

Effect of Task-induced Online Learning Behavior on Incidental Vocabulary Acquisition by Chinese Learners—Revisiting Involvement Load Hypothesis

Jiehui Li

Guangzhou Civil Aviation College, Guangzhou, China

Abstract—In the present study, Laufer & Hulstijn's Involvement Load Hypothesis was tested by a different methodology from that adopted by the previous similar studies. That is, instead of focusing only on the product of learning, the present study also attended to the details of learners' task-induced online learning behavior via a specially designed computer program. Eighty-one participants were randomly assigned to one of the four tasks with different amount of involvement load. Once completing the task, the participants were unexpectedly tested on the retention of the target words that appeared in the texts. Two weeks later they were given two delayed posttests. The data were analyzed both quantitatively and qualitatively. The results suggested that different tasks did elicit different patterns of on-line learning behavior in terms of frequency of look-ups and amount of time spent on target words. It was also found that tasks assumed with higher involvement load hypothesis did not necessarily lead to higher retention scores. It was concluded that the criteria that Laufer & Hulstijn have proposed for grading the task effectiveness may be too simplified and idealized, failing to reveal the sophisticated nature of the cognitive processes that the different task induces.

Index Terms—Involvement Load Hypothesis, online learning behavior, task type, incidental vocabulary acquisition, computer program

I. INTRODUCTION

A. The Rationale of the Study

Research on vocabulary acquisition in general, and on incidental vocabulary acquisition in particular, has increased considerably over the past few years. Aspects which form the focus of research include the resources and procedures applied in meaning inference (Haastrup 1991; Van Parreren & Shouten-van Parrenen 1981), the effect of dictionary use or glossing support (Cho & Krashen 1994; Hulstijn et al. 1996; Knight 1994; Luppescu & Day 1993; Wantanabe, 1997), the influence of different reading tasks and the effects of vocabulary learning (Hulstijn, Hollander, & Greidanus, 1996; Hulstijn & Trompetter, 1998; Wesche & Paribakht, 2000; Newton, 2001).

In explaining the superior effect of one approach or task over another, most researchers suggest that the superior tasks, that is, tasks that lead to better retention rate in incidental vocabulary acquisition, required a deeper level of processing of the new words than other tasks. Psychologists and language acquisition scholars working within the framework of cognitive psychology agree that the more a learner pays attention to a word's morphophonological, orthographic, prosodic, semantic, and pragmatic features and to intraword and interword relations, the more likely it is that the new lexical information will be retained (Hulstijn, 2001). Such close attention to word features, which is often associated with performing a vocabulary task, has also been referred to as deep processing, elaboration, and cognitive effort. Yet in a normal language learning situation, it is hard to decide whether one instructional task requires deeper processing than another. Hence research on task effectiveness would require the identification of criteria which could be observed, manipulated, and measured. The Involvement Load Hypothesis (ILH) proposed by Laufer and Hulstijn (2001) is deemed as the first attempt to operationalize traditional general labels such as noticing, attention, elaboration into task-specific components. Thus tasks may be graded for the processing depth they demand.

Revealing as the Involvement Load Hypothesis is, the empirical evidence for this hypothesis has been rare. Hulstijn and Laufer (2001) conducted two parallel experiments testing whether retention of vocabulary acquired incidentally is contingent on amount of task-induced involvement. They compared three learning tasks with varying involvement loads: reading comprehension with marginal glosses (index=1), reading comprehension plus fill-in (index=2), writing composition and using the target words (index=3). It turned out that the results of the Hebrew-English experiment fully support the hypothesis; the results of Dutch-English experiment, however, only partially support the Hypothesis.

As an attempt to test the Involvement Load Hypothesis in the foreign language classroom setting in China, Huang (2003) set out to investigate the effect of task types on L2 incidental vocabulary learning. On the basis of the hypothesis, it was predicted that tasks with higher involvement load should produce higher retention rate than those with lower

involvement load. However the results turned out to be very unexpected and mixed. Statistically significant difference of scores between the tasks was found only in the Immediate Posttest, and only between Task 6 and the other tasks; there was no significant difference between the other five tasks.

Keating (2008) also used three tasks with different involvement loads to assess the predictive nature of the hypothesis. Based on ILH, it was predicted that Task 3 (writing original sentences using target words) would outperform Task 2 (reading comprehension plus fill-in), which in turn would do better than Task 1 (reading passage with marginal glosses). The results strongly supported the hypothesis that the involvement load hypothesis can be generalized to low-proficiency learners, though no significant difference was found between the groups on Task 3 and Task 2 about their passive knowledge of the target words. Kim (2008) also provided empirical evidence for the hypothesis. The results of immediate post-test partially supported the hypothesis while those of delayed post-test fully supported this hypothesis. He also investigated whether tasks with equal involvement load would lead to equivalent initial and later retention of words by adult ESL learners at two different levels of proficiency. In line with other studies, the results showed that a higher involvement index leads to more effective initial and delayed vocabulary learning.

