

Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories

Ronan Brown
Seinan Gakuin University
Japan

Rob Waring
Notre Dame Seishin University
Japan

Sangrawee Donkaewbua
Rajabhat Mahasarakham University
Thailand

Abstract

This study examined the rate at which English vocabulary was acquired from the 3 input modes of reading, reading-while-listening, and listening to stories. It selected 3 sets of 28 words within 4 frequency bands and administered 2 test types immediately after the reading and listening treatments, 1 week later and 3 months later. The results showed that new words could be learned incidentally in all 3 modes, but that most words were not learned. Items occurring more frequently in the text were more likely to be learned and were more resistant to decay. The data demonstrated that, on average, when subjects were tested by unprompted recall, the meaning of only 1 of the 28 items met in either of the reading modes and the meaning of none of the items met in the listening-only mode, would be retained after 3 months.

Keywords: incidental vocabulary acquisition, graded readers, recurrence rate, vocabulary decay, extensive reading, reading-while-listening, extensive listening

Incidental learning is the process of learning something without the intention of doing so. It is also learning one thing while intending to learn another (Richards & Schmidt, 2002). In terms of language acquisition, incidental learning is said to be an effective way of learning vocabulary from context (Day, Omura, & Hiramatsu, 1991; Jenkins, Stein, & Wysocki, 1984; Nagy, Herman, & Anderson, 1985; Saragi, Nation, & Meister, 1978).

Among the early studies of vocabulary acquisition in first languages (e.g., Boettcher, 1980; Carey, 1982; Clark, 1973; Dale, O'Rourke, & Bamman, 1971; Deighton, 1959; Eichholz & Barbe, 1961; Gentner, 1975), the study by Nagy et al. (1985) is particularly significant. In the

course of their research they developed a methodology for measuring small gains in vocabulary knowledge. They detected that a single incidental encounter of a word would seldom lead to full knowledge or understanding of a word's meaning. Moreover, if learning the meaning of vocabulary from context does occur, Carey (1978) suggested that it must be on the basis of encounters perceived in an incidental way. Because of this, learning vocabulary is understood to be a gradual process (Deighton, 1959). Nagy et al. (1985) declared that when this gradual learning process is encouraged by the help of contact with a sufficient amount of written language exposure, incidental vocabulary learning in the first language can be substantial.

Studies on incidental vocabulary acquisition in the foreign language typically involve subjects in extensive reading. One goal of extensive reading is to read for pleasure, which will hopefully translate into general language improvement and a boost in reading motivation (Krashen, 1994). The general language-learning process from extensive reading is incidental, with few specific learning demands from the teacher (Widdowson, 1979). Some researchers suggest that extensive reading is mainly for the purpose of reinforcing partially known words so that they may move up to known words, rather than focus on building new vocabulary (Nation & Wang, 1999; Waring & Takaki, 2003). Nevertheless, this does not exclude the learning and the acquisition of new vocabulary entirely.

Extensive Reading

There is a strong connection between incidental vocabulary learning and extensive reading, perhaps because of the definition of extensive reading. According to Bright and McGregor (1970), Day and Bamford (1998), Harmer (2003), Krashen (1993), Nation (2001), and Waring (1997), extensive reading is a pleasurable reading situation where a teacher encourages students to choose what they want to read for themselves from reading materials at a level they can understand. Krashen's (2003) *comprehension hypothesis* claimed that comprehensible input is a necessary and sufficient condition for language development and extensive reading provides this condition. Through the provision of engaging language-learner literature, extensive reading programs aim to develop reading fluency, and reading skills in general, while at the same time consolidate knowledge of previously met grammatical structures and vocabulary.

There has been a reasonable amount of research on incidental vocabulary learning from extensive reading (e.g., Day et al., 1991; Dupuy & Krashen, 1993; Grabe & Stoller, 1997; Hayashi, 1999; Mason & Krashen, 1997; Pigada & Schmitt, 2006; Pitts, White, & Krashen, 1989; Waring & Takaki, 2003). Several studies of such extensive reading programs have cited gains in overall language development (e.g., Cho & Krashen, 1994; Elley, 1991; Hafiz & Tudor, 1990). Other studies have emphasized benefits such as increased motivation to learn the new language and renewed confidence in reading (e.g., Brown, 2000; Hayashi, 1999; Mason & Krashen, 1997). In addition, research has indicated that the productive skills of writing and speaking have similarly been enhanced (Cho & Krashen, 1994; Janopoulos, 1986; Robb & Susser, 1989).

Horst, Cobb and Meara (1998) claimed that through extensive reading learners can "enrich their knowledge of the words they already know, increase lexical access speeds, build network linkages between words, and...a few words will be acquired" (p. 221). In their vocabulary study,

a multiple-choice, immediate posttest measure indicated that of 23 new words available for learning in the graded reader *The Mayor of Casterbridge*, 5 words were learned, which is a gain of 22%. In a similar study conducted by Waring and Takaki (2003), a multiple-choice, immediate posttest measure indicated that of 25 new words available for learning in the graded reader *A Little Princess*, 11 words were learned (as measured by success on these tests), a gain of 42%.

