Critical Review of Research on Extensive Reading and Vocabulary Acquisition

Phillip Rowles

There has been a vast quantity of recent research into extensive reading, covering a range of studies from high to low quality. This previous knowledge is of particular interest to me as a teacher researcher as I am presently considering incorporating extensive reading into my curriculum. My teaching context involves instructing Japanese English as a Foreign Language university learners in their first, second and third years of tertiary education at an anonymous university in Tokyo, Japan.

The purpose of this paper is to critically review contemporary empirical research addressing the question, how does extensive reading affect the breadth of receptive second/foreign language vocabulary acquisition? The structural framework adopted in this paper starts with an introduction which contains a theory and concepts outline. The paper consists of two main sections: first, individual study summaries and critiques, and second, a discussion and conclusion. In the summaries and critiques section, the studies have been ordered chronologically, to give a developmental perspective, and the critiques examine the strengths and weaknesses of each study. The studies were selected on the basis of being contemporary studies of adults from major TESOL-related journals.

Krashen's (1981; 1982; 1984; 1985, in Ellis, 1990) Monitor Model, and his five related Hypotheses indicated a clean separation from audio-lingualism. One of these five Hypotheses, the Input Hypothesis, despite a lack of support
from empirical research, was promoted by Krashen as emphasizing, “the value of undirected pleasure reading as a source of comprehensible input” (Lightbown & Spada, 1999:39). The Input Hypothesis has been used to attempt to explain that acquisition, or subconscious development of linguistic competence, can take place where input is made comprehensible by the context. The symbols $i+1$ represent this input as just above the learners’ current language level, (Ellis, 1990).

Nation (2001) divides reading into two basic types: intensive and extensive. Intensive reading involves short texts usually less than 500 words long, with a focus on language, and where learners are familiar with less than 95% of the words. Extensive reading has two types: for fluency and for vocabulary acquisition. Extensive reading for fluency type involves fast reading, as 99 to 100% of the words are familiar. Extensive reading for vocabulary acquisition involves incidental vocabulary acquisition where learners are familiar with 95 to 98% of the words. This last type, extensive reading for vocabulary acquisition, is the focus of this paper.

In this paper, extensive reading is defined as reading that is self-selected and read by the learners, at a comfortable level where guessing from context is achievable, with a focus on meaning, conducted in large quantities, and over an extended period of time. Receptive vocabulary acquisition breadth (or size) refers to receptive intake of vocabulary (or words) that are incidentally acquired, as a by-product of focusing on meaning for comprehension while reading.

**Study One: Horst, Cobb & Meara (1998)**

**Summary**

Horst, Cobb & Meara (1998) conducted a quantitative study of a quasi-
experimental design investigating second language learners. This replication study of Saragi, Nation & Meister’s (1978, in Horst et al., 1998) first language study was in response to previous weak and methodologically problematic second language replication studies. Major problems with previous replications included relatively short texts, small gains in vocabulary acquisition compared to a control group, and incomplete details in the studies. The research hypothesis was recognition of more definitions and making more associations between vocabulary, after reading the relatively longer text of 21,232 words. The closest previous replication studies matched this in terms of text size was a text of only 6,700 running word length.

The 34 participants were low-intermediate English as a Foreign Language learners. Over a ten day period, a teacher read a 109 page simplified book The Mayor of Casterbridge (Jones, 1979, in Horst et al., 1998) to the participants in six sessions of about an hour each. They were pre-tested and post-tested using two tests; a 45-item multiple-choice and a 13-item word-association test. They found that a longer text showed more evidence of higher gain in vocabulary acquisition. In regards to frequency of encounters, they found that large vocabulary gains were likely to occur when the word item was encountered eight times or more.

