Learning Vocabulary through Extensive Reading: Word Frequency Levels and L2 Learners' Vocabulary Knowledge Level

May 2018 - Volume 22, Number 1

Anna C-S Chang
Hsing Wu University
<annachang@livemail.tw>

Hsueh-chao Marcella Hu
The Overseas Chinese University
<hhcm@ocu.edu.tw>

Abstract

This study investigates 62 EFL young adults' vocabulary learning rates immediately after reading ten graded readers, and their vocabulary decay rates after a three-month period. The participants were divided into two relative levels: higher (n =31) and lower (n =31). One hundred target words were semi-randomly selected from the reading texts and divided into three-word frequency levels (1,000, 2,000, and 3,000 plus off-list words). The levels had 31, 36, and 34 target words respectively. The target words were tested on three occasions: pre-, post- and delayed post-tests. The results show that the immediate learning rates for the higher-level group were 68%, 71% and 69% for the 1,000, 2,000 and the off-list words respectively, and were 20%, 19% and 21% for the lower-level group. The attrition rates in the three-word frequency levels were 7%, 11% and 22% in the higher-level group, and 31%, 34% and 36% in the lower level group. The results of the present study suggest that L2 learners learned any words that they encountered while reading without considering word frequency levels, and the higher learning rate of low frequency words could be attributed to the after-reading activities and the interesting stories in the graded readers.

Introduction

Some challenges involved in investigating vocabulary learning through extensive reading have been discussed by L2 researchers and teachers (Horst, 2005; Webb & Chang, 2015). The first challenge is that word knowledge measured before and after an Extensive Reading (ER) program should feature words that actually appear in the reading materials. The second is that learning outcomes should be solely from ER, meaning that learners do not have other learning opportunities (e.g., living in the target language countries or having target language friends) or other languages classes (e.g., speaking or listening) that may influence the outcome. The third challenge is that the participants should choose their own reading materials suitable to their language levels and interests. The first challenge is difficult to overcome unless some

compromise is made with the third. The second challenge is easier to overcome if the ER program is carried out in a foreign language context with students who are not used to reading extensively beyond their prescribed content. This study attempts to meet these three challenges in order to provide clearer evidence of the efficacy of ER. Comprehensive reviews for the theoretical benefits and empirical evidence of ER can be found in the works of Day and Bamford (1998), Grabe (2009) and Krashen (2004).

Vocabulary Learning Through Reading

In the past few decades L2 researchers have been attempting to determine whether vocabulary is learned more effectively through implicit reading or explicit instruction (e.g., Nation, 2001); however, there are many factors that may affect whether a word can be learned, such as frequency of occurrence, meaningfulness of the context of each word, similarity between L1 and L2, and learners' language proficiency. With few exceptions, most people agree that ER enhances vocabulary knowledge and that the more one reads, the more one acquires. However, many studies have shown that the rates of incidental vocabulary learning through reading a single text are low (see Horst, 2005 for a comprehensive review), ranging from two out of 28 words (Pitts, White, & Krashen, 1989) to six out of 12 (Rott, 1999). This is because these are read-and-test studies, meaning that students read a passage or a book and receive a test immediately after the reading. They are different from the studies of learning vocabulary through ER in many ways. Some salient differences are the reading quantity, with ER involving much more reading material at the learners' levels; the reading pace, ER students fully controlling their reading speed and re-reading if necessary; and ER learners determining their own reading processes, from consulting unknown words in the dictionary to taking notes of unknown words or ignoring them. These differences make a contrast between incidental vocabulary learning from reading and vocabulary learning through ER (Horst, 2005). Before reviewing vocabulary learning through ER, some relevant findings from previous studies on incidental vocabulary learning through reading a single text or laboratory-control research are summarized below:

- Language proficiency: Higher proficiency learners are able to acquire more words than lower level students (Tekmen & Daloğlu, 2006). Learners who have more vocabulary knowledge need fewer encounters with a word to learn it, while those who have less vocabulary knowledge need more encounters (Zahar, Cobb & Spada, 2001).

- Frequency of occurrence: There is no general agreement on the optimal number of encounters to guarantee learning; without word-focused activities, word frequency alone shows little effect on vocabulary gain (Laufer & Rozovski-Roitblat, 2011, 2015). The number of times needed to learn words may depend largely on the context in which they are encountered. For example, Horst et al. (1998) suggest eight times; Saragi et al. (1978), 10 times; and Waring and Takaki (2003), 20 times. Different dimensions of word knowledge need varying numbers of encounters (Webb, 2007). The recognition of spelling requires fewer exposures (Waring and Takaki, 2003), whereas word meaning is less affected by the degree of word frequency (Pigada & Schmitt, 2006).

- The use of dictionaries: Students who used a dictionary while reading learned more words and also comprehended better than those who guessed from context (Knight,

1994). Students who typed words into the digital dictionary to look for meanings scored higher in a spelling test (Liu, Pan, & Paas, 2014).

- Context type in which a word is: Researchers found no conclusive relationship between words learned and the type of context in which they appeared (Horst, 2000; Zahar, Cobb, & Spada, 2001). If unknown words repeatedly appear in more informative contexts, their meanings may be learned more quickly than those of words that appear in less informative or misleading contexts (Webb, 2008).