In a further study conducted by Horst (2005), a modified *vocabulary knowledge scale*, immediate posttest measure indicated that of 35 new words available for learning in self-selected graded reading materials, 18 words were learned: a gain of 51%. These gains are comparable to those achieved in the *A Clockwork Orange* investigation conducted by Saragi et al. (1978). In their study, subjects were able to correctly identify the meanings of 75% of the target words, especially the frequently recurring ones, in an unannounced multiple-choice test given immediately after the reading treatment. Since Saragi et al., approximately 10 other investigations have been undertaken to determine how much vocabulary is learned from reading in a foreign language. For a meta-analysis of these oft-cited, learning-from-context studies of vocabulary growth, see Horst or Waring and Nation (2004).

The study of Waring and Takaki (2003) is particularly significant. Like Nagy et al. (1985), they too developed a methodology for measuring small gains by having several test formats. Where other studies had used only one measurement, this study used three different kinds of measurements. The measurements were a simple *yes* or *no* sight-recognition test, a standard multiple-choice test, and a translation test into the first language. Their results showed that incidental vocabulary learning from reading occurred at several levels and the gain scores depended on the test type, but not much new vocabulary was learned.

Reading-While-Listening

A form of extensive reading that has recently been receiving more attention from language teachers and researchers is reading while simultaneously listening to an audio recording, or to the teacher reading a narrative aloud. The benefits cited have included increases in overall language proficiency, particularly listening comprehension, as well as the ability to acquire a greater sense of the rhythm of the language, which in turn can help learners to read and listen in meaningful sense groups rather than adopt a word-for-word strategy (Day & Bamford, 1998). Moreover, used as a strategy for promoting extensive reading, reading-while-listening can also pay dividends, provided that learners understand “it might take [time] for concentration to develop...eventually the moment will come when students are actually reading ahead of the teacher and at the end of the lesson students carry on reading and ask to take the books home” (Smith, 1997, p. 34).

Studies investigating the effectiveness of reading-while-listening for comprehension have claimed that because low-proficiency English as a foreign language (EFL) readers tend to break sentences into small incoherent parts while they read (thereby spoiling the sentences’ integrity and rendering them meaningless), the teacher reading aloud early on in a program helps retain that integrity by presenting larger semantic units, which in turn leads to better comprehension.

Thus, by adopting a more holistic approach, learners may realize that a higher level of comprehension is possible when engaged in reading while listening to larger chunks of texts rather than attempting to understand single words or unintelligible bits of sentences (Amer, 1997; Dhaif, 1990). In terms of vocabulary growth, the teacher reading aloud while the learners follow the written text created the conditions necessary for the incidental vocabulary acquisition gains of 22% in the Horst et al. (1998) study cited earlier. In this study, reading aloud focused the subjects' attention on the events in the story, and allowed the text itself and a few pictures to function as support for learning new words.

Extensive Listening

Research undertaken to determine the benefits of extensive listening (i.e., listening to long, easy texts for fluency and enjoyment) has largely been concerned with native-speaker populations, particularly early readers in elementary school. Reading stories to children is almost universally acknowledged as good pedagogy, and when it is done in an environment of shared reading or recreational reading, it also produces considerable gains in reading and listening skills (Elley, 1989; Senechal & Cornell, 1993). A further benefit of listening to stories is the potential for acquiring new vocabulary incidentally. In a set of studies conducted by Elley, it was found that oral story reading constituted a considerable source of vocabulary acquisition, whether or not the reading was accompanied by teacher explanation of word meanings. Subjects in one group showed gains of 15% from one story, without teacher explanation; while subjects in a second group, who did receive teacher explanations, showed gains of 40%. It was further found that these incidental vocabulary gains were relatively permanent, and that a key predictor of the successful acquisition of a word was its frequency of recurrence in the story.

Although the number of research studies on extensive listening in a foreign language is limited, there is a certain amount of didactic literature on the benefits and procedures of reading stories to students (e.g., Moody, 1974; Prowse, 2005). West (1953) argued that reading aloud to the class was "valuable for practice in understanding correctly spoken English and the appreciation of literature" (p. 21). In addition, Nation (2001) claimed that "there is a growing body of evidence that shows...that learners can pick up new vocabulary as they are being read to" (p. 117).

From the foregoing, successful learning of new vocabulary has been shown to take place when EFL learners are engaged in either an extensive-reading condition or extensive reading-while-listening condition. However, we know little about the rate at which vocabulary is picked up in these two modes. Would more vocabulary be learnt by reading only, or by reading while listening to a text? Moreover, as native-speaking children have been shown to acquire new vocabulary from listening to stories (Eller, Pappas, & Brown, 1988; Elley, 1985; Elley, 1988; Elley & Mangubhai, 1981), it is also pertinent to determine the rate at which foreign-language vocabulary is learnt while only listening to stories. This question is of vital importance as it can help determine how much reading or listening (and what type) needs to be done in foreign language learning. The investigation that follows, therefore, is primarily concerned with how foreign-language vocabulary acquisition rates compare across these three distinct input modes.

The main questions under investigation in this paper are as follows:

1. Do the subjects learn more vocabulary from reading, reading while listening, or listening to stories?
2. At what rate is this new vocabulary knowledge learned, and at what rate does it decay?
3. Are the subjects more likely to learn a word if they meet it more often?
4. Are there significant differences in acquisition rates depending on whether the test is a multiple-choice test or a meaning-translation test?
5. Do the subjects prefer to read only, read while listening, or listen only to stories?

