learning. The retention rate, however, was significantly different for the three groups: F = 30.732, p = .000. Given the fact that the three groups had the same conditions except the tasks, we may attribute the marked difference to the tasks, which vary in involvement loads. In other words, task-induced involvement loads did have a significant immediate effect on vocabulary retention. Furthermore, a post hoc Scheffe test indicates that both Groups B and S scored significantly higher than Group M (p = .000 in either case, see Table 4.2) but did not differ remarkably from each other as expected; rather, the former scored slightly higher than the latter.

Insert Table 4.2 here!

These findings partially support the Involvement Load Hypothesis, which predicted that Tasks B and S which induced a higher involvement load than Task M would be more effective for vocabulary retention. The results also corroborate Hulstijn and Laufer's findings in their Hebrew-English experiment (2001), in which "reading plus fill in" and writing tasks outperformed the comprehension task in the acquisition of me target words. Furthermore, the findings also seem to support Swain's output hypothesis (Swain, 1985, 1995), given that the difference between Tasks B and S and Task M in this study was actually the one between pushed output and comprehension because the former required students to infer the word meanings and use them, whereas the latter involved only the understanding of the target words.

Retrospective interviews with the task performers and the questionnaire data also provided explanations for this phenomenon. When asked how they had processed the target words while performing the tasks, Task M performers reflected that they focused mainly on word meanings and even if they sometimes paid attention to other aspects of word knowledge such as word class and word form, the purpose was still to get some clues for the inference of meanings. One interviewee said,

I mostly thought about meanings while performing the task. I cared little about the word spelling, its part of speech and context. Even though sometimes I paid attention to these aspects, it was mainly for the sake of inferring lexical meanings.

Task B and Task S performers, however, reflected that in order to complete the tasks, they had to pay careful attention to many aspects of word knowledge such as meanings, word classes and collocations, as one interviewee reported,

To use the word, I should know its meaning. Besides; I also paid particular attention to how it was used in the passage such as its part of speech and the words with which it appeared together.

Clearly, Tasks S and B performers attended to more aspects of word knowledge than Task M performers. According to many linguists and psychologists, processing new lexical information more elaborately (e.g., by paying careful attention to the word's pronunciation, orthography, grammatical category, meaning, and semantic relations to other words) will lead to higher retention than processing lexical information less elaborately (e.g., by paying attention to only one or two of these dimensions). Accordingly, we may conclude that more elaborate processing reduced by Tasks S and B leads to their superiority in the immediate posttest.

However, out of our expectations, the results reveal no significant differences between Tasks S and B although the former induced a higher involvement load than the latter. On the contrary, Task B yielded slightly, higher retention than Task S. This finding runs counter to the Involvement Load Hypothesis and also contradicts those obtained by Hulstijn and Laufer (2001) who found remarkable differences between the tasks with moderate and strong evaluation. The reasons for this divergence can be various. One possible explanation is that the time control for the two tasks is different in the two studies. In this study, time on task was kept identical. In Hulstijn and Laufer's study, however, time on task varied. "Reading plus fill in" performers spent 50-55 minutes on their task whereas "composition writing" performers 70-80 minutes. Clearly, the latter spent much more time than the former and this may contribute to the obvious advantage of writing task over "reading plus fill in" task in their study.

Another possible interpretation is that the measures adopted to examine the task effect are different in the two investigations. In this study, the researcher investigated on three aspects of word knowledge to explore the task value for vocabulary retention. In their study, however, Hulstijn and Laufer only examined the task effect on one aspect of lexical knowledge, namely meaning. The difference in measures may also bring forth different results.

Last but not least, it is also possible that Task S performers did not approach the task in the way the researcher had expected. Instead of the anticipated mental effort exerted in integrating new information with acquired knowledge, some students just simply imitated the sentences in the passage without giving too much thought. One of the interviewees from Group S said:

Although I was not quite sure about the meanings of the words, it was not difficult for me to compose sentences. On the whole, I made sentences by imitating the example patterns in the original text. The target words and their collocations were also used in the similar way as in the passage. I just simply changed some other words in the given sentences.