These factors seem to influence each other. No one single factor can determine the ease or difficulty of vocabulary learning. As a result, learning rates vary greatly from study to study, from the lowest rate of one out of 14 words (Zahar et al., 2001), through five out of 45 (Horst et al., 1998) and 85 out 300 (Horst & Meara, 1999), to the highest rate: six out of 12 (Rott, 1999).

Vocabulary Learning Through Extensive Reading

Compared to how words are learned incidentally through reading, how learners acquire vocabulary knowledge through ER is less well understood by researchers. One reason for this pointed out by Horst (2005) is that because most studies have used standardized vocabulary tests, the materials that the students read were not actually included in the assessment. This can be seen in the vocabulary measures column in Table 1. In comparison with studies on incidental vocabulary learning through reading, ER studies have much longer treatment periods—from one month to nine months. This feature allows researchers to explore learners' overall language improvement over time. Some ER studies, using various methods, have attempted to measure the vocabulary students learned from their reading materials (Cho & Krashen, 1994; Horst, 2005; Chang et al., 2011; Pigada & Schmitt, 2006; Webb & Chang, 2015). Due to their relevance to the present study, some of these often-cited studies will be reviewed in detail (see Table 1) because without understanding the learning process in each study—what was involved in the treatment procedure and what activities students did—little would be contributed to the understanding of vocabulary learning through ER.

Table 1. Selected Studies on Learning Vocabulary from Extensive Reading

Studies (by year)	Participants/context	Reading materials	Vocabulary measures	Treatment period / reading quantity per week	Other English input
Hafiz & Tudor (1990)	25 EFL secondary/ Pakistan	Graded readers	Using essay writing to analyze	23 weeks / 240 minutes per week	Yes
Cho & Krashen (1994)	4 ESL adults/USA	Simplified readers for NSs	Individualized vocabulary tests	Several months / varied from 2 to 5.75 per week	In English speaking milieu
Renandya, Rajan & Jacobs (1999)	49 EFL Vietnamese adults/Singapore	Graded readers	Integrated tests	Two months / varied	Yes
Sheu (2003)	65 EFL secondary/Taiwan	Mixed (Graded readers & children's books for NS of English)	Cambridge Key English Test	32 weeks / varied	Yes
Horst (2005)	17 ESL/Canada	Graded readers	Individualized vocabulary tests	6 weeks / 1.75 books per week	Yes/in English speaking milieu
Pigada & Shmitt (2006)	1 French learner/ England	Graded readers	Individualized vocabulary tests	1 month / four graded readers	Possible but unlikely
Lee (2007)	University students/Taiwan	Mixed	Vocabulary Levels Test	From 12 weeks to 9 months / varied	Possible but unlikely
Chang et al. (2011)	82 Junior college EFL students/Taiwan	Graded readers	Vocabulary Levels test & content vocabulary test	4 months / 10 books in total	No
Chang (2011)	19 Secondary EFL students/Taiwan	Mixed (Graded readers & children's books for NS of English)	Vocabulary levels test	26 weeks / varied from 28 to 38 books	Yes
Webb & Chang (2015) Suk (2016)	60 secondary EFL students/ Taiwan 171 Korean EFL university Students	20 graded readers Graded readers	Content vocabulary test Content vocabulary test	26 weeks 15 weeks	No Yes

In a small but in-depth study, Cho & Krashen (1994) asked four ESL adults (three Korean, one Spanish) living in America to read the *Sweet Valley* series for pleasure in their free time. They were asked to indicate the unknown words they encountered either by underlining them in the book or by listing them in a notebook. Participants read varying numbers of the books—eight, 18, 23, and 10—with means ranging from 2 to 5.75 books per week. To measure their vocabulary learning rates, individualized form-meaning vocabulary tests were developed, based on the unknown words the students had listed or underlined. Totals of 535, 396, 275 and 175 words were tested for the four individual participants respectively, and they answered correctly on 299, 316, 189 and 71 words respectively. The percentages of words correct were 56% (299/535), 80% (316/396), 69% (275/189) and 43% (71/ 175), with an average of 62%. Broadly speaking, it seems from these results that the more one reads, the higher the learning rate is, unless other factors are involved, e.g., using a dictionary. The learning rates in this study were very inspiring, and the findings shed light on how much vocabulary readers may possibly gain through reading. However, although we are told that these four participants had been learning English for many years and had been in America for periods of time ranging from a few months to seven years, the researchers did not specify their levels of English proficiency or vocabulary knowledge. This lack of information limits the understanding we can gain from the study of whether learners with more vocabulary obtain more from ER than those with less vocabulary knowledge.