To recapitulate, the Involvement Load Hypothesis was basically verified in previous studies (Laufer & Hulstijn, 2001; Keating 2008; Kim 2008), but it was only marginally supported in Huang (2003). The reasons for inconclusive and contradictory findings appear to lie in the following two interrelated major weaknesses of the above-mentioned studies. First, in all of the previous studies, the researchers assumed that the participants would do what the tasks required them to do without observing what the learners were actually doing. In other words, they only examined the learning products without giving any consideration to how they were related to the learning process. Secondly, since no observation was made of the learners' behavior when performing language tasks, it is very difficult to tell whether it is the task-induced involvement that leads to better retention of the target words or it is the time-on-target words that leads to better learning. This is because an important question related to task effectiveness is whether it is the nature of the task that induces effective learning, or the time learners spend on it (Hill & Laufer, 2003). Apparently the major drawback of the previous studies is their failure to observe the extent to which learners' behaviors were in accordance with the amount of cognitive load inherent in task as specified by the Involvement Load Hypothesis.

In recent years considerable attention has been given to vocabulary learning in Computer-assisted language learning. Some L2 reading specialists suggest adopting user-behavior tracking technology, because a tracking feature allows the researchers to explore learners' look-up behaviors and simultaneously examine the process and final product involved in learning vocabulary. (e.g., Collentine, 2000; Hulstijn, 2000). To overcome the limit of the previous studies special computer programs can be designed whereby not only the final learning products but the process involved in incidental vocabulary learning as well can be examined. More specifically, such computer programs would enable the recording of every action by learners when interacting with the available glossed entries.

The solution to the above problem is of great significance for foreign language vocabulary teaching. Theoretically, it makes an empirical contribution to the long-contested issue of task efficiency with respect to incidental vocabulary learning. Pedagogically, it is hoped that the findings of the present study may assist the teachers and learners alike in developing tasks that foster vocabulary learning in an incidental setting.

B. Theoretical Framework: Involvement Load Hypothesis

Laufer & Hulstijn (2001) put forward the Involvement Load Hypothesis, in which they proposed a motivational-cognitive construct of involvement, consisting of three basic components: need, search, and evaluation.

The *need* component is concerned with the need to achieve. Need is moderate when it is 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. Need is strong when imposed on the learner by him or herself. *Search* is the attempt to find the meaning of a word by consulting a dictionary or another authority. *Evaluation* entails 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 does or does not fit its context. If the evaluation entails recognizing differences between words, or differences between several senses of a word in a given context, we will refer to this kind of evaluation as "moderate". If, on the other hand, evaluation requires making a decision about additional words which will combine with the new word in an original sentence or text, we will refer to it as "strong" evaluation.

Vocabulary learning tasks are therefore described in terms of "involvement load" rather than as being input or output type tasks. Greater involvement occurs when there is a high degree of need, a search for answers, and evaluation of the word and context. In any given task, these factors can be present or absent (+/-), moderate (+), or strong (++). The basic contention of the involvement load hypothesis is that retention of unfamiliar words is claimed to be conditional upon the amount of involvement while processing these words. Words which are processed with higher involvement load will be retained better than words which are processed with lower involvement load.

C. Key Research Questions

The problem that the present study attempts to solve is concerned with the effect of task-induced online learning behavior on incidental vocabulary acquisition. A further scrutiny of this problem reveals that its solution is dependent on the answers to the following three related key questions:

- (1) Do different task types induce different online learning behavior?

(2) Do different online learning behaviors lead to different learning results in terms of retention rates of newly learned target words?

(3) To what extent can the observed online learning behavior and learning results be accounted for within theoretical framework of Involvement Load Hypothesis?

II. RESEARCH DESIGN

A. Participants

The participants were 81 adult Chinese EFL second-year non-English majors from Guangdong University of Foreign Studies. To make sure that these participants were similar in English proficiency level, English Language Proficiency Assessment (Hever, 2001, downloaded from internet¹) was given. The means of the four groups ranged between 47.19 and 51.22, indicating that the participants had an average vocabulary size of around 3000 words. One-way ANOVA was performed on the means and no significant difference was found between the groups: $F(4, 81) = 0.453$ ($p = 0.716$). To verify the participants' unfamiliarity to the selected target words, a pretest in the form of a checklist was developed. Learners received a list of 50 lexical items (9 target words and 41 distracters) and were instructed to attempt a translation of each of the words and skip those items that they had never encountered before. The results suggested that the mean score of the four groups ranged between 0.667- 0.773, which means that on average the participants knew less than one target word. One-way ANOVA was then conducted on the means of the four groups and it was found there was no significant difference between them. $F(3, 81) = 0.085$ ($p = 0.968$).