4.1.2 Immediate effects on the retention of different word knowledge types

To further explore the immediate task effects on the retention of different word knowledge types, the scores on the three parts were displayed and compared among the three groups respectively. The results were summed up in Tables 4.3 and 4.4.

Spelling. Table 4.3 shows that in terms of spelling, Group B scored higher than Group S, which in turn, scored noticeably higher than Group M. The difference among the three groups reached a significant level (F = 54.882, p: .000), suggesting that the tasks had a great impact on the students' recall of word spellings. A post hoc Scheffe (see Table 4.4) further indicates that both Groups B and S outscored Group M significantly, but they did not differ markedly from each other, p =.094. This means that Task B was slightly more conductive than Task S in prompting spelling retention and

both of them were significantly more effective than Task M in this respect. These findings partially support the involvement Load Hypothesis.

Insert Table 4.3 and 4.4 here!

The obvious superiority of Groups B and S over Group M in spelling may be due to two reasons. First, the higher involvement load induced by Tasks B and S may possibly push the students to process the lexica] information with more mental efforts and this may facilitate the retention of word spellings. The follow-up interviews with Task B and S performers confirmed this speculation, as one interviewee explained:

I paid little attention to the spellings of the words because they were all listed on the exercise paper, what I paid attention to were, actually, the meanings of the words, their parts of speech and collocations. During the test, .however, I was surprised to find that I could retrieve the spellings of most of the words. This was mainly due to the painstaking efforts I had exerted on them and the deep impression they left on me. Consequently, I could spell out the words without much difficulty in the test.

Secondly, while Task M only required the students to make their choice from the given options, Tasks B and S provided the students a chance to write the words. Clearly, this may also contribute to the superiority of these two tasks in the spelling measure.

As to the question why Task B yielded higher retention in spelling than Task S although it induced a lower involvement load, the interviews with the students may provide the possible explanation. Some interviewees who performed Task B explained that in order to put the target words into the appropriate given contexts, they studied and compared these words again and again, thus having a deep impression of them. Task S performers, however, explained that after inferring the word meanings, they exerted much effort in making the decision about additional words that could combine with these new words in the original sentences, and hence less attention was paid to the forms of these new words. This being the reason, we may possibly conclude that Task B could facilitate the retention of spellings more efficiently than Task S.

Collocation. The task effect on the collocation retention patterned similarly to that on spelling retention with the exception of the advantage of Task S over Task B. ANOVA results again reveal that there was a marked difference among the three groups, indicating that the tasks also played a significantly different role in facilitating the retention of collocation. Also, the post hoc Scheffe again indicates that both Groups B and S outscored Group M significantly. However, no marked difference was found between Groups S and B although the former did slightly better than the latter. This means that of the three tasks, Task S was the most beneficial to developing collocation knowledge, Task M the least and Task B in between. Both Tasks S and B differed from Task M significantly in this respect.

These findings partially confirm the Involvement Load Hypothesis. They are also consistent with Swain's output hypothesis (1985, 1995), Given that Tasks B and S were both output tasks, whereas Task M was an input task. According to Swain, using the language, as opposed to simply comprehending the language, may force the learner to move from semantic processing to syntactic processing (1985: 249). Hence the advantage of Tasks B and S over Task M in the collocation measure may attribute to their ability to push the students to pay more attention to form (collocation, in this case).

Another aspect of the findings that may deserve due attention is that the contrast between Tasks B and S in the retention of collocation was not as acute as had been expected. One possible explanation is that Task S was not demanding enough to produce a superior result than Task B, as discussed earlier. An alternative interpretation is that Task B could also direct learners' attention to word collocations. As mentioned above, more than 60% of Task B students responded they had paid attention to collocations. Although this percentage was lower than that of Task S performer (72.2%), the difference was rather small (p = .362).

Meaning. A different picture emerges for the task effect on the retention of word meanings. In contrast to the Involvement Load Hypothesis, the current findings did not show any significant differences among the three groups in the meaning measure; rather the difference was quite small (F = .032, p = .969).