The method used by Cho and Krashen (1994) to measure learning of specific words from each participant's reading materials was very time-consuming. A different approach was used by Horst (2005), who scanned portions of 37 graded readers ranging from 100% (a whole book) down to 18%. A vocabulary knowledge confidence level test using three rating options (know, not sure, do not know) was developed on the basis of the scanned text. The test contained 100 items, of which 50 words were from the 1,001-2,000 list, and 50 were off-list words. The six-week ER program was built into the regular three-hour ESL classes with 21 international students, whose language proficiency levels ranged from elementary to high intermediate. Each week, one hour was devoted to a variety of activities to support ER, including discussing books in pairs, adding entries to vocabulary notebooks, completing worksheets, or just reading silently. The learning rates were 76% for the 1,001-2,000 level words and 62% for the off-list words, with an average of 69%. The vocabulary confidence test using the vocabulary knowledge scale showed a rate of 51% (18 out of 35 words rated as fully or partially known). Compared to the gains documented in other studies of incidental vocabulary learning through reading, the learning rates for different level words in Horst's study were very high. It was unclear whether all the students tended to acquire more words from the 1,001-2,000 level or if the high-level students acquired more from the off-list words. Suk (2016) also selected different word levels (2,000, 3,000 and 4,000) to test her EFL university students, who were given 30 minutes per week for ER for a period of 15 weeks. Overall, the ER students gained 13.07 words out of 120 tested, their levels of knowledge improving from 51.63/120 to 64.70/120. However, the author did not report the gains from each level, which limits our understanding of whether students acquired more words from the higher frequency levels. This is a gap the present study is intended to fill.

More recently, another line of ER research (assisted reading with MP3 recordings of the texts) was carried out by Chang, Pang and Chang (2011) with 82 EFL young adults. These students

had never had any experience of doing outside reading beyond their formal textbooks, so graded readers were used as common class texts. The students were classified into two language proficiency levels, high and low, according to their in-house placement test results, and further by a vocabulary levels test administered by the first researcher. Lower-level students who needed more assistance were assisted with oral rendition while reading, whereas the high-level group adopted silent reading without assistance. Both groups read a total of 12 graded readers over a four-month period. Students' vocabulary learning was measured with a test of 120 target words chosen from the 12 graded readers. The learning rate was 32% for the high-level group and 30% for the lower level group. The difference between the two groups narrowed from 17% at the pre-test to 13% at the post-test. It is evident that the additional assistance received by the lower level group might have increased their learning effectiveness. The result supports a previous study by Brown et al. (2008), who found that reading and listening to graded readers at the same time enhanced vocabulary learning rates and retention. The target words test was not analyzed in terms of word frequency levels, which limits the understanding we can gain from the study of whether word levels affect the learning rates of students whose language levels are different.

The three aforementioned studies examined only students' vocabulary knowledge in meaning recognition, which, as many scholars comment, does not capture other types of word knowledge being learned (Laufer, 1997; Nation, 2001; Webb, 2005). Pigada and Schmitt (2006) carried out an in-depth study with an adult French learner living in England. Four graded readers were read during a period of a month. A total of 133 words were tested in three aspects (spelling, grammar and meaning) through a one-on-one interview. The results showed that 66 words (50%) were improved in one type of word knowledge, 13 (10%) in two types, and eight (6%) in three types, with improvement of some kind in word knowledge on a total of 87 words (66 + 13 + 8) out of 133 (65% of the words). In contrast to the previously mentioned studies, the participant neither consulted the unknown words in a dictionary nor took notes on them; however, the participant was aware that he would be tested after reading, which might have led him to focus on vocabulary learning. Though gains were made in knowledge of 65% of the words overall, this has to be interpreted cautiously because much of this gain came from the realm of orthography, not word meaning. The results may thus not be comparable to those from the three other studies, which focused on word meanings.

The four studies described above measured learners' vocabulary knowledge gained from ER using tests based on the actual texts that the students read. They provide researchers and language teachers with a quick picture of learners' vocabulary knowledge immediately after reading, but they do not show the incremental nature of the long-term learning process because none of the studies involved delayed post-tests. This limitation was overcome by Webb and Chang (2015), who recently carried out a two-term study with a group of students divided into high, intermediate and low-level subgroups according to their prior knowledge of vocabulary. The students were tested on 200 individual words (100 in each term) selected from 20 graded readers that the participants studied. The results showed that in Term 1 the relative gains on the immediate post-test were 63.18%, 44.64% and 28.12% for the high, intermediate, and low-level groups respectively. The researchers found that the learning gains were quite durable on the Term 1 three-month delayed post-test. Results were similar for

Term 2. More importantly, the study showed that prior knowledge had a great impact on learning vocabulary through extensive reading.

Compared to the rates found in studies on incidental vocabulary learning through reading (see Horst, 2005), the learning rates shown through studies of ER were much higher. From among these studies, the research design of Cho and Krashen (1994) would not be feasible for a larger sampling. Horst scanned only parts of the study materials, which led to an incomplete lexical profile. In addition, the vocabulary confidence measure using three bands (Yes, Not Sure, No) seems to be efficient, but it relies greatly on student honesty. At times, it may not be accurate because what students think they know might not actually be correct, leading to an overestimated result. The measurements used in the study by Webb and Chang (2015) were sounder and broader, but some questions remain unanswered: At what levels were the words that students of different levels of language proficiency acquired most or least? What were the attrition rates for the words in each frequency level for students of different levels? The present study was intended to look at how these two factors' (L2 learners' vocabulary knowledge and word frequency levels) impact on learning rates in learning vocabulary through ER. The study lasted for one semester in an EFL context, with 62 young adults divided into two relative vocabulary knowledge levels (high and low). All the materials were scanned and analyzed. Two research questions were addressed:

> RQ1: To what extent did word frequency levels affect the vocabulary learning rates from ER of L2 learners of different levels of language proficiency?
> RQ2: What were the attrition rates (after three months) in each word frequency level for the high and low-level students?