Method

In this study, 35 subjects in three experimental groups read and listened once to three stories in graded-reader form, each of which was approximately 5,500 words long. The reading and listening treatments took place during three regular 90-minute classes at intervals of 2 weeks. The subjects were then assessed on their recognition and recall of the target vocabulary items with varying frequency of recurrence rates that they had met in each story. Similar to the Waring and Takaki (2003) study, it was decided that the vocabulary acquisition would be assessed at two levels and over three test periods. Eighty-four target words (3 sets of 28) were selected from three 400-headword-level graded readers. These words, which represented already known common concepts to the subjects (e.g., *letter*, *restaurant*, *family*), were then changed into substitute words. See Table 1 for an overview of the study.

Table 1. *An overview of the study*

Text	Group A ($n = 12$)	Group B ($n = 14$)	Group C ($n = 9$)
<i>The Elephant Man</i>	Listen (Week 2)	Read + listen (Week 4)	Read (Week 6)
<i>One-Way Ticket</i>	Read (Week 4)	Listen (Week 6)	Read + listen (Week 2)
<i>The Witches of Pendle</i>	Read + listen (Week 6)	Read (Week 2)	Listen (Week 4)

Participants

Thirty-five Japanese students of English literature from a medium-sized private university in Kyushu, Japan, completed all aspects of the study. The ages of the 32 females and 3 males ranged from 18 to 21 years old. They had studied English for 7.5 years on average (including 6 years at junior and senior high school). The study began with 68 subjects, but 33 were omitted due to absence or incomplete data. The 35 subjects that saw the study through to its conclusion had been randomly assigned to three experimental groups. In Group A, there were 12 subjects from a 1st-year reading skills class; in Group B, there were 14 subjects from another 1st-year reading skills class; and in Group C, there were 9 subjects from a 3rd-year speaking skills class. All the subjects had pre-intermediate- or intermediate-level competence in English. This was determined by their classwork and homework assignments, as well as by two standardized tests: a 90-item Vocabulary Levels Test (Nation, 2001) and the paper-based version of the Test of English as a Foreign Language (TOEFL).

To test for differences in proficiency between the groups, we administered a combined test of four versions of the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) at the 2,000-word level. Group A's mean score was 64.83 ($SD = 9.3$), Group B's mean was 63.14 ($SD = 7.9$), and Group C's mean was 63.56 ($SD = 7.9$). There was no significant difference between the groups, $F(2, 32) = 0.14, p = .87$. The means of the subjects' most recent TOEFL scores were as follows: Group A, $M = 454$ (range: 407–483); Group B, $M = 448$ (range: 390–483); and Group C, $M = 460$ (range: 420–510).

The subjects were initially told that they would take part in a vocabulary-learning strategies program in which they would read and listen to some stories and that by using background knowledge, context, and co-text, they were to try to infer the meanings of any unknown words. They were also told that after reading and listening to a story, they would have to write some brief comments on their impressions of the experience and on how they felt about the content of the stories.

Materials and Design

The approach taken in this study was to use graded readers that were well within the subjects' current reading-ability level (i.e., texts in which 96% to 99% of the running words were already known). This would constitute ideal conditions for successfully inferring the meanings of unknown words from context (Laufer & Sim, 1985). The test items were embedded within the reading and listening texts. A 400-headword graded reader should not have presented any major lexical problems for the pre-intermediate- and intermediate-level subjects. In this way, it could be assumed that the surrounding co-text for the test items would be familiar, and therefore investigating the rate of acquisition that took place based solely on the test items could proceed. Three graded readers from the 400-headword, high-beginner level of the Oxford Bookworms Library were selected: *The Elephant Man* (Vicary, 1989), a true and tragic story set in 19th century England; *One-Way Ticket* (Bassett, 1991), a contemporary, human-interest collection of adventures on European trains; and *The Witches of Pendle* (Akinyemi, 1994), a true and dark story set in 17th century England. Prior to the study, all the copies of *The Elephant Man*, *One-Way Ticket*, and *The Witches of Pendle* in their original graded-reader form held at the university library were removed along with the original audio recordings. It was further determined that none of the subjects had read or listened to these stories before, nor had they seen the movie version of *The Elephant Man*.

Rationale for the use of substitute words. For the purposes of this study, adjustments were made to the texts of each story. The spellings of the 28 test items in each of the three books (total 84) were changed, replicating the design reported in Waring and Takaki (2003). Henceforth called *substitute words*, these words refer to the change in spelling of an already known word representing a common concept. For example, the words *happy*, *book*, and *skin* from *The Elephant Man* are rendered *mird*, *hoult*, and *labin* respectively in their substitute forms in the texts and tests. Words being symbols of meanings, a change in the symbol (its spelling), provided it conforms to normal spelling and collocational conventions, has both construct and face validity as it represents the matching of a new form for a given concept (i.e., learning a word in the traditional sense). As Nation (2001) noted, “at the simplest level, the unknown word

may represent a familiar concept and so the new label for that familiar concept is being learned” (p. 240). In a recent study on the effects of reading and writing on vocabulary knowledge, Webb (2005) used a similar approach by replacing target words with nonsense words.