In trying to account for the discrepancy with Hulstijn and Laufer, several potential explanations present themselves. First, the contrast in findings may be due to time on task, as mentioned earlier. A second explanation could be the different measures used to assess the meaning retention. While Hulstijn and Laufer (2001) seemed to be testing for productive knowledge of the words by asking students to produce translations of the target items, the researcher was more interested in detecting the receptive retention of meanings by adopting the multiple-choice test. The third possible explanation could be that most of the students, whichever task they performed, processed the meaning aspect of the target words deeply because all of the tasks were mainly meaning-driven. This speculation is supported by the findings obtained from the questionnaire and interview data. The questionnaire results indicate that in each group, more than 92% of the students reported their attention to the lexical meanings.

To sum up, the findings partially support the Involvement Load Hypothesis, in that Tasks B and S yielded significantly higher retention than Task M in the overall immediate posttest as well as the spelling and collocation measures but they did not differ significantly. Furthermore, the three tasks showed no marked differences in their immediate effects on meaning retention.

4.2 Delayed effects

This section will report and discuss the findings of the delayed effects of the tasks on vocabulary retention.

4.2.1 The overall delayed effects

To investigate whether there was any overall difference among the three groups in the delayed posttest, one-way ANOVA was performed using the delayed post-test scores.

Insert Table 4.5 and 4.6 here!

Table 4.5 showed that Group S scored the highest in the delayed posttest, Group M the lowest and Group B in between. The difference among them had reached a significant level (F = 6.277, p = .002), indicating that the tasks still had a great influence on vocabulary retention in spite of time. The post hoc Scheffe (see Table 4.6) further reveals a marked difference between Groups M and S (54.315 vs. 62.833, p= .002). However, no significant difference was observed between Groups M and B or between Groups B and S. This means that of the three tasks, Task S was the most effective in facilitating long-term retention and its effectiveness was considerably superior to that of Task M whereas Task B was more conductive than Task M but not significantly conductive.

These findings only support the Involvement Load Hypothesis to a limited degree. That is, Task S still kept its superiority over Task M as time went by, suggesting that Task S could not only help the students to produce more words immediately after the treatment, but also allow them to store more of these words in their long-term memory. This result is also consistent with that obtained by Hulstijn and Laufer in their two parallel experiments (2001).

However, contrary to expectations, Task B lost its obvious advantage over Task M in the delayed posttest (58.519 vs. 54.315, p = .202). We may explain this phenomenon from the perspective of generative model (Slamecka & Graf, 1978). Task B performers, unlike their Task S counterparts, were not required to generate. That is, they were not asked to use the target words in original contexts; rather they reacted to experimenter-provided stimuli, merely recognizing the differences among the words and put them into the given contexts. Probably, this kind of learning would efficiently facilitate immediate word gain. However, its positive effect would drop dramatically over time.

4.2.2 Delayed effects on the retention of different word knowledge types

Tables 4.7 and 4.8 sum up the task's effect on the students' performances in different parts of the delayed posttest.

Insert Table 4.7 and 4.8 here!

Spelling. With regard to word spellings, Group S scored the highest, Group B lower and Group M the lowest. The differences among them reached a statistically significant level (F = 9.233, p = .000), implying that such differences were not due to chance. A post hoc Scheffe test (see Table 4.8) shows that Group S differed from Group M significantly (14.042 vs. 9.241, p = .000). No marked difference, however, existed between Groups B and M, or between Groups B and S. These results mean that Task S was the most beneficial to long-term retention of word spellings whereas Task B failed to sustain its significant superiority over Task M

Collocation. The delayed task effects on collocation retention resembled those on spelling retention. Specifically, there was a significant task effect on collocation, measure (F = 5.159, p = .006). Again, the post hoc Scheffe indicates that Task S performers outperformed Task M performers significantly (18.208 vs. 14.482, p = .006), Still, no marked difference was found between Tasks B and M or between Tasks B and S. Clearly, Task S again proved the most effective in promoting collocation retention. The questionnaire results reflected that the students also held the most positive attitudes towards the effectiveness of Task S in collocation measure

Meaning. A different picture appears in the case of the delayed task effects or meaning retention. No significant difference was found among tile three groups; instead most of the students demonstrated a high level of retention in recognizing word meanings, which implies that the three tasks had similar delayed effects on the receptive retention of word meanings.