Methodology

The Participants

The participants were 98 Year-10 Taiwanese senior high school students from two intact classes, who had been classified into two relative levels (higher and lower) by the school administrators on the basis of their results of an in-house reading placement test. At the end of the treatment, however, only 62 of the students had fully completed the reading tasks and all tests. The data of the other 36 students who did not read all the books or who missed any of the tests were excluded from the study. An additional vocabulary levels test (VLT, Schmitt, Schmitt & Clapham, 2001) containing words from the 1,000[1], 2,000 and 3,000 levels was also given to all the participants by the first researcher to ensure that they were placed in the right classes. The high-level group scored 51/90 on the VLT, with 79%, 52% and 38% correct in the 1000, 2,000 and 3,000-word levels respectively. The lower level group scored 39/90, with 63%, 36%, and 32% respectively in the three-word levels. A *t*-test showed a significant difference between the two groups: *t* (60) = 6.78, *p*

The Study Materials

The participants in this study had never read any English stories and had never experienced reading independently, so choosing their own reading materials was difficult at first. On the basis of some scholars' suggestions (Hill, 2001; Webb & Chang, 2015) and the researchers' teaching experience, we started the program by using graded class readers. There are at least two advantages in adopting graded readers as common class texts: 1) the teacher can guide the whole class and influence the quality of reading, and 2) it is easy to exploit the content for discussion. Hence, ten level-one graded readers from the Oxford Bookworm series were used as class readers. These books had been used a few times by previous students of the first researcher; these students had reported the books to be very interesting. All ten books were scanned and were analyzed using Range computer software[2] and the British National Corpus (BNC). The vocabulary profile is presented in Table 2. As can be seen from the table, the total number of words in the 10 books was 56,188; around 93% of these words were from the first and second 1,000-word lists, and the average number of words in each T-unit[3] was 9.11 (SD = 6.26).

Table 2. Lexical Profile of the Ten Level-One Graded Readers

Word list	Tokens	%
1,000	48,527	87.17
2,000	2,964	5.28
Lower frequency words	2,579	4.59
Proper nouns	2,118	3.77
Total	56,188	

Dependent measures. To choose 100 target words from level-one graded readers (written within 400 headwords) to be tested was challenging for the researchers because there were not many words that were unknown to the participants. The researchers first read through the 10 graded readers, then checked the word lists in the students' junior high school textbooks and finally selected some words from the graded readers that might not be known by all students. To motivate students to complete the tests, some known words were included. Through text analysis (see Table 2), 100 words were randomly selected (see Appendix A) and were further analyzed with the Range computer software. Among the 100 target words, 31 words were from the 1,000 level, 36 from the 2,000 level, and only 3 from the 3,000 level; 30 were not on any lists. The three words from the 3,000-word level were grouped with those not on any list as "off-list words". The target vocabulary profile is shown below:

Word level	**Number of words / %**
1,000	31 words / 31%
2,000	36 words / 36%
Off-list	33 words / 33%

The target words were divided into 10 blocks, with around 10 target words and one distractor in each. The test format used was matching multiple-choice items. Students had to choose one correct Chinese translation for each target word. An example is shown as below. For the post-tests, all the words were rearranged. Students were unaware a post-test or delayed post-test would be given. Each class was allowed 30 minutes to complete the 100 items.

____ shiver	a. 面紗
____ maid	b. 女傭
____ cow	c. 掀開
____ whisper	d. 審判
____ pale	e. 顫抖
____ married	f. 乳牛
____ veil	g. 水桶
____ uncover	h. 已婚的
____ trial	i. 蒼白的
____ arson	j. 耳語
	k. 縱火

Data Analysis

SPSS version 23 software was used to analyze the data. The data used were only for those students who had attended the class regularly and had never missed any of the tests, i.e., those who had completed the pre-, post- and delayed post-tests, so data for a total of 62 students out of 98 in the two classes were used. The pre-, post- and delayed post-tests are referred to below as Time 1, Time 2 and Time 3. Students' immediate learning rates were calculated by the formula [(raw score for correct answers at Time 2 - raw score for correct answers at Time 1) / (total words tested - raw score for correct answers at Time 1)] * 100. In other words, learning rate was the gain in answers correct from the pre-test to the post-test as a percentage of the number of words not known in the pre-test. The attrition rates were calculated by the formula [(raw score for correct answers at Time 3 - raw score for correct answers at Time 1) / (raw score for correct answers at Time 2- raw score for correct answers at Time 1)] * 100. To

examine the differences in learning at the three-word frequency levels within each group, t-tests were performed.

Procedure

In the first week of the semester, the vocabulary levels test (Schmitt et al., 2001) and the pre-test consisting of 100 items were administered to all the students. Weeks 2 to 16 were the treatment period. In week 17 students were given a post-test, and a delayed post-test in week 29.