Critique

Horst, Cobb & Meara’s (1998) study doesn’t match the definition of extensive reading outlined in the introduction. Although they used a relatively longer text of 21,232 words, they only used one book, which was read by the teacher to the students over about 6 classes of one hour each, conducted over a 10 day period. The level of appropriateness of the researcher-chosen reader for all participants is in question as the results imply it was, “roughly on target.
although some in the group must have found it challenging,” Horst et al. (1998:213). Further, “the match of the text and reader was less than perfect in the study, at least for the purposes of incidental vocabulary acquisition,” (Horst et al., 1998:219).

As for noticing, Horst et al.’s (1998) study allowed for this crucial first step, as learners were told to circle words they had problems with. In this study, the teacher read to the class, but there is no mention of the speed that the text was read at except, “the pace did not allow for looking up words in dictionaries,” (Horst et al., 1998:213). This brings into question whether the low-intermediate English as a Foreign language learners had enough time to notice unfamiliar words, circle them, and keep up with focusing on meaning in the text, set by the teacher-led reading pace.

As for acquisition Horst et al.’s (1998) study found more gains than previous problematic second language replication studies which indicated gains of 1 in 5 to 1 in 17. On one hand, these gains are correctly attributed to the longer text that was used. Using a longer text is a step in the right direction, however, as previously mentioned, it is still only one text, read over a short period of time. In testing, the pretest of the 45 item multiple choice test showed that a mean of 21.64 of the target words were known previously. This left 23 words up for acquisition. The post-test mean of 26.26 showed about five words (or 1 in 5) were incidentally acquired. The pre-test of the more difficult 13 item word association test showed a mean of 5.53 acceptable associations and a post-test mean of 6.71 showing a gain of about one association between three words.

As for retention of gains, there was no delayed post-test used so no claims for retention of acquired vocabulary were made in this study. The use of a delayed post-test would have provided this important information.

As for frequency of words in the text in relation to vocabulary acquisition,
a computer analysis determined that tested target words occurred 2 to 17 times. More acquisition was found when there was eight or more encounters with a word. Unfortunately, in this study only six words were found to occur in the text eight or more times. A text with more words that occurred eight or more times would have provided more chances for vocabulary acquisition.

**Study Two: Leung (2002)**

**Summary**

Leung (2002) conducted a qualitative diary study/case study in response to the multitude of quantitative research examining extensive reading. Data was collected using triangulation in the second stage of the study, using: learner diary entries, audio tape recordings and vocabulary tests. The purpose was to examine the experiences encountered while engaged in extensive reading, including vocabulary acquisition, reading comprehension, and attitudes.

There was one participant, Wendy, in this study who was also the researcher and author. Her first language was Chinese and her second language was English as she had lived in Hong Kong for 20 years. She had studied Japanese language there 10 years before this study. At the time of this study she had lived in North America for 10 years. In this study she was restarting as a beginner Japanese as a Foreign Language learner. The study covered two Stages of 9 and 11 weeks, with a two and a half month gap in between these two Stages. She read 32 books (mainly children’s simple storybooks, some comic books, and a few children’s textbooks) covering 1,260 pages. An average of about one hour was spent each day reading and she made one or two diary entries a week describing her experiences and progress. In Stage 2 she kept her regular study schedule, plus expanded her learning opportunities with a Japanese friend who tutored her for half an hour to an hour each
week. Audiotape recordings monitored reading sessions, questions, and tutor comments. These were used for identifying patterns and commenting on in her diary. She was tested twice with a test developed by a Japanese graduate student: at Week 16 and Week 20. The tested words represented a random sample from the textbook, Japanese for Busy People. Version A (Test 1) and Version B (Test 2) consisted of 50 words which were self-rated on a scale of usage ability. The results indicated a 23.5% improvement in one month.

Critique

Leung's (2002) study matches the criteria for extensive reading as set out in the introduction. In this study the participant read 32 books, covering 1,260 pages over two stages of 9 and 11 weeks (totaling 20 weeks). As she was at a low reading ability, she correctly chose books at an appropriate level that were of interest to her and her 2 year old daughter who she often read to. Therefore she read mostly simple children's story books, some comic books and children's textbooks.