Controlling the word-frequency variable. Other than Horst et al. (1998), Saragi et al. (1978), and Waring and Takaki (2003), few studies have investigated what types of words are learned in the reading treatment. Moreover, a single gain figure is generally given for the total number of words learned, irrespective of whether the words appeared frequently or not in the reading material. The present study, however, controlled for the word-frequency variable, in the hope that it would lead to greater accuracy in determining how many times a word needs to be met in reading and listening for it to be acquired. Therefore, in addressing Research Question 3 (*Are the subjects more likely to learn a word if they meet it more often?*), it was necessary to select words of differing frequencies of recurrence. In addition, it was necessary to decide what types of words should be selected. Nouns and adjectives were chosen because they are generally easier to guess than adverbs (Higa, 1965; Laufer, 1997; Rodgers, 1969). Verbs were not selected because they appear with their inflections and in various tenses, which can make it difficult to determine whether the word is known and to ascertain how frequently the word type has occurred in the text. Moreover, in order to get reasonably reliable data, it was necessary to test at least 25 words that the subjects would have to infer from context.

After looking at the recurrences of words in several 400-headword-level graded readers, *The Elephant Man*, *One-Way Ticket* and *The Witches of Pendle* were selected as the most appropriate titles for this study because the distribution frequencies in these titles had a good spread of words at different frequency bands. Each band had 7 test words. The frequency bands emerged from the natural frequency occurring in these books. The 28 words—seven words from four frequency bands—from each book were replaced with different spellings to ensure the words were unknown (the substitute words). Seven words occurred between 15–20 times in a given book; seven words appeared 10–13 times; seven words, 7–9 times; and seven words, 2–3 times. When more than seven words were in a given frequency band, the words were chosen randomly. This configuration of frequency groups and substitute words also ensured that a satisfactory coverage rate of running words could be maintained, as indicated in Table 2.

Table 2. *Lexical coverage of the running words by recurrences and types*

Text	Running words	Recurrences of test items	Coverage of running words (%)	Types	Coverage by types (%)
<i>The Elephant Man</i>	5,415	272	95.0%	574	95.1%
<i>One-Way Ticket</i>	5,522	272	95.0%	569	95.1%
<i>The Witches of Pendle</i>	5,765	264	95.4%	651	95.7%

The coverage rates in Table 2 refer to the percentage of the total running words assumed to be known by the subjects. For example, for *The Elephant Man*, 5,143 (which is 5,415 subtracted by 272 total recurrences for the 28 test items) of the 5,415 words in the book makes 95% coverage. When calculating the percentage of coverage by types, we calculated the total number of types minus the 28 types used as substitute words (i.e., 574 - 28 = 546) and then divided it by the total of types, which resulted in 95.1%.

In calculating the above coverage rates, as has been mentioned, it was assumed that because they

were meeting 400-headword-level texts, the pre-intermediate- and intermediate-level subjects would know all the other words. Clearly, however, this would not be true for all subjects, and for all words, especially considering the range of subjects' proficiency.

It should also be noted that as the subjects read and listened to the stories, many of the high-frequency substitute words would soon be recognized and learned as they got further and further into the narrative, thus the coverage rate would steadily increase. See Appendix A for the list of the substitute words and their English equivalents.

Instruments

In addressing Research Question 4 (*Are there significant differences in acquisition rates depending on whether the test is a multiple-choice test, or a meaning-translation test?*), separate tests were required in order to measure different types of word knowledge. Following Waring and Takaki (2003), two tests were selected, namely, a multiple-choice (prompted recognition) test and a meaning-by-translation (unprompted recognition) test to assess various levels of word knowledge.

The two tests were extensively piloted with a group of 40 subjects of similar ability and background who were not part of the main study. The aim of the piloting was to confirm that the test words were pronounceable for Japanese subjects, that the tests contained enough words, and that the stories were not too long and could be read or listened to in about 1 hour.

The multiple-choice test was a standard, prompted recognition four-choice test with the correct meaning and three distracters. An *I do not know* option was added to allow subjects to indicate when they did not know an item so as to reduce the effect of guessing. The subjects were asked to circle the words they thought were nearest to the substitute words' meanings. These choices were the same part of speech. For example, the substitute word *grift* means *leg*. *Leg* is a concrete noun, so the four choices were concrete nouns. Care was taken to ensure that the distracters came from different semantic sets so as to allow small amounts of knowledge to be demonstrated (Donkaewbua, 2008; Joe, 1994, 1998; Joe, Nation, & Newton, 1996). A sample extract from the test appears in Appendix B.

The meaning-translation test presented the 28 substitute words in a list. The subjects were asked, "What do these words mean? Write the meaning in Japanese." Subjects were required to either provide the exact meaning or give a plausible approximate answer, such as a near synonym. For instance, the exact meaning of *hoult* in *The Elephant Man* is *book* ("hon" in Japanese). However, if subjects wrote *story* ("monogatari" in Japanese), they would be given credit. Thus, half marks were given for partial knowledge of the meanings of the substitute words. Moreover, to further encourage a response, subjects were given two chances to provide an answer. A sample extract from the test appears in Appendix B. Finally, in order to prevent the transfer of knowledge from one test type to another, the meaning-translation test was given first and the multiple-choice test given second.