Of the original cohort, the data from 14 subjects were excluded for two reasons: (1) exhibiting knowledge of more than two target words, and (2) being absent either in the immediate posttest or in the delayed posttest. The final sample size was eighty-one.

B. Instrumentation

Tasks

Four tasks were designed on the basis of the notion of task-induced involvement.

Task 1: Read and answer irrelevant comprehension questions. In this task, the participants were required to read a passage on the computer screen, and then answer the comprehension questions (true/false questions). The questions were designed in such a way that they could be answered without reference to the highlighted target words. This task, according to Involvement Load Hypothesis had a total involvement load of zero.

Task 2: Read and answer relevant comprehension questions. This task involved reading the same passage on the computer screen, but the true/false questions were designed in a way that they could not possibly be answered correctly without referring to the target words. This task had a total involvement load of 2 (+ n, + s, - v).

Task 3: Read and fill in missing words (cloze). The participants read the same passages as those in Task 2, but the target words to be learned were deliberately left out from the text. They were then required to fill in the missing words by making correct choices from a list of 14 words, five of which served as distracters. After that they also had to answer the same true/false comprehension questions as those in Task 2. The task had an involvement load of 3 (+ n, + s, + v).

Task 4: Read and make sentences. The participants read the same passages as those in the other three tasks. They are required to write original English sentences using the target words whose meaning had been used in the reading passages. The involvement index in this task was 4 (+ need, + search, ++ evaluation).

The involvement loads of the four tasks were summarized in Table I below.

TABLE I.
INVOLVEMENT LOADS OF THE TASKS USED IN THE PRESENT STUDY

Task	Components of Involvement			Amount Index
	need	search	evaluation	
Reading & irrelevant questions	-	-	-	0
Reading & relevant questions	+	+	-	2
Reading & cloze	+	+	+	3
Reading & sentences making	+	+	++	4

Material

Two reading passages were taken from a collection of College English Test of Band 4 (CET-4) examination papers (Liuxiang, January, 1998). The passages were of an appropriate level of difficulty with a lexical density that would allow general comprehension. Hu and Nation (2000) suggested that in intensive reading of short passages, less than 95% coverage may be suitable for developing language and the use of reading strategies. The passages in the study therefore included 94% of words (which were later verified) that were familiar to the participants. The text was so modified that the number of occurrence for each target word to be learned would appear only once.

¹ The net address is www.educ.goteborg.se/usam/pforum/elpa/ar00.1.html

The selection of the target words to be learned from the reading texts was based on the following criteria. (1) Each target word was supposed to have more than three different senses so that we can investigate whether the participants evaluate the different meanings against the context when performing the assigned tasks. (2) No synonyms or explicit definitions of the target words appeared in the reading passages, and the context of the reading passage should not provide sufficient clues for learners to infer the words' meanings. (3) It was unlikely that the participants had learned these target words. Nine words and expressions were selected for investigation. They were: prodigy, distinct, nurture, discipline, significant, shrinking, sophisticated, lag behind, and provincial.

C. Data Collection and Investigative Procedure

The collection of data was conducted over a period of 4 weeks.

(1) In Week 1, administer the vocabulary size test to make sure that the participants were similar in proficiency level. In addition, a vocabulary checklist test was administered to make sure that the target words were unfamiliar to all participants. Participants were also told not to use dictionaries or any other reference books in the test. Nor were they allowed to write down the words in their notebook. This session lasted about 40 minutes.

(2) In the second week participants were randomly assigned to different tasks designed on the principles of Involvement Load Hypothesis, and their on-line performance of the tasks were recorded by using specially designed computer program. Once the task was completed, the participants were asked to do the immediate posttest in paper-and-pencil form. In the immediate test there were 18 words altogether, among which 9 words were target words and 9 words served as distracters. The participants were supposed to write the Chinese meaning of each target word. This session took about 45 minutes.

(3) In the fourth week, delayed posttest test 1 and delayed posttest test 2 were administered to the same participants during the regular class hours without any previous notice. Dictionary use was not allowed. It took about 10 minutes. The delayed posttest 1 and the immediate test took the same format. The delayed posttest 2 differed from the other two tests in that the target words appeared in sentences instead of being isolated. Participants were required to write the Chinese meaning of each underlined word.