All the students had a total four sessions of English classes per week of 45 minutes per session. None of the students had other English input from other courses or outside the class, and the four sessions were completely devoted to extensive reading. All books were provided by the first researcher and read in the classroom, usually one book per week. Due to the students' lack of experience with independent reading in the past, the reading for the first two or three chapters proceeded very slowly because the instructor had to make sure that students understood the names of the characters or places where the story took place; otherwise students would have lost interest in continuing to read the story. All students read while listening to the audio recording of the book. This helped wean them away from word-by-word reading (Hill, 2001) and to guess unknown words from context rather than checking unknown words as they encountered them. Most students enjoyed the sound effects of the audio recording for its different accents, pitches and paces. There were some after-reading activities for students to do: sharing personal opinions, keeping a learning journal, noting down words they wanted to learn, or writing a book report or a summary about the story, but all these activities were voluntary (see also Horst, 2005). The instructor did not mark or correct students' book reports. One of activities that students enjoyed most was sharing personal feelings about each story. For example, students shared their personal opinions about why the elephant man still deeply missed his mother many years after she gave him away to someone who treated him badly. A similar question was used for comparing views when they read *The Phantom of the Opera*, in which one of the characters who was born with a very ugly face and his mother did not give him love but a mask to cover his deformed face.

A simple reading comprehension check was given after students finished each graded reader. Two forms of test are provided by the publishers. One form contains true/false questions, putting story events in order and easy cloze tests, while the other test contains 50 multiple-choice questions. Though there were many test items, the instructor usually only chose only 10 to 20 questions to be tested. It took students only a few minutes to complete the test.

Results

The descriptive statistics for the vocabulary test results are presented in Table 3. Percentages are used in the table because there was a slight difference in terms of the number of words in each word frequency level. The actual raw scores are shown in Appendix B. Before the intervention (the pre-test), the high-level group scored correctly averages of 64%, 65% and 57% on the 1,000, 2,000 and off-list word frequency levels respectively, whereas the low-level group scored 40%, 41% and 35% on the three levels respectively. After the intervention

(the post-test), the high-level group scored 88%, 90% and 86% items correct at the three levels. The improvements from Time 1 to Time 2 were statistically significant within each word frequency level (t-tests were performed and confirmed that each word level improved significantly, α set at .01); however, the learning rates among word levels were not significantly different from each other, which meant that word frequency level did not affect the higher-level group's learning rates. Students in the low-level group scored 52%, 52% and 48% in the three levels respectively at Time 2. The improvements from Time 1 to Time 2 were also significant within each word level (α set at .01). As with the high-level group, the differences among the learning rates at the various word levels were not significant. The overall learning rates differed between the high and low groups, but for each group gains at each word level were similar; in that respect, there were no significant differences.

The learning rates for each of the three-word levels in each group immediately after the treatment were quite similar in the case of each group: 68%, 71% and 69% for the high-level group, and 20%, 19% and 21% for the low-level group. The average learning rate for the high-level group was 69%, but it was only 20% for the low-level group. The difference in learning rates was thus 49% between the high and low-level groups. This higher learning rate may be an outcome of the fact that learners with more vocabulary knowledge comprehend texts better (Horst, 1998; Tekmen & Daloğlu, 2006; Webb & Chang, 2015). Another interesting pattern for both high and low groups was that the learning rate was the highest for the 2,000-word level, followed by the rate for those words on the off-list, and the lowest for the 1,000-word level. The differences, however, were marginal; there were no statistically significant differences between the levels, and this pattern may imply that when students read, they do not acquire high frequency words first—any words appearing in the text are likely to be learned.

Table 3 (the last column) shows that the attrition rates in the high-level group were 7%, 11% and 22% in the three-word frequency levels respectively, the average attrition rate from previous gains being 14%. The attrition rates for the low-level group were 31%, 34% and 36% in the three-word frequency levels, the average being 34%, which meant that at Time 3 the students had lost one third of the words they had learned three months earlier. It is evident that the low-level group forgot more words than the high-level group, and both groups also lost more words from the off-list. Another consistent pattern was that for both groups the retention rate was higher for the 1,000-word level, followed by the 2,000 level and off-list words. It is clear that when students continued to read, they had more opportunities to re-encounter the high frequency words. As shown in Table 2, more than 95% of the words in the graded readers were 1,000 and 2,000-level words plus proper nouns.

Table 3. Descriptive Statistics of Vocabulary Test Results for 1,000, 2,000 and Off-List Words for Students of High and Low Levels [Raw Scores (Maximum= 100) and in Percentages] Across Three Occasions.