Procedure

The subjects were told that the main purpose of this "vocabulary-learning strategies program"

(i.e., the study) was to determine whether they learn vocabulary better from reading, reading-while-listening, or listening to stories. It was explained that they would read and listen to three stories in which certain words had been changed. The rationale for, and examples of substitute words were explained, but none of the actual test items were cited. They were told to enjoy reading and listening to the stories and to do their best to guess the meanings of the substitute words. Afterwards, they would have to answer some questions. Neither dictionary use nor note-taking was allowed. Moreover, during the reading and listening sessions, no questions on the content of the stories were permitted. On completion of the whole program (the study), the researcher would individually inform the subjects which mode was best for them when acquiring new vocabulary in English. The research schedule in detail is set out in Table 3.

Table 3. *Research schedule in detail*

Group	Week							3 Month Delay
	1	2	3	4	5	6	7	
A (n=12)	PDVT	S1 L	Test	S2 R	Test	S3 R + L	Test S3-2	Tests 1-3 2-3 3-3
		Test	S1-2	Test	S2-2	Test	Essay	
		S1-1		S2-1		S3-1		
B (n=14)	PDVT	S3 R	Test	S1 R + L	Test	S2 L Test	Test S2-2	Tests 1-3 2-3 3-3
		Test	S3-2	Test	S1-2	S2-1	Essay	
		S3-1		S1-1				
C (n=9)	PDVT	S2 R + L	Test	S3 L	Test	S1 R	Test S1-2	Tests 1-3 2-3 3-3
		Test	S2-2	Test	S3-2	Test	Essay	
		S2-1		S3-1		S1-1		

Note. PDVT = profile data vocabulary test; R = Reading-only mode; R + L = Reading-while-listening mode; L = Listening-only mode.

S1 (Story 1): *The Elephant Man*;

S1-1: Story 1, Posttest 1;

S2 (Story 2): *One-Way Ticket*;

S2-2: Story 2, Posttest 2;

S3 (Story 3): *The Witches of Pendle*;

S3-3: Story 3, Posttest 3.

The reading-only mode and the reading-while-listening mode. For the purposes of this study, the full texts of *The Elephant Man*, *One-Way Ticket*, and *The Witches of Pendle* with their substitute words were printed and put into book form. In the reading-only mode and the reading-while-listening mode, the subjects were asked to read (and listen to) the stories as usual and enjoy them. Short written introductions to the stories (150 words approximately) were given in each of the three modes; however, these words were not counted in the figures for the main experiment. These introductions were added to provide schematic background for each book.

Furthermore, to control for consistency of coverage rate, key words in each story that fell outside the 400-headword range and that appeared in the books' glossaries were written on the chalkboard with their Japanese translations. Subjects could consult these lists (8 words per story) if they needed to as they read or listened. A short, verbal preamble was given for each story to orientate the subjects towards its topic, setting and background, but without mentioning anything about the storyline or characters. Maps were used to help set the scene when necessary.

The listening-only mode. The full texts of the three stories were read aloud and recorded on audiocassette by the second author. Care was taken to ensure that the narration was as clear and as natural as possible. Piloting determined that a mean speech rate of 93 words per minute (wpm) was appropriate for the subjects as they had never before listened to a long narrative on audiocassette in English (e.g., Hirai, 1999). These recorded versions of the stories had a mean duration time of 63 minutes. In the listening-only mode, the subjects' supplementary-text support was a short written introduction (150 words approximately) and a set of six or seven illustrations (without captions) both from the original book. Subjects were asked to listen to the audiocassette and to look at the pictures while listening to help them follow the narrative. There was a mid-session interval of 3–4 minutes during which the subjects could stand up and stretch. Because of the long duration time of the listening treatment, it was hoped that general fatigue or attention-span limitations would not have a detrimental effect on word learning. Such long listening sessions are not uncommon, however, especially in commercial testing and when listening to university lectures. If we compare, for example, the current generation TOEFL, the Internet Based Test (iBT), we find that it has a listening section that is between 60–90 minutes long and contains up to six lectures and three conversations.

Data Collection

After reading or listening to the stories, as mentioned, the two tests were given in this order: (a) meaning-translation test, and (b) multiple-choice test. These instruments formed the test set. The test set was administered three times: Posttest 1, immediately after the story reading or listening sessions; Posttest 2, 1 week later; and Posttest 3, 3 months later. The test items used in each administration were the same, but the item order was rotated so as to control for a potential learning effect from the tests. All of these test administrations were unannounced. The subjects took the tests without seeing or hearing the story again, and they never met the substitute words again.

In the listening-only mode, because the subjects had not read but had heard the substitute words in a recording of the story, the test instrument for this mode necessitated the recording of the prompts on audiocassette. It was considered important to test the subjects in the way that they had learned so as to maintain reliability of data. Thus, at test time, the subjects listened to the prompts and marked their responses on paper. The mean duration time of the listening test set was 20 minutes. The reading-only and the reading-while-listening test sets were the same instrument and took subjects approximately 10 minutes to do.

At the beginning of Posttest 1 (as shown in Table 3), the time taken to read or listen to the story was written down by each subject. A questionnaire asked subjects to indicate on a six-point attitude scale (5–0): (a) if they thought the story was easy or difficult to read or listen to; (b) if they knew most or only a few the words; (c) if they understood most or only a little of the story; and (d) if they thought the story was interesting or not. An open-ended question asked what they thought of the story.