Generalizing from the above results, we may conclude that in terms of the delayed task effects on vocabulary retention, this study only provided limited support or the Involvement Load Hypothesis. That is, Task S still enjoyed its significant superiority over Task M in promoting the overall retention and retention of word spellings and collocations one week later. However, Task B did not yield significantly higher retention than Task M as predicted. No marked difference existed between Tasks S and B either.

4.3 Different tasks contributing to vocabulary acquisition through reading

Laufer and Hulstijn (2001) hold that task with a higher involvement load will be more effective than task with a lower involvement load in terms of vocabulary retention. In the study, it was also predicted that if other factors being equal, tasks with a higher involvement load will be more effective for vocabulary acquisition than tasks with lower involvement load. The aim is to test whether this assumption can apply to Chinese learners of English or not.

Insert Table 4.9 here!

From Table 4.9, we can see that in the immediate test, the highest mean score among these four tasks is that of the reading and composition group, which was 16.32. So in the immediate test, the performance in the reading and composition group was higher than that in the reading and filling group and reading and guessing group, which, in turn, was higher than that in the reading and comprehension group. And a significant task effect between groups (F=15.615, p=.000< .05) was obtained. The results proved that in the immediate test Task 4 (reading and composition) with higher involvement load promoted better word acquisition than Task 1, 2, and 3.

In the same way, Table 4. 9 shows that in the delayed test, the highest mean score is task 4. And there was also a significant group difference (F=16.345, p=.000<.05). The results support Involvement Load Hypothesis that tasks with a higher involvement load will be more effective than tasks with a lower involvement load in terms of vocabulary retention.

Task 2 and Task 3 have the same involvement load index, but do they equal in vocabulary retention? Or is one task superior to the other in the immediate test or in the delayed test? To see whether the difference between mean retention scores of Task 2 and Task 3 was statistically significant, the mean scores in the two experiments in the immediate test and delayed test were then submitted to a t-test for Independent Samples (shown in Table 4. 10).

Insert Table 4.10 here!

The results revealed in the immediate test showed that the difference between Task 2 and Task 3 was significant (p=.014<.05). It is unexpected that Task 2 had better acquisition of vocabulary in the immediate test. In the delayed test, there was no statistic significance between them. Drawing on the mean retention scores in Table 4-9, it appears that the group performing Task 2 got significantly better scores than the group doing Task 3. But in the delayed test, although the mean retention score of Task 4 (9.49) is higher than that of Task 3 (8.3), the difference was not statistically significant. The reason might lie in that the participants in the reading and guessing meaning group just guess meaning of the thirty target words when they did the task, but after collecting back the materials, they didn't pay attention to check the correct meaning of the target words when the teacher delivered the translation list of the target words

Therefore, tasks with higher involvement load generally but not necessarily lead to better retention. Task 4 with the highest involvement load resulted in the best retention result. As for Task 2 and Task 3 with the same involvement load, although there is statistically significant difference between them in the immediate test, the overwhelming gains of Task 2 disappeared in the delayed test.

4.4 Different tasks having different effects on vocabulary acquisition

It was predicted that in the same task, with need and search controlled, tasks with higher evaluation will produce better retention than those with lower evaluation.

From Table 4-9, we know that L2 learners could gain the knowledge of the target words by incidental learning and reading tasks could facilitate vocabulary acquisition. However, different tasks have different effects on vocabulary acquisition.

First, in order to determine whether there was statistically significant effect of each factor, the mean retention scores of the immediate test and the delayed test were then submitted to One-way analysis of variance (ANOVA) respectively. The ANOVA results of the immediate posttest are listed in Table 4-11 and the ANOVA results of the delayed posttest in Table 4-12.

Insert Table 4.11, 4.12 and 4.13 here!