One problematic issue was the reference to a beginning learner of Japanese. In actual fact, the participant had studied some Japanese language about ten years previously (covering writing and pronouncing hiragana and katakana characters). In addition to this her first language was Chinese, so she had some grounding in the basics of kanji characters. A more accurate description may be to label the participant as a false beginner of Japanese. Another concern is that there was a two and a half month break between Stage 1 and 2. The question of whether extensive reading continued to occur during this time or not, was not addressed.

Another problematic issue for some may be the fact that in this diary study/case study, the participant, researcher and writer of the study was one
person, Wendy. This may be too subjective for some researchers, especially those that have a leaning towards quantitative research. As this qualitative study was conducted in response to the vast amount of quantitative research already published, it is beneficial and important as it offers new insight and a fresh perspective, where quantitative studies have not covered due to their limitations concerning objectivity.

Problems also arise in the lack of some details provided. In the second stage, Wendy triangulated data by using tape recordings of tutorials, diary entries and a vocabulary test. The tape recordings included conversations about reading passages from a beginner-level Japanese textbook, questions, comments and oral reading. The test included a sample of words taken from the textbook, *Japanese for Busy People*, (AJLT, 1995, in Leung, 2002). It is not clearly defined if the reading passages in the tutorial tape recordings come from *Japanese for Busy People* or not. As has been clearly explained in the study, the extensive reading books read in this study were mostly simple children’s books, so it is difficult to make claims for vocabulary acquisition from extensive reading when the tested words were from a completely separate textbook that was studied in tutorial time.

The results of the test indicated that Wendy’s vocabulary acquisition improved by a reported 23.5% in one month from the words taken from the textbook, *Japanese for Busy People*. An examination of results table reveals that the score of 57/200 (28.5%) was obtained on the pre-test and a score of 69/196 (35.2%) was obtained on the post-test. This seems to indicate a vocabulary acquisition gain of only 6.3% (instead of the reported 23.5%). The pre-test, or more correctly the first test, was given in Week 16 and the post-test was given in Week 20. This only shows vocabulary acquisition over the last month of the study. The study lacks a diagnostic vocabulary test which could
estimate vocabulary level, help in book selection and combined with a post-test could give a more comprehensive overview of vocabulary acquisition over the entire study.

In the conclusion, Leung (2002:78) notes that, "Although the instructions Wendy received from her textbooks through self-study and her tutor contributed to her vocabulary knowledge and reading comprehension, reading extensively also played an important role in her learning process." Unfortunately because of the test used we don’t know exactly what level of importance extensive reading played.

Study Three: Gu (2003)

Summary

Gu (2003) conducted a qualitative case study in response to a number of quantitative studies which examined vocabulary learning strategies, but had failed to indicate how a specific kind of strategy is used in building vocabulary. The purpose was to examine the vocabulary learning strategies employed during and after reading by high-level participants. Two adult Chinese participants of a non-English major background were studied. Both were successful English as a Foreign language learners in the third year of university and they were selected on the basis of a national College English Test. The first participant, Chi Wei was selected as representative of learners who favored close study of textbooks, while Chen Hua represented learners who enjoyed reading many books outside of class. Chen Hua, the one who enjoyed prolific extensive reading, was also the fastest reader. They were given an intensive reading book passage to read, and after this, they were involved in audio-taped think-aloud protocols and interviews. A training session occurred before the study and inter-rater reliability was achieved by having two raters and a third-party rater to resolve
any disputes.