	Time 1	Time 2	Time 3	Gain rates (Time 1-2)	Attrition rates (Time 2-3)
Higher level (n =31)					
1,000	64.48%	88.24%	86.47%	68%	-7%
2,000	65.23%	89.87%	87.19%	71%	-11%
Off-list	56.50%	86.31%	79.86%	69%	-22%
Total	61.81%	88.19%	84.55%	69%	-14%
Lower level (n = 31)					
1,000	39.85%	51.61%	47.97%	20%	-31%
2,000	40.86%	52.24%	48.39%	19%	-34%
Off-list	34.51%	48%	43.11%	21%	-36%
Total	38.45%	50.65%	46.52%	20%	-34%
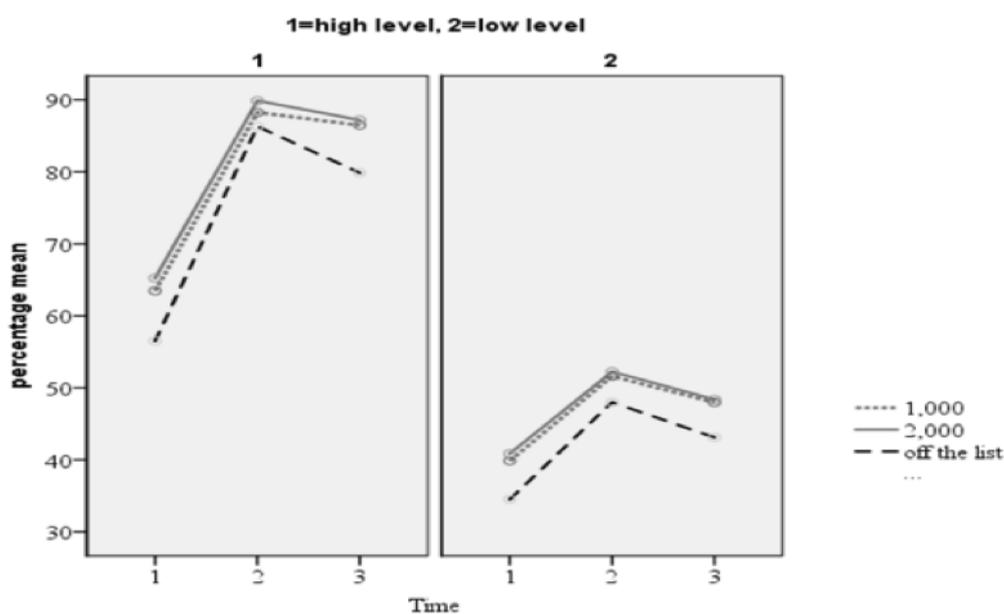
Figure 1: Vocabulary test results for 1,000, 2,000 and off-list words across three occasions for students of high and low levels (in percentages)

Discussion and Conclusion

The results of reading 10 level-one graded readers are summarized below:

- The average immediate learning rate of the high-level group was 69%, and the learning rates were quite similar across three-word levels, with 68%, 71% and 69% increases in scores in the 1,000, 2,000 and off-list levels respectively.

- The average immediate learning rate of the low-level group was 20%, with 20%, 19% and 21% in the three-word levels respectively.

- Though no significant differences in learning rates were found between word frequency levels, a consistent pattern shows that both high level and low-level groups learned slightly more words from the 2,000-word level, followed by those from the off-list, and the least from the 1,000-word level.

- The retention rates show a consistent pattern for both groups: the higher the word frequency level, the higher the retention rate. For the high-level group, the attrition rates were 7%, 11% and 22% for the 1,000, 2,000 and the off-list words respectively. For the low-level group, attrition rates were 31%, 34% and 36% respectively. The average attrition rate was 14% for the high-level group, and 34% for the low-level group.

Based on the results above, the answer to the first research question regarding learning rates in relation to word frequency levels was that students' learning rates were comparable across the three-word frequency levels, no statistically significant differences being found among them.

The second research question regards differences in the retention rates for each word level between high and low-level students. The results show that high-level students' attrition rate was 14% on average. The 1,000-word level was retained the most, followed by the 2,000 level, and the off-list words were retained the least. Low-level students' vocabulary knowledge decayed by an average of 34%. For the three levels, the attrition rates were 31, 34% and 36% respectively. The attrition pattern was the same as with high-level students, that is, the 1,000 level was retained the most and the off-list words the least.

The results of this study seem to corroborate the studies by Horst (2005) and Cho and Krashen (1994). The average learning rate in Cho and Krashen's study was 62%. Because Cho and Krashen did not analyze word levels, further comparison is limited. Coincidentally, the learning rate for off-list words in Horst's study was also 62%. Even though Pigada and Schmitt (2006) claimed that their word knowledge measurement was more comprehensive, the overall learning rate was again about the same at 65%. The learning rates of the high-level students in the present study are comparable to all the previous studies that measured specific words drawn from the material read. In terms of learning rates at each word level, Horst's students showed a higher learning rate for the 1,001 to 2,000-word level (76%); however, the higher-level students in the present study showed a higher rate for the off-list words than Horst's. The difference may be difficult to compare because different words were tested, and different test methods were used.

Despite the encouraging learning rates shown above for higher-level students, the much less successful learning rates by the low-level group appear worrying. Compared to the results in the studies by Chang et al. (2011), Pigada and Schmitt (2006) and Webb and Chang (2015), the learning rates for low-level students in the present study are much lower. A few salient differences in learning process between high and low-level groups were observed. One was that very few low-level students kept a learning journal (or took notes), whereas nearly all the high-level students did. A learning journal is an important bridge between the participants and the researchers or the instructors. As Little (2007) notes, journals move to the very center of the learning process when language learners keep journals in their target language, using them to capture all their learning, and cumulatively the journal becomes the story of the individual's language learning. It was likely that when the low-level students finished reading a book, they did not give much thought to what they had read. Another reason might be related to the first. Due to the absence of note-taking on what was learned, it was also highly likely that students seldom looked up unknown words in a dictionary, which further suggests that their motivation to learn words from reading was low. As Beck et al. (1983) notes: "If [students] identify an unknown word, it seems that only highly motivated students will choose to interrupt their reading to check on its meaning" (p. 180). Previous research (Liu, Pan, & Paas, 2014; Luppescu & Day, 1993; Knight, 1994; Cho & Krashen, 1994) has also shown that students who look up unknown words while reading remember them better than those who do not use a dictionary. Comparing the learning processes from the first researcher's observation, it is apparent that the high-level learners were involved in many more activities than the low-level students. These after-reading activities would have contributed greatly to higher rates of vocabulary learning because, as Nation (2001) notes, "Learning rates can be increased considerably by some deliberate attention to vocabulary" (p. 238). It has to be noted that all of these activities were voluntary, not required by the instructors, and it is difficult to quantify the degrees of the involvement load.