The figures in Table 4. 6 show that in the immediate test the mean difference of Task 2 (reading and composition) and Task 4 (reading and blank-filling group) is 2.13, the difference is insignificant (p=.949>.05). But both the mean differences between Task 1 and Task 2, Task 1 and Task 4 are 8.35, 6.22 respectively and the significance levels are 0.000 (p<.05), which indicates that there is significant difference between Task 1 and Task 2, and Task 1 and Task 4. From the post hoc test for task difference in the immediate test, we know that the mean score of the group performing Task 2 is superior to that of Task 1 and Task 3, which means participants who complete Task 2 get better retention scores than students who finish Task 1 or Task 3 in the target words retention check immediately after the experiment.

In terms of the delayed test, the mean differences between Task 1 and the other three tasks are significant. The mean differences between Task 2 and Task 4 are not statistically significant, but they still have some differences in mean

scores. Since Task 2 is with higher evaluation than Task 3 and Task 4, from Table 4-9 we can find that in the immediate test, the mean retention score of Task 2 (mean=17.32) is higher than Task 3 (mean=12.14) and Task 4 (mean=15.19), while in the delayed test, the similar result can also be found, that is, the mean retention score of Task 2 was still superior to the other three tasks. Therefore, both the results in the immediate and delayed test proved the fourth prediction that Task 2 (reading and composition) with higher evaluation would yield the highest retention scores, which also support the Involvement Load Hypothesis.

5. Conclusion

Based on the above results and discussions, the following findings emerge from the present investigation:

1)   As far as the immediate task effects on vocabulary acquisition are concerned, the results partially support the Involvement Load Hypothesis. That is, Task B and S, which induce higher involvement load than Task M, yield significantly higher acquisition in the overall immediate posttest as well as the spelling and collocation measures. However, Task S does not produce acquisition significantly superior to Task B as predicted; rather, the latter enjoys a slight advantage over the former in overall acquisition, especially the acquisition of word spellings. Furthermore, the three tasks do not differ markedly in producing the receptive acquisition of word meanings.

2)   In terms of the delayed task effects on vocabulary acquisition, the results support the hypothesis only to a limited degree. As is expected, Task S has greater effects than Task B, which in turn, has superior effects to Task M. the difference between Task S and M has reached a statistically significant level. However, no measurable difference exists between Tasks S and B, or between Task B and M.

3)   Time has a great impact on vocabulary acquisition. That is, whichever task the students perform, they generally show a significant decrease in word knowledge with the exception of meaning over one week. In addition, the effect of Task B proves to be subject to diminution.

4)   Students' English proficiency only influences the immediate task effects on vocabulary acquisition. As for high proficiency students, Task B produces significantly higher acquisition than Task S in the immediate posttest whereas these two conditions do not differ markedly for the low proficiency counterparts

In light of the findings of the present study, we may find some useful implications for vocabulary teaching and learning in China.

First, the results of this study suggest that teachers should design a variety of reading-based tasks that can induce the need for the attention to target words to develop learners' vocabulary knowledge.

Secondly, teachers could design or select tasks varying in involvement load for different words depending on the type of reinforcement they want to provide.

Third, due to the significant time effect on vocabulary acquisition as revealed by the current study, teachers need to provide opportunities for students to practice the vocabulary they have learnt so as to help them to better anchor the words in memory.

Fourth, the findings of this study also suggest that writing with new words could serve as an efficient means to extend and consolidate learners' vocabulary.

Finally, it would be highly desirable to communicate the findings of this study to Chinese learners as well, so that they will be aware of the effect of task-induced involvement loads on vocabulary acquisition and thus being able to make a better decision as to what kind of tasks they select to meet their individual needs of lexical learning.

References

Coady, J. (1997). *L2 Vocabulary Acquisition through Extensive Reading*. In J. Coady and T. Huckin (Eds.), *Second Language Vocabulary Acquisition A Rationale for Pedagogy*. Cambridge: Cambridge University Press, 174-200.

Ellis, N. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition,* 21, 285-301

Gass, S. & Selinker, L. (2001). *Second Language Acquisition-An Introductory Course*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Goulden et al. (1990). How large can be a receptive vocabulary be?. *Applied Linguistics,* (11):341-363.