D. The Computer Program

The computer program was so designed that once the participants logged in by name, they would be randomly assigned to one of the four tasks. Whatever task the participants are assigned, the screen would display the text in which the target words are underlined and highlighted in white (the text itself was in blue). Unfamiliar words can be looked up by clicking on them with the mouse. The participants can return to any word any time while reading the text for further information. However, the instructions for the participants varied with the nature of the task they are assigned to. Specifically, for the participants who are assigned to Task 1 and Task 2, the following instruction appears on the screen:

reading the text **reading the question** — The participants can click either of them according to their own reading habits.

return to reading passages — The participants can click this column and go back to the passages if they are not sure about their answer to the comprehension questions.

done — The participants click on this column when they finish the comprehension questions of the first reading passage.

go on — Click on this column to go on reading the second passage.

exit — The participants click on this when they finish the task

For the participants who are assigned to Task 3, the following instructions are available. On the top of the screen, seven words are available for the participants to choose from and fill in the blanks. These words are all underlined and the meanings of these words can be looked up by simply clicking on them. On the left side of the screen, a reading passage was presented, and on the right side of the screen are the following types of information:

item	
1. *choosing answer	answering questions
2. *choosing answer	
3. *choosing answer	back to cloze — while answering the
4. *choosing answer	questions, the participants can go back
5. *choosing answer	to the reading passage or go on doing
	the cloze task.

Notice: **done** — The participants could not click this item unless they have finished both cloze and reading comprehension questions.

In Task 4, the participants were asked to make sentences with the underlined target words. When the participants finished reading the passage, they click on the column **writing original sentences**, the target words in the text would pop up on another screen. The participants can make sentences by using all the target words. If they are not sure about the meaning of some of the words in the passage, they can click on the column **go on reading** to return to the passage. When they finish reading the first passage, they can click **go on reading the second passage**

Whenever a participant selects a component of task by clicking on it, the log registers the click. The online behavior of the participants during the task performance that are recorded by the program are as follows: the words that have been chosen to view, the amount of time spent on each target word, the amount of time spent on each task, and the number of times that the target word has been selected. When the task is completed participants clicked on the label *Done*. All the collected data would then appear on the results screen. Table II below illustrates what the logs look like at the end of a task. We will take Task 1 for example.

The participants' task-induced need, search and evaluation can be inferred from the number of clicks on the target words and the time-on-target words. For Tasks 1 and 2, the focus of observation would mainly be on whether the tasks induced need and search, and for Tasks 3 and 4, the focus would be on all of the three components.

TABLE II.
COMPUTER'S RECORDING OF PARTICIPANTS' ON-LINE LEARNING BEHAVIOR

I. Name	II. Task1	III. Log files
Yanna Lu	Passage 2	StartTime:15:55:58#ReadPassage:15:56:21#ReadQuestion:15:57:11#ReadPassage:15:57:18#agging:15:57:27#ReadQuestion:15:57:31#stopItem:15:57:33#provincial:15:57:48#ReadPassage:15:57:52#ReadQuestion:15:58:50#ReadPassage:15:59:13#ReadQuestion:16:00:05#ReadPassage:16:00:27#ReadQuestion:16:01:31#end:16:01:43#

E. Scoring

The same scoring system was used for the pretest, immediate posttest and the two delayed posttests. Specifically, only when the participants were able to remember or use the target words in the same meanings as those used in the reading passages were their answers regarded as correct, and awarded one point. The incorrect answers were scored no points. The scoring was implemented by two independent raters. Whenever there was difference in the scores given, they resolved the problem through discussion.

III. THE RESULTS

A. Effect of Different Task on Online Learning Behavior

The first key research question is: Do different tasks induce different online learning behavior as revealed in the number of times the target words selected and the amount of time spent on target words? The results for this question will be presented in two parts: one from a quantitative perspective, and the other from a qualitative perspective.

Quantitative Results

As can be seen from Table III, the participants who performed Task 1 and 2 on average clicked the nine target words 4.28 and 5.27 times respectively. This suggested that in Task 1 and Task 2 each target word was looked up less than once (i.e., 0.47 and 0.59 respectively). By contrast, in Task 3 and 4, each target word was looked up more than once (i.e., 2.10 and 1.63). It can therefore be inferred that Task 1 and Task 2 induced neither need nor search, while Task 3 and Task 4 induced both need and search. We compared the four tasks on dictionary activity by ANOVA and found the difference between them was significant ($F=29.38$). A *post hoc* test (Scheffé) was run to find out where the differences lie. Table IV showed that the mean numbers of target word selection in Task 1 was significantly different from those in Task 3 and 4, and it was also true of Task 2 when it was compared with Task 3 and 4. However, there was no significant difference between Task 1 and Task 2, nor was there any significant difference between Task 3 and Task 4.