At the conclusion of the reading and listening (story) sessions, and on completion of Posttest 2 in Week 7, the subjects were asked to write a brief essay describing how they felt about the program (i.e., the study). In so doing, they were asked to consider these three points: (a) the story

they liked the most, and why; (b) the story that was easiest, and why; (c) the mode they preferred, and why. The data collected from the subjects' responses were examined in order to address Research Question 5 (*Do the subjects prefer to read only, read while listening, or listen only to stories?*).

Marking

On the multiple-choice test, correct answers were given one point each. On the meaning-translation test, correct answers were given one point and a word with a similar meaning was given a half point. For example, if the test word's correct answer was *book*, one point was given, but if the subject supplied *story*, because it is a near synonym, a half point was awarded. A total of only 41 (0.46%) of all the possible responses were given a half point for the 35 subjects over the three test administrations and thus did not significantly affect the overall results. Moreover, 99.1% of the subjects used only one blank to provide a translation. The first author and a native Japanese speaker scored the test.

Results and Discussion

Research Question 1: Do the subjects learn more vocabulary from reading, reading while listening, or listening to stories?

Table 4 summarizes the data for the three input modes and the two test types at the immediate posttest (i.e., at Posttest 1). The data are presented graphically in Figure 1. Data for the delayed tests are reported later. The data by test type are reported first. All standard deviations are in parentheses. Across all texts, the mean scores for the multiple-choice (MC) test are: reading-only mode 12.54 (5.03), reading-while-listening mode 13.31 (3.90), and listening-only mode 8.20 (2.82). The mean scores for the meaning-translation test are: reading-only mode 4.10 (4.02), reading-while-listening mode 4.39 (3.29), and listening-only mode 0.56 (1.13).

Table 4. Mean scores for all texts for the two tests by the three input modes at Posttest 1

Text	Reading-only		Reading-while-listening		Listening-only	
	MC	Translation	MC	Translation	MC	Translation
<i>Elephant Man</i> <i>n</i> = 12	18.67 (3.61)	8.11 (4.41)	12.67 (4.15)	4.06 (2.90)	7.83 (1.52)	0.00 (0.00)
<i>One-Way</i> <i>n</i> = 14	9.58 (3.20)	1.71 (2.18)	11.25 (3.52)	2.13 (1.87)	7.93 (2.64)	0.82 (1.49)
<i>Witches</i> <i>n</i> = 9	11.14 (3.66)	3.57 (3.11)	15.50 (3.06)	6.54 (3.23)	9.11 (4.23)	0.89 (1.05)
All Texts <i>n</i> = 35	12.54 (5.03)	4.10 (4.02)	13.31 (3.90)	4.39 (3.29)	8.20 (2.82)	0.56 (1.13)

Note. Standard deviations are in parentheses. Max = 28.

The MC test results for the reading-while-listening mode across all texts indicate that an impressive 48% (13.31) of the 28 words were learned (compare gains of 22% in the study by Horst et al., 1998). MC gains made in the reading-only mode were similarly impressive standing at 45% (12.54). Gains made in the listening-only mode, however, were less remarkable standing

at 29% (8.20).

Of the two tests, the meaning-translation test is probably the one that most closely indicates whether a subject actually knew the meaning of the word while reading and listening. This is because it shows that the subject is not only capable of recognizing the word but can also assign a meaning to it without being prompted. In Table 4, the meaning-translation test results across all texts show that 16% (4.39) of the 28 words were learned in the reading-while-listening mode. This rate of acquisition is followed closely in the reading-only mode, which yielded gains of 15% (4.10) of the 28 target words. This reading-only rate matches that in the Waring and Takaki (2003) study, in which the meaning-translation test scores showed that 18% of the 25 target words were learned. In the present study, gains in the listening-only mode were minimal with only 2% (0.56) of the 28 words learned.

Table 4 also displays the mean scores of the input modes by text and test type, and these scores help indicate which modes were easier or harder for the subjects. We find that of the 28 new words presented in this study, the most outstanding gains of all were those achieved when the subjects read *The Elephant Man* (18.67 on the MC test and 8.11 on the translation test). These were followed by the reading-while-listening gains for *The Witches of Pendle* (15.30 on the MC test and 6.54 on the translation test). Conversely, it can be seen that on listening-only to *The Elephant Man*, the subjects did not register any perceptible gains on the translation test. With regard to *One-Way Ticket*, it can be seen that most of the test scores across the three input modes were quite close, with the test scores for listening-only being marginally better than those attained when listening-only to *The Elephant Man*. Interestingly, although the story was generally reported not liked, the test scores for listening only to *The Witches of Pendle* yielded the best overall results in this mode (9.11 on the MC test and 0.89 on the translation test).