Henriksen, B. (1996). Semantisation, Retention and accessibility: Key concepts in vocabulary acquisition. *Paper presented at the AILA Congress*, Jyvaskyla, Finland.

Huckin,T.& Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition,* (21):181-183.

Hulstiin, J. & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition, *Language Learning*.

Hulstijn, J. (2003). *Incidental and Intentional Learning .In The Handbook of Second Language Acquisition*. Maiden, MA: Balckwell. 349-381.

Knight, S. (1994). Dictionary use while reading: the effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal.* (78): 285-299.41

Laufer, B. (2001). Reading, word–focused activities and IVA in a second language. *Prospect,* 16(3):44-54.

Muncie. (2002). *Process writing and vocabulary development: Comparing lexical frequency profile across drafts.* System, 30, 225-235

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: a matter of topic familiarity? *Language Learning,* 54(3): 469-523.

Richards, J. (1976). The role of vocabulary teaching in the English syllabus. *TESOL Quarterly*, l0(l): 77-89.

Schmidt, R. (2001). *Cognition and Second Language Instruction.*Cambridge: Cambridge university press.

Wesche, M.B. & Paribakht, T.S. (Guest Eds.). (1999). Incidental L2 Vocabulary Acquisition: Theory Current Research and Instructional Implications. Special Issue: *Studies in Second Language Acquisition,* 21(2):175-335.

Wilson, P.T. & Anderson, R.C. (1986). *What they don't know will hurt them: the role of prior knowledge in comprehension*. In J.Orasanu (Eds.), *Reading comprehension: from research to practice*[C]. Hillsdale, NJ: Lawrence Erlbaum Associates

Huang Linhui. (2007). *The effects of tasks with different involvement loads on incidental vocabulary acquisition through reading*. Guangxi Normal University.

Table 4.1 The overall immediate task effects on vocabulary retention

Group	N	Mean	SD	F	P
M	50	55.87	14.27		
B	50	73.65	10.64	30.732	.000
S	52	71.09	10.02		

Note. M = multiple-choice questions; B = blank-filling; S = sentence-making;

N = number of students; total possible score for the test = 90.

Table 4.2 Scheffe post hoc comparisons for immediate task effects on vocabulary retention

Group	dif.	P
M vs. B	-13.74	.000
M vs. S	-12.16	.000
M vs. S	1.60	.609

Note. dif. = mean difference between the groups.

Table 4.3 Task effects on different parts of the immediate posttest

Immediate Posttest	Group	Mean	SD	F	P
Supply Spelling (Max. = 33)	M	13.22	7.75		
	B	24.51	6.47	54.882	.000
	S	21.97	7.17		
Matching (Max. = 24)	M	16.59	6.20		
	B	20.29	5.23	16.464	.000
	S	21.08	3.83		
Select Definition (Max. = 33)	M	30.15	4.52		
	B	30.00	3.31	.032	.969
	S	30.04	3.58		

Note. M = multiple-choice questions; B = blank-filling; S = sentence-making; Max. = maximum score.

Table 4.4 Scheffe post hoc comparisons for immediate task effects on the retention of spellings and collocations

Immediate Posttest	Group	dif.	P
Supply Spelling	M vs. B	-11.28	.000
	M vs. S	-8.74	.000
	B vs. S	2.54	.094
Matching	M vs. B	-3.70	.000
	M vs. S	-4.49	.000
	B vs. S	-.795	.645

Note dif. = mean difference between tile groups.

English Language Teaching

Table 4.5 The overall delayed task effects on vocabulary retention

Group	N	Mean	SD	F	P
M	81	54.32	16.16		
B	80	58.52	14.46	6.277	.002
S	72	62.83	13.68		

Note. M = multiple-choice questions; B = blank-filling; S = sentence-making;

N = number of students; total possible score for the test = 90.

Table 4.6 Scheffe post hoc comparisons for delayed task effects on vocabulary retention

Group	dif.	P
M vs. B	-4.21	.202
M vs. S	-8.52	.002
B vs. S	-4.32	.204

Note. dif. = mean difference between the groups.