Chen Hua was infatuated with English literary imagery. She read the tested text twice on her way to learning vocabulary. First she read quickly and silently, and the second time she read for knowledge depth. For unfamiliar vocabulary she either: 1) guessed, ignored or underlined; 2) guessed, checked her dictionary and recorded meanings; or 3) marked them unknown, guessed, checked a dictionary, recorded the meanings in the margin, and recorded words on vocabulary cards. After reading she reviewed her new words again in the text and in her notes. As for list learning she used two types; 1) vocabulary word cards and 2) bought word lists. She found reading was more rewarding than memorizing from lists. While in middle school, she prolifically read nearly every simplified reader in China. The conclusion explained how she read for global understanding of a text and only zeroed in on new vocabulary that was interesting. She spent a lot of time on extensive reading, which provided further encounters with vocabulary she had tried to pick up in the intensive reading session. However, this didn't stop her from trying examination focused strategies. She was very flexible in strategy use, adopting intentional or incidental learning strategies when needed.

Critique

Gu's (2003) case study examined the vocabulary acquisition of two learners: one who preferred intensive reading and one who preferred extensive reading. Chen Hua, the participant who favored extensive reading certainly fit into the criteria set in the introduction. In her middle school days she had read almost all of the graded readers on offer in China. Even though no explicit numbers were given, we might be safe to assume a lot of extensive reading had been carried out in her free time outside of school. As she grew older and
moved into tertiary education, she started to read original works of English novels supported by Chinese notes. At the time of this study her reading level had improved to the point where she was reading only original English works in her free time for pleasure.

Gu (2003) states that this study was focused on intensive reading, as this was the most popular English input source in the Chinese English as a Foreign Language setting. Although Gu (2003:79) specifies that, “all subjects read an intensive reading passage,” this is questionable. To his benefit, Gu (2003) conducted a pilot study, with a random sample of 13 university non-English major third-year students (like the two participants) which showed the text familiar to unfamiliar word ratio was 43.7:1 or, “about 98% of vocabulary coverage, the safe threshold for text comprehension and learning,” (Leung, 2003:80). As the two participants were very high level, successful earners, we might assume that they would have higher coverage than 13 randomly selected learners. Therefore they would have higher than 98% coverage. This would put the so-called intensive reading passage into the territory of extensive reading for fluency zone as outlined by Nation (2001) in this papers’ introduction.

On the positive side, Gu (2003) had the participants undertake a training session for the audio-taped think-aloud session. Interviews were conducted straight after to examine any other strategies that were not voiced during the think-aloud part. A further strength was inter-rater reliability was achieved by having two raters and an objective third party to settle any disputes in coding the data.

Detractors of Gu’s (2003) study might say it was limited to examining only two highly successful learners of English. On the contrary, as this was a case study, the depth of insight achieved from focusing on only two participants is a strength. Chen Hua was the fastest reader and had a lot of extensive reading
practice to the point that she had memorized many textbooks after repeated readings.

**Study Four: Waring & Takaki (2003)**

**Summary**

Waring & Takaki (2003) conducted a quantitative experimental study in response to previous studies that had design problems with: scarce retention data, vocabulary depth, tests, and how the number of encounters influenced incidental vocabulary acquisition. The purpose of examining how much vocabulary is acquired in a foreign language was broken down into four research questions: 1) how much vocabulary can be acquired and retained?, 2) do more encounters lead to greater acquisition? 3) how quickly is vocabulary forgotten?, and 4) is there an effect of different tests on vocabulary gain scores?

Fifteen participants, all Japanese female university students aged 19 to 21, of lower-intermediate level, and mostly English majors, volunteered for the study. A graded reader *A Little Princess* (Oxford University Press) at Level 1 (400 headwords, with 5,782 running words) was chosen for ease of learner comprehension. Twenty five words with different frequencies of occurrence in the text, were changed into ‘substitute’ words by having their spelling changed, for example, house became windle, yes became yoot, and face became mand. After reading, every participant was informally interviewed. They were tested at this time using three tests: 1) word-form recognition, 2) multiple-choice (prompted recognition), and 3) meaning by translation (unprompted recognition). They were post-tested again using these same three tests one week later and delayed post-tested three months later. Results indicated answers to the four research questions as: 1) participants could learn some new vocabulary from reading, but the majority of the new words were not learned. Further,
the participants didn’t retain most of the words read and acquired. After three months, the participants acquired an average of one new word per hour of extensive reading, 2) more encounters lead to more acquisition, 3) about half of the words acquired were forgotten after three months, and 4) different test types showed different test scores.