We also compared the time on target words across the four tasks by ANOVA and found that the differences are significant ($F=27.27$). It is likely that the participants evaluated the meanings of target words only in Task 3 and 4, as on average they spent about half a minute on each target word. By contrast, in Task 1 and 2, the time spent on each target word was only 1 to 2 seconds. Therefore, it can be inferred that Task 3 and Task 4 induced "evaluation", while Task 1 and Task 2 did not.

TABLE III.
MEAN NUMBER OF CLICKS; MEAN TIME AND STANDARD DERIVATIONS ON TARGET WORDS

Task	Clicks on Target Words		Time on Target Words	
	Mean	SD	Mean	SD
1	4.28	3.03	17.56	11.72
2	5.27	3.39	25.73	15.47
3	18.90	7.36	243.67	85.79
4	14.70	8.09	248.95	157.22

TABLE IV.
SCHEFFÉ TEST ON DIFFERENCES BETWEEN MEAN NUMBERS OF SELECTION ON 9 TARGET WORDS ACROSS 4 TASKS

	Task 2	Task 3	Task 4
Task 1	-0.99	-14.63**	-10.42**
Task 2		-13.63**	-9.43**
Task 3	13.63**		4.20

Qualitative Results

According to Rieder (2002), one of the shortcomings of the current research on incidental vocabulary acquisition is that the selective and quantitative attitude of most studies tends to lead to results that do not grasp the qualitative nature of the acquisition process. To further scrutinize the details of how different tasks induced different learning behavior, the number of clicks on each target word and the amount of time on each target word by each participant under each task condition were re-examined. A summary of online behavior when learning the target word *prodigy* is given below.

As can be seen from Table V, in Task 1 and Task 2, almost all learners (94.4% and 90.9%) clicked the target word only once. In Task 3, the clicks ranged from once to three times, and the participants who clicked the words twice accounted for 71.4% of the group. In Task 4, the clicks on the words ranged between 1 and 5 times, and 25% of the participants who performed this task clicked the target words four or five times. As for the time on the target words, Table VI indicated that in Task 1 and Task 2, the time that most learners spent on target-words ranged between 4- 7 seconds; 3-13 seconds respectively, while in Task 3 and Task 4 the range becomes 5-102 seconds;18-175 seconds respectively. Therefore, the number of clicks on target words and the time on target words rose correspondingly as the tasks became more difficult and demanding.

TABLE V.
NUMBER OF CLICKS ON PRODIGY BY PARTICIPANTS ACROSS TASKS

Task 1	No. of clicks	0	1	2	3	4	5
	participants	1	17	0	0	0	0
	% within Task	5.6	94.4	0	0	0	0
Task 2	No. of clicks	0	1	2	3	4	5
	participants	2	22	0	0	0	0
	% within Task	9.1	90.9	0	0	0	0
Task 3	No. of clicks	0	1	2	3	4	5
	participants	0	3	15	3	0	0
	% within Task	0	14.3	71.4	14.3	0	0
Task 4	No. of clicks	0	1	2	3	4	5
	participants	3	4	5	3	3	2
	% within Task	15	20	25	15	15	10

TABLE VI.
AMOUNT OF TIME(SECOND) ON PRODIGY BY PARTICIPANTS ACROSS TASKS

Task 1	Amount of time	0-3	4-7	9	9-175
	participants	1	16	1	0
	% within Task	5.6	88.8	5.6	0
Task 2	Amount of time	0-3	4-13	17	17-175
	participants	3	18	1	0
	% within Task	13.7	81.8	4.5	0
Task 3	Amount of time	0-7	10-52	84	102
	participants	4	16	1	1
	% within Task	18.2	72.2	4.8	4.8
Task 4	Amount of time	0-17	18-52	77-99	175
	participants	3	13	3	1
	% within Task	15	65	15	5

Although different tasks did induce different incidental vocabulary learning behavior, yet the patterns of behavior for the tasks were not exactly what the Involvement Load Hypothesis predicted them to be. As was clear from Table VII, only Task 1 and Task 3 induced the predicted pattern of behavior, while Task 2 and Task 4 did not.

TABLE VII.
A COMPARISON OF PREDICTED AND OBSERVED TASK-INDUCED PATTERN OF INCIDENTAL VOCABULARY LEARNING BEHAVIOR

Task	Components of involvement							
	Need		Search		Evaluation		Index	
	P	O	P	O	P	O	P	O
1	—	—	—	—	—	—	0	0
2	+	—	+	—	—	—	2	0
3	+	+	+	+	+	+	3	3
4	+	+	+	+	++	+	4	3

Legend: P: predicted task-induced behavior and total involvement index

O: observed task-induced behavior and total involvement index

B. Effect of Different Task-induced Online Learning Behavior on Vocabulary Acquisition

The second research question is: Do different task-induced online learning behavior lead to different learning results in terms of retention rates of newly learned target words? This question is also approached from both quantitative and qualitative perspectives.