The learning rates found in these studies (Cho & Krashen, 1994; Horst, 2005; Webb & Chang, 2015; and the present study) ranged from 60% to 70%. These rates are much higher than those reported in studies of incidental vocabulary learning through reading (see Horst, 2005 for a review). A major difference lies in the exposure amount and learning process. If students read a quantity of 60,000 words, they repeatedly encounter the same patterns of words or combinations of words over a period of time, which certainly enhances the opportunity to learn. The reading process, as described earlier, can make a large difference. In addition, the present study also involved students listening to the corresponding audio CDs while reading. With audio-assisted reading, the speaker segments the texts for the readers, which not only helps readers comprehend better but also helps readers concentrate better on the task (see Chang & Millett, 2015), which may lead to higher vocabulary gain. Some other after-reading activities may have had a great impact on vocabulary learning and retention. As shown in the present study and those by Horst (2005) and by Cho & Krashen (1994), while students were reading extensively, they did not just read: they took notes on the words they did not know, discussed the plots of stories, and made comments about what they read. All of these post-reading activities were helpful in enhancing learning effectiveness, and this may be the main reason why these ER studies have shown higher learning rates than studies involving incidental vocabulary learning through reading a single text.

Limitations and Future Research

A few limitations of this study should be noted. First, the learning rates reported were limited to full knowledge of word meaning, and partial gains were not measured. This was not sufficient to capture a whole picture of what students actually learn; a variety of sensitive measures are needed (see, for example, Nation, 2001; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007; Nation & Webb, 2011; van Zeeland & Schmitt, 2013). Having said this, caution is necessary about the learning effect of taking many tests with the same or similar content, leading to an overestimation of the results. Furthermore, students' perceptions of taking many tests should be taken into consideration, because students may become fed up with these tests and not exert much effort to answer. For these reasons, some compromise between breadth and depth may have to be considered in future research. Second, all the target words were tested in a decontextualized manner, which may have affected student recall of the meaning of those target words. Third, word repetition in relation to learning gains at each word level was not explored. Some recent research (Chen & Truscott, 2010; Horst, Cobb, & Meara, 1998; Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003; Webb, 2007) has shown that repetition of encounters with a word determines to a great extent whether a word is learned, with the number of encounters adjudged necessary varying from six (Rott, 1999) to twenty (Waring & Takaki, 2003; Pigada & Schmitt, 2006). Apart from the number of repetitions of a word, how a word is repeated (the distribution of word encounters) is also an important factor. For example, sledge appears 49 times in the corpus of books read in the present study, but only in The Coldest Place on Earth, while remember appears a total of 67 times but with occurrences spread over nine out of the ten books. These variations make comparisons difficult in terms of word repetition frequency. This phenomenon was also acknowledged by Pigada and Schmitt (2006), who found that the distribution of word encounters made it very difficult to determine whether a word will be noticed or learned. Despite this limitation, unlike some texts written particularly for pedagogical purposes, the texts read in the ER program are more natural, as Zahar et al. (2001) noted: "natural text seems to be remarkably well-designed for learning purposes, which is not surprising since most native speakers of a language are able to expand their vocabularies through reading" (p. 557).

The low learning rate of low-level students should be of concern. It is possible that they were not ready to read extensively; direct instruction for effective vocabulary development may need to be considered. Another approach, assisted repeated reading (repeatedly reading and listening to short passages), suggested by Webb and Chang (2012), may be useful in developing the reading skills of beginner readers before starting extensive reading. If, after repeated reading, students' reading comprehension, fluency and motivation improve, then there may be greater progress in ER, and in turn, a higher rate of vocabulary learning.

This study sheds additional light on learning vocabulary through extensive reading. It not only supports previous research findings that L2 learners can acquire sizable vocabulary knowledge from reading extensively, but also reveals how learners acquire words in different frequency levels, and how much word knowledge can be retained across different word frequency levels. The study also suggested reasons why high-level students acquired more words than low-level students. If learning processes had been similar, then perhaps

comparable gain rates could have been expected for the two groups of students. Whether learning rates can be replicated at all levels is a matter for future research.

Notes

1. The first 1,000-word level was developed by Professor Paul Nation of Victoria University of Wellington, New Zealand and the Chinese translation was done by Karen Wang, Nation's former student. [back]

2. The vocabulary was analysed with the Range program, with the GSL/AWL List downloaded from Professor Paul Nation's website: http://www.victoria.ac.nz/lals/staff/paul-nation.aspx [back]

3. T-units measure the overall syntactic complexity and refers to a sentence, including all subordinate clauses. The tool for T-unit analysis is from Professor Tom Cobb's website, https://www.lextutor.ca/tools/ex_sent/ [back]

About the Authors

Anna C-S Chang is Professor in the Department of Applied English at Hsing Wu University, New Taipei, Taiwan, teaching courses on English listening, reading and vocabulary. Her main research interests focus on listening and reading development, and vocabulary learning. She has published extensively with internationally refereed journals.