It was indicated that for there to be a 50% chance of vocabulary recognition after three months, readers have to encounter the word at least eight times.

Critique

Waring & Takaki’s (2003) study doesn’t match the criteria set out for extensive reading in the introduction to this paper. Despite a criticism of earlier second language replication studies like Horst, Cobb & Meara’s (1998) study as, “showing modest but positive gains,” “rather short and the text was often quite difficult.” (Waring & Takaki, 2003:131), Waring and Takaki (2003) chose only one text at the 400 headword level with only 5,782 running words, which took an average time of 56.3 minutes to read. They assumed that choosing an easy text would be beneficial as, “all the other words in the book would be known as it was at a reading level far below what learners of their ability should have been capable of,” (Waring & Takaki, 2003: 137). However, despite this in the reading time there were some comments from the subjects, such as there are many unfamiliar words, but they were only told “please enjoy it”.” (Waring & Takaki, 2003:140). Further, they were not, “sure that the non-tested items (i.e., the surrounding co-text) were all known,” (Waring & Takaki, 2003:135). Another weakness was all learners were forced to continue reading the book that contained unusually-spelt substitute words, making for a possibly confusing, or not very pleasurable reading experience. Waring and Takaki (2003) try to differentiate the substitute words used in this study from nonsense
words used in other studies. They define nonsense words as a word describing an imaginary thing, like a ten-legged elephant, whereas substitute words are real things that have been given new spellings, so house was called *windle*. This lead to confusion as participants may have found the learning of substitute words more difficult if they had already learnt the usual English word and meaning previously. Especially a word like *windle* was confused with the words wind or window. In the end, the use of substitute words seems problematic.

As for noticing they did not require the participants to, "read and underline every word they did not know before reading the text," (Waring & Takaki, 2003:137). They didn’t pre-test the participants knowledge of the real words behind the substitute words as they didn’t want to focus attention on these words. Still, getting the participants to underline unknown words would have helped them notice them, also the researchers would have had some clearer idea of the participants pre-treatment ability level.

As for the tests used, to their credit the researchers did not test words in context, as these could be guessed from context at test time. Instead the participants were tested with three tests: word-form recognition, multiple-choice recognition and a meaning/translation test, to estimate gains in vocabulary. Again to their credit, after the first round of tests immediately following the reading, another round of tests was given seven to ten days later and then round three was given about three months after the reading.

As for repetitions of unfamiliar words, there was considerable variability. On the word form recognition test, if learners encountered the word eight times or more, they had a 50% chance of recognition three months later. A weakness of this study was that only 15 words out of 25, occurred eight times or more. So this limited the scope for repetitions and also testing retention of acquisition in a delayed post-test.
As for acquisition and retention of the substitute words the participants could, "learn some new words from their reading but the majority of the new words were not learned. Moreover, the subjects forgot the vast majority of their words they read and learned." (Waring & Takaki, 2003:148). This is hardly surprising as these were not real words but substitute words, which the participants had no chance of encountering as repetitions again. Further the statement that, "about half of the words learned in the reading are forgotten after three months," (Waring & Takaki, 2003:148) underlines the importance of encountering repetitions for acquisition and retention. Even if they did meet the new word more than eight times, they only experienced it in one reading session of an average of 56.3 minutes. This seems like an extremely limited chance for acquisition and retention, especially combined with the 14 other words that occurred eight or times.

**Study Five: Horst (2005)**

**Summary**

Horst (2005) conducted a pilot study of a quantitative experiment in response to the unanswered question in the published literature of, to what extent do adult extensive reading participants experience the vocabulary breadth that usually comes from an increased amount of reading? To determine if measuring an increase in vocabulary via extensive reading was possible, three research questions were examined: 1) is computer scanning whole texts viable?; 2) how often do off-list items (infrequent words) and words of other frequencies, occur in graded readers?; and 3) is constructing personal self-report checklist tests that sample each individuals reading choices viable?