Quantitative Results

Table VIII summarized the group means of the four tests across 4 tasks. It showed a sharp increase in the mean from the pretest to the immediate posttest for all tasks, suggesting that a uniform effect of task on the incidental vocabulary learning. That is to say, whatever the involvement load of the task, the learners universally benefited from the task performance. It should be noted, however, the retention rates of target words differ significantly among the tasks in the immediate posttest ($F=3.58$), with the participants performing Task 3 gaining the highest score (Mean = 6.24), followed by those performing Task 4 (Mean = 5.40), Task 2 (Mean = 4.82), and Task 1 (Mean = 4.11) respectively. Also noteworthy is the delayed effect on the four tasks. The differential gains obtained in the 4 tasks disappeared in delayed posttest 1, as shown by the non-significant difference among the group means of the 4 tasks ($F=0.12$), but were preserved to a certain degree in delayed posttest 2 ($F=3.78$), though in a different pattern. Thus, it appears that the tasks assumed to induce different involvement load did not lead to the expected learning results, especially when the delayed effect of learning is taken into consideration.

TABLE VIII.
GROUP MEANS AND STANDARD DEVIATIONS OF FOUR TESTS FOR 4 TASKS

	Pretest		Immediate Posttest		Delayed Posttest 1		Delayed Posttest 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Task 1	0.67	0.69	4.11	1.78	3.22	1.96	4.72	1.74
Task 2	0.77	0.81	4.82	2.42	3.27	1.28	5.00	1.35
Task 3	0.71	0.64	6.24	1.92	3.43	1.60	6.05	1.63
Task 4	0.75	0.64	5.40	2.21	3.15	1.35	4.60	1.43

Qualitative Results

To further explore the relationship between online learning behavior and learning results in terms of qualitative analysis, a comparison is made of the participants' look-up behavior about each target word and the correct responses on each target word, as shown in Table IX. It was revealed that words that were looked up most did not necessarily lead to better retention rate. For example, although the number of clicks on the words *distinct*, *discipline*, and *provincial* was greater than those on *lag behind* in Task 2, the retention scores of former three words turned out to be lower than that of the latter.

TABLE IX.
A COMPARISON BETWEEN NUMBER OF CLICKS ON EACH TARGET WORD AND NUMBER OF CORRECT RESPONSES IN FOUR TASKS IN THE IMMEDIATE POSTTEST

	Task 1(n=18)		Task 2(n=22)		Task 3(n=21)		Task 4(n=20)	
Target Words	Clicks	CR	Clicks	CR	Clicks	CR	Clicks	CR
prodigy	17	15	20	17	42	15	45	13
distinct	10	7	15	8	41	14	46	9
nurture	7	8	16	14	49	20	37	13
discipline	8	2	12	6	49	12	34	6
provincial	10	4	12	3	47	11	27	6
significant	3	6	7	14	48	15	31	14
lag behind	9	18	9	16	49	16	20	18
shrink	5	6	15	13	32	13	30	15
sophisticated	7	8	10	15	38	15	23	14

Legend: CR = number of correct responses

IV. THE RESULTS

A. Involvement Load Hypothesis Revisited

The first research question aimed to investigate whether different tasks induced different online learning behavior. The results indicated that tasks assumed to have different involvement load did elicit different online learning behavior as reflected in the number of clicks on target words and time-on-target words. However, the patterns of behavior are not quite what the tasks were supposed to generate. Specifically, the involvement index in Task 2 (read, and then answer relevant questions) and Task 4 (read and then make sentence) were hypothesized to be 2 and 4 respectively. However, the observed task-induced involvement index for the two tasks was 0 and 3 respectively. Only Task 1 and Task 3 induced the predicted pattern of behavior; while Task 2 and Task 4 did not. This partly explains why the learning results were often not the same as predicted by the conditions in which the learners performed the task, a phenomenon that was also observed in other similar studies (e.g., Huang, 2003; Laufer & Hulstijn 2001). The discrepancies between the learning behavior predicted by the Involvement Load Hypothesis and the actual observation of the online learning indicated that it is difficult, if not impossible, to operationalize the concepts of need, search, and evaluation.

Secondly, the investigation of effect of task-induced on-line learning behavior on incidental vocabulary learning results showed that the participants performing the tasks with higher involvement load achieved significantly higher scores than those performing the tasks supposed to have lower involvement load in the immediate posttest, but the order of task effectiveness is not quite what the Hypothesis predicted. According to the Hypothesis, the order of task effectiveness are expected to have a pattern of Task 4>Task 3>Task 2>Task 1. However, the pattern of the observed

task effectiveness turned out to be Task 3>Task 4>Task 2>Task 1. Therefore, the Involvement load hypothesis was only partially verified. In addition, the advantage gained in the immediate posttest disappeared in delayed posttest 1 and was preserved to a lesser degree in delayed posttest 2, and the hypothesis was not verified in the delayed posttests.