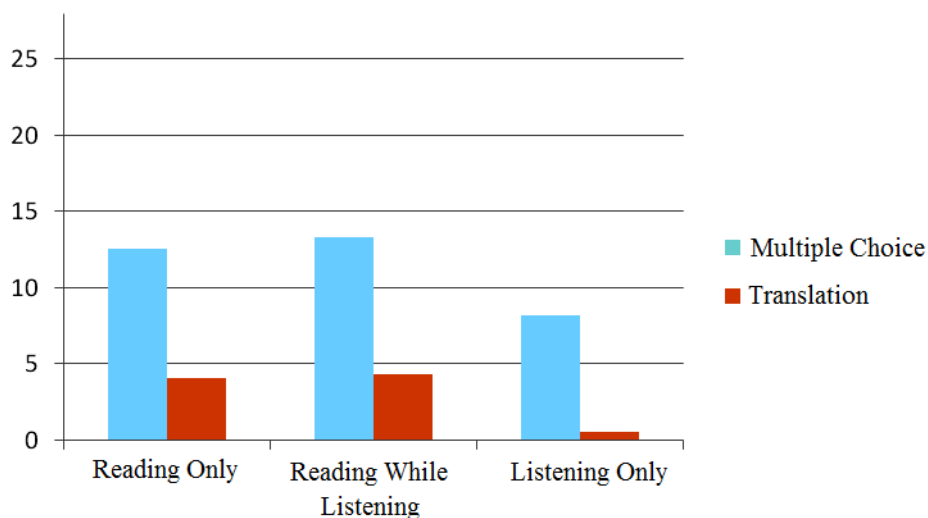


Figure 1. Overall mean scores for the two tests by the three input modes at Posttest 1.

ANOVA administrations revealed significant differences between the MC tests and the meaning-translation tests for the three modes (reading-only, reading-while-listening, and listening-only). Significant differences in test scores emerged in the three modes for the MC test, $F = 13.32$, $p < .001$, and the meaning-translation test, $F = 16.38$, $p < .001$.

To determine where the differences between the tests were, *t* tests were conducted for the two tests by three input modes. The results are presented in Table 5. There was a significant difference between the reading-only and listening-only modes, as well as for the reading-while-listening and listening-only modes for both test types. This suggests that it is far more difficult to pick up words from listening-only than from either the reading-only or reading-while-listening modes. There was, however, no significant difference between reading-only and reading-while-listening modes.

Table 5. *T-test data for the two tests by three input modes at Posttest 1*

Test	Reading-only and listening-only	Reading-while-listening and listening-only	Reading-only and reading-while-listening
MC	6.93*	7.23*	0.86
Translation	5.66*	7.24*	0.41

Note. * $p < .05$.

Reading-only mode versus reading-while-listening mode. The scores the subjects attained in these two modes were similar across the tests. The mean test scores for the three books varied relatively little depending on the test type (even after 3 months). Given the almost equal expected learning outcome from each of these modes, it would seem that the selection of preferred input mode should rest with the learner.

Listening-only mode. It seems rather obvious that the listening-only mode should be the most difficult to acquire new vocabulary from (especially given the length of the listening task). In this study, the results of the meaning-translation test at the immediate posttest for the listening-only mode showed that only 2% (0.56) of the 28 target words were learned (compared with 15% and 16% in the other two modes). Moreover, as we shall see in detail later, when asked which input mode they preferred, 0% of the subjects chose listening-only.

The subjects, it seems, displayed a critical lack of familiarity with spoken English. As they listened to the story, they had to pay constant attention to a stream of speech whose speed they could not control. Because they were incapable of processing the phonological information as fast as the stream of speech, they may have failed to recognize many of the spoken forms of words that they already knew in their written forms.

A possible reason for this is that the subjects' phonological knowledge of English varied from the phonological system employed by native speakers. The Japanese language has a different syllable structure to English and is often said to be mora-timed; therefore, Japanese learners may expect to hear words pronounced in this manner and thus may have considerable problems interpreting spoken English. McArthur (2003) claimed that Japanese learners have great difficulty in speaking and listening to English because of this "tendency not only to pronounce English in terms of Japanese syllable structure but also to adapt English words syllabically into Japanese" (p. 21).

A second reason might have been a lack of skill in detecting word boundaries in connected speech (i.e., skill in the lexical segmentation of the input signal). On reviewing the comments made by the subjects regarding the listening-only mode, it became apparent that a major

challenge for them was negotiating the seamless nature of connected speech. Because of the way one word runs into the next seamlessly “without any little silences between the spoken words compared with the way there are white spaces between written words” (Pinker, 1994, p. 159), subjects may have found it particularly difficult to tell where one word ended and the next began. In terms of second-language listening, Field (2003) characterized the lexical segmentation of streams of speech as “arguably the commonest perceptual cause of breakdown of understanding” (p. 327).

A third reason might have been that the subjects were required to listen at a coverage rate (95%) that was set for reading and not listening. The data suggest that the coverage rate was too low for the listening-only mode, rendering the task of inferring the meanings of the 28 target words as too great a challenge. Although no statistical data was provided, Nation (2001) claimed that “it is likely that for extensive listening the ratio of unknown words to known words should be around 1 in 100” (p. 118).

Research Question 2: At what rate is this new vocabulary knowledge learned, and at what rate does it decay?

The decay data for the three input modes at the three test times are shown in Table 6. Decay data for each test are shown graphically in Figures 2 and 3. These data show relatively little decay from their initial learning.

Table 6. *Decay data by input mode over the three test periods*

Mode	Immediate posttest		One-week delay		Three-month delay	
	MC	Translation	MC	Translation	MC	Translation
Reading-only	12.54 (5.03)	4.10 (4.02)	12.46 (4.25)	2.34 (2.39)	11.37 (3.10)	0.97 (1.47)
Reading-while-listening	13.31 (3.90)	4.39 (3.29)	12.37 (3.41)	1.83 (1.94)	12.14 (2.86)	1.14 (1.32)
Listening-only	8.20 (2.82)	0.56 (1.13)	9.06 (2.65)	0.74 (1.12)	10.09 (2.72)	0.37 (0.72)

Note. Standard deviations are in parentheses. Max = 28.