The participants were 21 adults who were English as a Second Language immigrant learners in Montreal, Canada. They were from mixed backgrounds
with some recently arrived, while others had lived in Canada for up to five years. Reading was offered as an extracurricular activity outside of three hour classes that were held twice per week. There were 150 graded books, with five or more titles to choose from at each of six levels ranging from 400 to 3,800 headwords. Over the six week program, the mean amount of books checked out was 10.52 books (SD = 6.71). Therefore, on average each participant borrowed just under two books per week. Scanning of books into a computer started a few weeks before the project. As this was time-intensive it was decided to scan only the first 20 pages. After the participants were pre-tested they commenced borrowing books. To avoid exposure from other sources, infrequent words, or off-list words, were focused on. These were words not on the 2,000 most frequent wordlist (West, 1953 in Horst, 2005) or the Academic Word List (Coxhead, 2000, in Horst, 2005). The tests were designed to measure self-rated pre-test and post-test knowledge of off-list items read in the extensive reading books. In total, 37 books were computer scanned of the 62 titles read (or 67.12%). However due to time constraints, only the first 20 pages of each book was scanned into VocabProfile, a vocabulary software program. The words were divided into four frequency ranges. A pre-test of 100 items was given, including 50 items from the 1,001 to 2,000 frequency list and 50 from the off-list words. Six weeks into the study, personalized post-tests for 17 of the 21 participants were created. These were in the same format as the pre-tests but used words that were known to have been read in four of the books that each individual participant had borrowed. The tests took only 10 minutes to complete and were taken seriously. The results of the 17 participants indicated vocabulary gains for the 1,001 to 2,000 frequency words and the off-list words. As a further check, they were tested on another personalized test, based on Wesche and Paribakht’s Vocabulary Knowledge Scale (1996). Once again, vocabulary gains
were measured in over half of the pre-tested unknown off-list words.

**Critique**

Horst’s (2005) study does match the criteria for extensive reading set out in the introduction. The participants self-selected books from a wide variety of about 70 different titles, ranging in level from 400 to 3,800 headwords. Over the six week period, the mean amount of books borrowed by each participant was about two titles per week. One participant borrowed 33 books over the period. There was a great wealth of texts available for the participants to enjoy.

In order to create individualized tests, graded reader texts were computer scanned. This was one of the strengths of the study. On the other hand, a weakness of this was that due to time restrictions, only two thirds of the borrowed books had the first 20 pages computer scanned. As a result, six weeks into the study only 17 of the 21 participants had personalized post-tests developed. The tests themselves may be questioned by some people, as a personalized checklist requires the participant to be honest in their self-assessment. A strength of the study is that they required participants to do a second personalized test which measured partial knowledge, this was based on Wesche and Paribakht’s Vocabulary Knowledge Scale (1996 in Horst, 2005).

The discussion proposes that this way of measuring extensive reading is viable for the future, based on this pilot study. The test results indicated that over half (51.43%) of unfamiliar words were acquired. However, how long they were retained is not addressed in this study.

The real strength of this study was allowing the participants to read a great volume of texts. This has rarely been achieved in quantitative studies of extensive reading. Some of the reasons why are the controlling nature of quantitative studies, concerns about measurement through testing, with fears
of outside contamination, objectivity and generalizability. If the mean of 10 new words from four books was transferred to the, "average 10.5 titles read in the pilot study and to whole books rather than 20-page excerpts, it is possible to claim that the participants learned dozens of new words during the six-week experiment. But volume is clearly crucial," (Horst, 2005:378).

Discussion

To synthesize the critiques of the studies together, six points are presented to integrate the information together. The six points are: extensive reading criteria, noticing unfamiliar words, repetitions of unfamiliar words, guessing from context, acquisition, and retention. These are based partly on four basic assumptions proposed by Laufer (2003) that are fundamental to the stance that the main source of second language acquisition is reading.