Finally, in the present study, it was found the time learners spent on target words across the four tasks were significantly different. Therefore it might not be the nature of the task that induced effective learning, the time learners spend on target words also played a role in better vocabulary learning results.

B. Task-induced Dictionary Behavior and Incidental Vocabulary Learning

It was found that the mean number of clicks of the four tasks was significantly different from each other, with Task 3 inducing the most clicks (18.9), followed by Task 4 (14.7), Task 2 (5.27), and Task 1 (4.28) respectively. In the immediate posttest the retention rates of target words were of the same order as their dictionary activity, with the participants performing Task 3 gaining the highest score, followed by those performing Task 4, Task 2, and Task 1 respectively.

How do our results compare with previous studies of dictionary behavior in incidental vocabulary acquisition? Monica Hill & Batia Laufer (2003) asked 96 Chinese EFL learners to read a text containing twelve unfamiliar target words and performed one of the three tasks: a form-oriented production task, a form-oriented comprehension task, and a meaning-oriented production task. The results showed that the two form-oriented tasks were more effective for vocabulary learning than the meaning-oriented task. The tasks were different in the amount of dictionary activity they provoked. The order in effectiveness (Task 1– the worst, Task 3–the best) was the same as the order in the dictionary activity. Task 1 provoked the lowest activity and Task 3 the highest. They conclude, therefore, that an important factor determining task effectiveness for vocabulary learning is the amount of word-related activity that the task induces. In Monica Hill & Batia Laufer's (2000) study, they investigated the learners' dictionary look-up patterns and their retention of these words. It was found that the correlation between the number of selection and retention is weak and insignificant in the case of both Israeli learners and Hong Kong learners.

Compared with findings of the previous studies, our study supported that of Monica Hill & Batia Laufer (2003), for it was found that different tasks induced significantly different amount of dictionary activity, which, in turn, lead to different retention rates. The order of task effectiveness was of the same order with the amount of dictionary activity.

Why did different tasks induce different dictionary behavior, which in turn led to different learning results? We might explain the learning results from the perspective of Rider's "combined model of incidental vocabulary acquisition." According to Rieder (2002), the conditions for the incidental acquisition of an unknown word which a learner encounters in a text appear to be the result of two complementary determinants: the availability of the resources required for specifying the contextual meaning successfully, and the necessity for focusing on this particular word.

The observed online learning behaviors and learning results in the present study can be accounted for from the perspective of "necessity of focus on word". When performing Task 1 and Task 2, the learners' primary purpose would be the text comprehension. Therefore, they would simply stay on the textual level while trying to figure out the meaning of the unknown words. Even if the learners invest effort in inferring the meaning of the new word, they would still be operating on the textual level. In Task 3 and Task 4, the learners were asked to do the cloze exercise or produce their own sentences by using these target words. These two tasks are word-focused production tasks. Therefore, the learners felt obliged to invest greater efforts on target words by clicking them on computer. At this moment their focus would shift from the text level to the word level, and they would use all resources available, including background knowledge, dictionary information, to achieve the specification of word meaning. Therefore, Task 3 and Task 4 induced significantly more dictionary look-ups than Task 1 and Task 2 did. In addition, the learners' active shift from the text level to the word level would provide a better chance for the acquisition to take place.

V. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

In the present study, Laufer & Hulstijn's Involvement Load Hypothesis was tested by a different methodology from that adopted by the previous similar studies. That is, instead of focusing only on the product of learning, the present study also attended to the details of learners' online learning behavior via a specially designed computer program. This improvement in the research methodology enables a direct observation of how three major components of the model, i.e., need, search, and evaluation manifest themselves under different task conditions, which, in turn, provides the empirical basis for establishing the link between learning behavior and learning results in L2 incidental vocabulary acquisition. Although the Involvement Load Hypothesis was only partially verified in the present study, we gained deeper understanding of the nature of the three task-specific components (need, search, evaluation) specified in the Hypothesis. By observing learners' task-induced online learning behavior, it was found that the criteria that Laufer and Hulstijn have proposed for grading the task effectiveness may be too simplified for the sophisticated nature of the cognitive processes that the different task induce. The Involvement Load Hypothesis might be regarded only as an idealized conceptualization of how the traditional general labels such as noticing, attention, and elaboration can be operationalized into task-specific components.