Hsueh-chao Marcella Hu is assistant professor in the General Education Center at Overseas Chinese University. Her main research interests include vocabulary learning and teaching, lexical inferencing, and the role of cognition played in L2 acquisition.

References

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *The Elementary School Journal, 83*(3), 177-181.

Brown, R., Waring, B., & Donkaewbua, S. (2008). Incidental vocabulary learning from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, 136-163.

Chang, C-S. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching, 19*, 43-64.

Chang, C-S., Pang, C-L., & Chang, C. (2011). A comparison of gains to L2 learners from extensive reading while listening and reading only with story books. *Studies in English Language and Literature, 27*, 255-265.

Chang, C-S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System, 52*, 91-104.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary learning. *Applied Linguistics, 31*, 693-713.

Cho, K.-S., & Krashen, S. D. (1994). Learning of vocabulary from the Sweet Valley Kids series: Adult ESL learning. *Journal of Reading, 37*, 662-667.

Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom.* Cambridge, UK: Cambridge University Press.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* New York, NY: Cambridge University Press.

Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 Learning. *System, 18*, 31-42.

Hill, D. R. (2001). Graded readers. *ELT Journal, 55*, 300-324.

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

Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review, 56*, 308-328.

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

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary learning for students of different verbal abilities. *The Modern Language Journal, 78*, 285-299.

Krashen, S. D. (2004). *The power of reading: Insights from the research.* Portsmouth, NH: Heinemann.

Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, learning and pedagogy* (pp. 140-155). Cambridge, UK: Cambridge University Press.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391-411.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19*(6), 687-711.

Lee, S-Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal, 38*, 150-170.

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching, 1*, 14-29.

Liu, T. C., Fan, M. H. M., & Paas, F. (2014). Effects of digital dictionary format on incidental acquisition of spelling knowledge and cognitive load during second language learning: Click-on vs. key-in dictionaries. *Computers & Education, 70*, 9-20.

Luppescu, S., & Day, R (1993). Reading, dictionaries and vocabulary learning. *Language Learning, 43*, 263-297.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Pigada, M., & Schmitt, N. (2006). Vocabulary learning from extensive reading: A case study. *Reading in a Foreign Language, 18*, 1-28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language, 5*, 271-275.

Renandya, W. A., Rajan, B. R. S., & Jacobs, G. (1999). Extensive reading with adult learners of English as a second language. *RELC Journal, 30*, 39-61.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary learning through reading. *Studies in Second Language Learning, 21*, 589-619.

Saragi, T., Nation, I. S. P., & Meister, C. F. (1978). Vocabulary learning and reading. *System, 6*, 72-128.

Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviors of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55-88.

Sheu, S. P. H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter, 36*, 8-26.

Suk, N. (2016). The effects of extensive reading on reading comprehension, reading rates, and vocabulary acquisition. *Reading Research Quarterly, 52*(1), 73-89.

Tekmen, E. A. F., & Daloğlu, A. (2006). An investigation of incidental vocabulary learning in relation to learner proficiency level and word frequency. *Foreign Language Annals, 39*, 220-243.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

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

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

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1): 46-65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*, 232-245.

Webb, S., & Chang, C-S (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review, 68*, 276-290.

Webb, S., & Chang, C-S (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(4), 651 − 675.

Zahar, R., Cobb, T., & Spada, N, (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review, 57*, 541-572.

Appendix A: Target words

The 100 target words:
arson, arthritis, attic, autumn, bar, beach, Bible, blood, brave, bridge, candle, castle, cemetery, centimetre, chairman, chandelier, chess, clay, clever, club, cow, crash, creature, crowd, curse, detective, diamond, director, empty, enormous, factory, ferry, field, ground, hanging, horrible, inspector, judge, kitchen, knee, lawyer, lie, lift, lighthouse, lonely, machinery, magic, map, married, mask, milkmaid, millionaire, mines, mystery, opera, pale, passage, paw, peddler, plateau, pocket, poor, pretty, princess, prison, reach, remember, rope, scream, secret, sergeant, servant, shadow, shiver, shout, skier, skin, sledge, snowstorm, soldier, stage, stairs, stick, stream, stupid, suit, tablet, tears, temperature, tent, torture, trial, uncover, veil, vodka, whisky, whisper, wine, witch, wooden.
[back]

Appendix B: The Raw Scores at Each Word Level Across Three Test Occasions

	Time 1	Time 2	Time 3
Higher level			
1,000	20 (4.70)^2	27.35 (4.06)	26.81 (4.38)
2,000	23.48 (4.31)	32.35 (3.53)	31.39 (5.14)
Off-list	18.65 (4.62)	28.48 (4.52)	26.35 (5.07)
Total	61.81 (12.63)	88.19(11.64)	84.55(13.99)
Lower level			
1000	12.35 (2.85)	16.00 (4.43)	14.87 (6.55)
2,000	14.71 (4.05)	18.81 (5.08)	17.42 (6.88)
Off-list	11.39 (2.51)	15.84 (4.58)	14.23 (6.15)
Total	38.45 (7.91)	50.65(12.88)	46.52(18.82)

Note: The numbers in the parentheses are standard deviations. [back]