The criteria for extensive reading was defined in the introduction of this paper. Briefly, it involves self-selection, pleasurable reading, guessing from context, a focus on meaning, in large quantities, and over a long time. Overall the quantitative studies had a hard time meeting these criteria. Horst, Cobb & Meara (1998) and Waring & Takaki (2004), two quantitative studies, didn't meet the criteria because of the use of only one book, over a short time and not self-selected by the participants, so all participants were in effect assumed to be at exactly the same reading level, which they were not. One quantitative study, the most contemporary one, by Horst (2005) met the criteria by providing a large selection of books from which the participants could choose freely over a relatively long period of six weeks. The two qualitative studies, Leung (2002) and Gu (2003) also met the criteria for extensive reading. A recommendation is made for all studies to examine genuine extensive reading when they set out to research into this field. In the past quantitative studies examined one book over
a short time in a classroom setting. A quantitative study such as Horst's (2005), shows it is possible to move away from this limited setting, examine genuine extensive reading and still carry out a quantitative analysis.

Noticing unfamiliar words is a very important first step towards acquisition. Attention needs to given otherwise the participant may simply ignore the unfamiliar word. "The first process encouraging learning is noticing, that is giving attention to an item. This means that learners need to notice the word, and be aware of it as a useful language item," (Nation, 2001:63). Nation (2001) suggests that noticing can be encouraged by underlining, using word cards, or orally repeating unfamiliar words. Horst et al (1998) encouraged noticing to an extent by suggesting that participants circle problem words but as the pace was quite fast, it is questionable how much noticing took place. Leung (2002) used noticing by recording word meanings in journals and getting the tutor to orally pronounce unfamiliar words. One of Gu’s (2003) participants, Chen Hua, the extensive reading expert, used some strategies for noticing, for example, underlining and recording on word cards. Waring & Takaki (2003) didn’t encourage noticing of unfamiliar words in their study. Horst (2005) was not aware of what the participants did to notice. Noticing is a crucial tentative step towards acquisition and is recommended in studies in order that unfamiliar words are not just simply ignored.

The exact number of repetitions of unfamiliar words required for acquisition is not easy to determine, as the correlation between the two is moderate. Studies have found that anywhere from five to sixteen repetitions are needed for acquisition (Nation, 2001). There was a pattern present in the studies, as Horst et al (1998) and Waring & Takaki (2003) both found a minimum of eight repetitions as the threshold point for marked vocabulary acquisition. In Gu’s (2003) study although no figures were given, Chen Hua
spent, "much more time on extensive reading, which ensured the natural recurrence of words she had tried to read in intensive reading." (Gu, 2003:95). Both Horst et al. (2005) and Leung (2002) did not include repetitions of encounters relative to acquisition in their results. As repetitions help participants to encounter new words in their step-by-step climb towards acquisition, a focus on repetitions is recommended for studies.

When guessing from context, credit for partial knowledge is important as vocabulary acquisition is a gradual, incremental process (Nation, 2001). Therefore the decision of which test to use in studies should take this process into consideration. An example of a test that measures partial knowledge of words is Wesche and Paribakht’s (1996 in Nation, 2001) Vocabulary Knowledge Scale, where participants can display knowledge on a 5 point scale. Two studies, Leung (2002) and Horst (2005) used tests based on this scale. Horst et al. (1998), Gu (2003), and Waring & Takaki (2003) didn’t include this type of scale measuring partial knowledge. As guessing from context is not on an all-or-nothing scale, a more gradually scaled test of vocabulary acquisition, like the Vocabulary Knowledge Scale, is recommended.