The study shows that reading supplemented with word-focused activities, such as reading and filling in the text, plays a crucial role in building the learners' lexical knowledge. Teachers should be encouraged to design such tasks to foster

vocabulary learning. In addition, some reinforcement activities should be provided in teaching to consolidate the learners' knowledge of the words. Last but not least, students should be encouraged to use the multimedia program to assist vocabulary learning. The computer program kept a log of learners' learning behavior. Learners can be invited to see the results screen and feedback would be offered to them immediately after reading.

The present study is liable to the following weaknesses. First of all, the participants were not trained in different tasks they were going to perform. Some learners might click wrong button due to their unfamiliarity with the on-line learning. Future computer program should include a warm-up session before participants received the treatment. Secondly, the study is limited in the choice of target words. In the study some students had already known one sense of some target words before the study began. This might influence their learning processes and learning results. The target words chosen for the investigation are expected to be totally unfamiliar to learners in future studies.

ACKNOWLEDGMENT

I am deeply indebted to Professor Xudong Wu for his guidance and advice throughout developing this article. I am also grateful to all the learners who participated in the study.

REFERENCES

- [1] Cho, K.S. and Krashen. (1994). Acquisition of vocabulary from the Sweet Valley Kids Series: adult ESL acquisition. *Journal of Reading*, 37, 662-667.
- [2] Collentine, J. (2000). Insights into the construction of grammatical knowledge provided by user-behavior tracking technologies. *Language Learning and Technology*, 3(2), 44-57.
- [3] Haastrop, K. (1991). Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English. Tübingen: Gunter Narr.
- [4] Hu, Marcella and Paul Nation. (2000). Vocabulary density reading comprehension. *Reading in a Foreign Language*, 13, 403-430.
- [5] Huang Haiyan. (2003). Effects of task on incidental vocabulary learning by adult Chinese learners of English--Testing the Involvement Load Hypothesis. M.A. Dissertation, Guangdong University for Foreign Studies.
- [6] Hulstijn, J. H., Hollander, Merel & Greidanus, Tine. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use and reoccurrence of unknown words. *Modern Language Journal*, 80, 327-339.
- [7] Hulstijn, J.H., & Trompeter, P. (1998). Incidental learning of second language vocabulary in computer-assisted reading and writing tasks. In D. Albrechtsen, B. Henrikse, I.M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy*. Odense, Denmark: Odense University Press, 191-200.
- [8] Hulstijn, J.H. (2000). The use of computer technology in experimental studies of SLA: A survey of some technique and some ongoing studies. *Language Learning and Technology*, 3(2), 32-43.
- [9] Hulstijn, J. H. & Batia Laufer 2001. Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, 22, 1-26.
- [10] Hulstijn, J.H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (Ed), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 258-286.
- [11] Hulstijn, J. H. & Batia Laufer. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- [12] Keating, G. (2008). Task effectiveness and word learning in a second language: the involvement load hypothesis on trial. *Language Teaching Research*, 12 (3), 365-386.
- [13] Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58 (2), 285-325.
- [14] Knight, Susan. (1994). Dictionary use while reading: The effect on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, 78, 285-299.
- [15] Liuxiang. (2000). Collection of CET-4 examination papers. Tianjin: Tianjin University Press.
- [16] Luppescu, S. & Day, R. R. 1993. Reading, dictionary, and vocabulary learning. *Language Learning*, 43(2), 263-287.
- [17] Monica Hill and Batia Laufer. (2000). What lexical information do L2 learners select in a Call dictionary and how does it affect word retention? *Language Learning and Technology*, 3(2), 58-76.
- [18] Monica Hill and Batia Laufer. (2003). Type of task, time-on task and electric dictionaries in incidental vocabulary acquisition. *International Review for Applied Linguistics*, 41, 87-106.
- [19] Newton, J. (2001). Options for vocabulary learning through communication tasks. *English Language Teaching Journal*, 55, 30-37.
- [20] Parreren, C.F. van & Schouten-van Parreren, M.C. (1981). Contextual guessing: a trainable reader strategy. *System*, 9(3), 235-241.
- [21] Rieder, Angelika. (2002b). Web Transcription Tool. http://www.univie.ac.at/Anglistik/views/02_1&2/AR.PDF. (accessed 20/3/2013).
- [22] Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19, 287-307.
- [23] Wesche, M. B., & T. S. Paribakht. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *The Modern Language Journal*, 84, 196-213.



Jiehui Li born in Shan'xi, China in 1978, she received M.A degree in Linguistics and Applied Linguistics from Guangdong University for Foreign Studies in Guangzhou, China in 2005. She has been teaching College English for about 5 years in Shanxi Normal University, and 9 years in Guangzhou Civil Aviation College, Guangzhou. She has published more than 10 research papers about language teaching and learning. Her research interests include: Second Language Acquisition; EFL teaching in the Chinese context.