Table 4.7 Task effects on different parts of the delayed posttest

Delayed Posttest	Group	Mean	SD	F	P
Supply-Spelling (max.= 33)	M	9.24	6.61		
	B	11.76	6.87	9.233	.000
	S	14.04	7.29		
Matching (max.= 24)	M	14.48	7.76		
	B	16.20	7.41	5.159	.006
	S	18.21	6.11		
Select-Definition (max. = 33)	M	30.59	4.31		
	B	30.56	2.95	.001	.999
	S	30.58	3.65		

Note. M = multiple-choice questions; B = blank-filling; S = sentence-making.

Table 4.8 Scheffe post hoc comparisons for the delayed task effects on retention of spellings and collocations

Delayed posttest	Group	dif.	
Supply Spelling	M vs. B	-2.52	.072
	M vs. S	-4.80	.000
	B vs. S	-2.29	.128
Matching	M vs. B	-1.72	.316
	M vs. S	-3.73	.006
	B vs. S	-2.01	.228

Note. dif. = mean difference between the groups

Table 4.9. Descriptive statistics for scores of the four treatments in immediate post-test and delayed post-test

Test		N	Mean	Std. Deviation	F(between groups)	Sig.(between groups)
Immediate test	Task 1		8.99	5.51		
	Task 2		15.20	5.60	15.615	.000
	Task 3		12.12	4.78		
	Task 4		16.32	5.80		
Delayed test	Task 1		4.91	2.55		
	Task 2		9.49	4.00	16.345	.000
	Task 3		8.30	3.96		
	Task 4		11.00	4.33		

Notes: the mean difference is significant at the .05 level.

Table 4.10. Independent Samples Test (comparing the mean retention scores of Task 2 with that in Task 3 in the immediate test)

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	Df	Sig. (2-taile)	Mean Difference	Std. Error Difference
The IM	Equal variances assumed	1.153	.286	-2.524	72	.014	-3.05	1.21
The DT	Equal variances not assumed			-2.524	70.28	.014	-3.05	1.21
	Equal variances assumed	.005	.973	-1.285	72	.203	-1.19	.93
	Equal variances not assumed			-1.285	71.9	.203	-1.19	.93

Notes: the IM=the immediate test, the DT=the delayed test.

Table 4-11. One-way ANOVA on the retention scores of the immediate test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1387.434	3	462.478	15.716	.000
Within Groups	4090.413	139	29.427		
Total	5477.846	142			

Table 4-12. One-way ANOVA on the retention scores of the delayed test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	699.109	3	233.036	16.345	.000
Within Groups	1981.716	139	14.257		
Total	2680.825	142			

To see the task effect on the retention scores among groups, a post-hoc-test for the one-way ANOVA was performed. The results are presented in Table 4.13.

Table 4-13. Multiple Comparisons of the mean scores of the four tasks in the immediate test and the delayed test

Dependent variable	(I)Task	(J)Task	Mean difference (I-J)	Std. Error	Sig.
Immediate test	Task 1	Task 2	-8.35*	1.31	.000
		Task 3	-3.16	1.28	.064
		Task 4	-6.22*	1.28	.000
	Task 2	Task 1	8.35	1.31	.000
		Task 3	5.19	1.29	.000
		Task 4	2.13	1.29	.347
	Task 3	Task 1	3.16	1.28	.064
		Task 2	-5.19	1.29	.000
		Task 4	-3.05	1.26	.073
	Task 4	Task 1	6.22	1.28	.000
		Task 2	-2.13	1.29	.347
		Task 3	3.05	1.26	.073
Delayed test	Task 1	Task 2	-6.09	.91	.000
		Task 3	-3.38	.89	.001
		Task 4	-4.57	.89	.000
	Task 2	Task 1	6.09	.91	.000
		Task 3	2.70	.90	.014
		Task 4	1.51	.90	.330
	Task 3	Task 1	3.38	.89	.001
		Task 2	-2.70	.90	.014
		Task 4	-1.19	88	.528
	Task 4	Task 1	4.57	.89	.000
		Task 2	-1.51	.90	.330
		Task 3	1.19	.88	.528

Notes: * the mean difference is significant at the .05 level.