Contemporary views of vocabulary acquisition from incidental learning in extensive reading admit it happens but the amount can be rather small as expressed by, "There is no doubt that incidental learning occurs, particularly through extensive reading in input-rich environments, albeit at a rather slow rate," (Read, 2004:147). Most of the studies found relatively small gains. Horst et al. (1998) found a gain of about 1 in 5 in the multiple choice test, some may question this finding because of the effects of guessing in a multiple choice test. Leung (2002) reported a high gain of 23.5 % in one month, but this is really a 6.7% gain. Gu (2003) was examining how acquisition took place, rather than how much acquisition took place. Waring & Takaki (2003) admitted that
not much vocabulary was acquired. Horst (2005) claimed that more than 1 in 2 unfamiliar words were acquired, however this claim comes with several warnings. Therefore, vocabulary acquisition can occur through extensive reading, but it may be quite small.

Retention of vocabulary acquisition refers to how long the meaning is retained in long term memory (Laufer, 2003). Only one of the studies, Waring & Takaki (2003) attempted to measure this with two delayed post-tests of one week and three months after the treatment. However because Waring & Takaki (2003) were using substitute words, with no chance of repetition outside of this study, they were unreasonably measuring participants capacity to remember 25 substitute words from an average 56.3 minute reading of a solitary text, several months before. The other studies didn’t include delayed post-tests. Retention of acquisition is important but it should be linked to the criteria of extensive reading which requires a lot of reading, with many repetitions, over an extended period of time.

Ideally future studies investigating the effect of extensive reading on receptive second/foreign language vocabulary acquisition should take into consideration certain recommendations. Initially they should meet the criteria outlined for extensive reading, that is, texts would be self-selected and read at a comfortable level. A long study involving many texts, for example, one text read per week, over a whole year would be very enlightening. Encouragement could be provided to notice unfamiliar vocabulary like underlining or circling. The texts themselves could be screened beforehand to ensure there were adequate repetitions or encounters with unknown words. The participants could be diagnostically tested using a test such as Nation’s (1983, in Nation 2001) Vocabulary Levels Test to obtain an estimate of receptive vocabulary breadth, and also to assist in text pre-selection. Partial knowledge could be measured
using a test such as Wesche and Paribakht's (1996, in Horst 2005) Vocabulary Knowledge Scale. Retention could be measured by several delayed post-tests given after the treatments.

There are small reported gains in many of the studies. This may be because of two reasons. First, there are only small gains available. If we look back at Nations (2001) criteria for extensive reading for vocabulary acquisition, 95 to 98% of words are familiar. This means there is a very small and precious opportunity, that is, only between 2 to 5% of the words in the text available for acquisition. For acquisition to occur many processes have to occur as outlined before, with the six points starting with noticing. Second, in regards to first language learners, “Incidental learning via guessing from context is the most important of all sources of vocabulary learning,” (Nation, 2001:232). Nation continues with, “It should also be true for second language learners, but many do not experience the conditions that are needed for this kind of learning to occur,” (Nation, 2001:232). Indeed two of the five studies examined here did not meet the criteria set for extensive reading and therefore the conditions needed for this kind of learning were absent.

**Conclusion**

Extensive reading may not provide great gains in vocabulary acquisition, but this is no reason to dismiss it as worthless. Extensive reading does provide opportunities for widening the breadth of receptive second/foreign language vocabulary acquisition, even if it may be limited in scope. In addition, it has other benefits. Two other benefits may be improving reading comprehension ability and the recycling of familiar vocabulary (Nation, 2001).

In regard to my original concern about incorporating extensive reading
into my curriculum, I will implement it into the curriculum. However, to achieve a further widening of the breadth of receptive second/foreign language vocabulary acquisition, a combination of approaches may be required. “Although learners certainly acquire word knowledge incidentally while engaged in various language learning activities, more direct and systematic study of vocabulary is also required,” (Read, 2004:146). For increased vocabulary acquisition, a combination of incidental learning (indirect forms) and intentional learning (direct forms) may be required. Indeed, Hulstijn (2001, in Read 2004) sees these two forms as not opposed, but actually complementary in the classroom context.

References


Laufer, B. (2003) Vocabulary Acquisition in a Second Language: Do Learners