Table 6 shows that there was relatively little decay over a 3-month period in the scores for the reading-only, reading-while-listening, and listening-only modes for the two test types. The scores remained about the same irrespective of the mode or the test, except for the meaning-translation test scores, which dropped more considerably or stayed very low in all three modes. Thus, the knowledge needed to complete a translation test seems to be far higher than simply selecting the best answer on an MC test.

ANOVA administrations were carried out to determine if there were any significant differences between the scores across the three data times for the two tests for each mode. Here are the results: on the translation test, the reading-only mode, $F = 11.11$, $p < .01$, reading-while-listening, $F = 19.52$, $p < .01$, and listening-only $F = 0.88$, $p = .42$; and on the MC test, the reading-only mode, $F = 0.76$, $p < .50$, reading-while-listening, $F = 0.84$, $p = .43$, and listening-only, $F = 4.20$, $p < .05$.

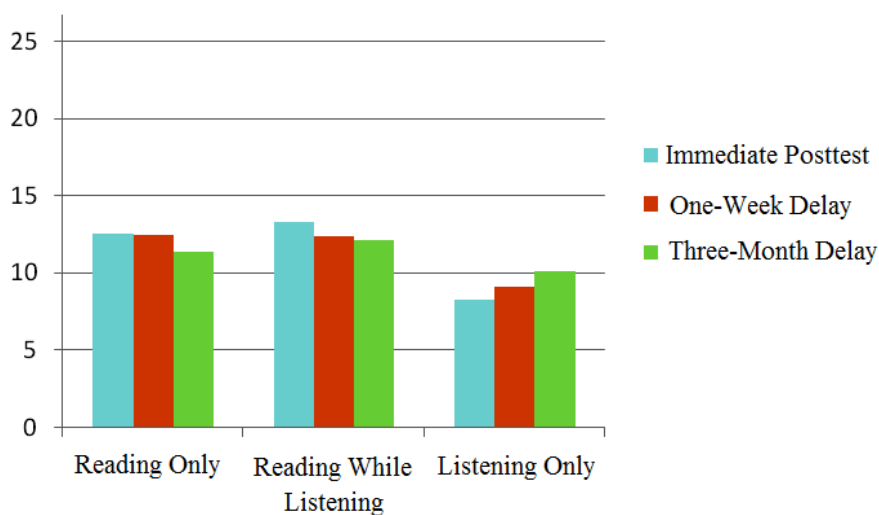


Figure 2. Decay data for the MC test over the three test periods.

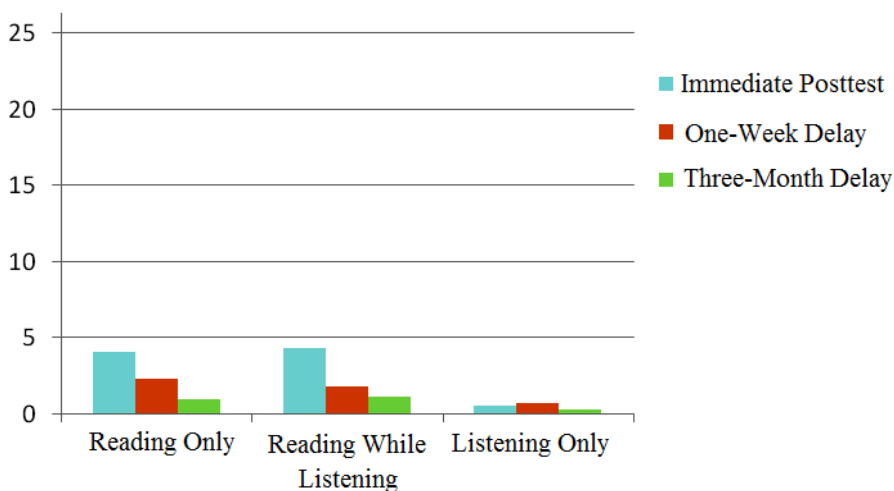


Figure 3. Decay data for the translation test over the three test periods.

The ANOVA scores suggest that there were significant differences for many of the translation tests, but not for the MC tests; so *t* tests were performed on the data to determine where the differences were. Table 7 presents the data for the translation test, and Table 8 presents the data for the MC test. The translation test scores tended to drop over time while the MC test scores did not, for both the reading-only and reading-while-listening modes. However, in the listening mode, the scores fluctuate, but given the small data set, the small number of subjects, and the possibility of floor effects, we should not read too much into these data.

Table 7. The t-test scores for the three modes across the three data times for the translation test.

Mode	Immediate posttest → One-week delay	Immediate posttest → Three-month delay	One-week delay → Three-month delay
Reading-only	3.67*	5.41*	5.24*
Reading-while-listening	5.83*	7.17*	2.70*
Listening-only	.89	.94	1.8

Note. * $p < .05$.

This seems to suggest that the prompted-meaning recognition knowledge is better retained than the unprompted knowledge. In other words, learners are much more likely to forget the meaning of a word if they are not primed for its meaning. This suggests that teachers should ensure that the learners meet words very often and that they be primed to remember words before reading a passage again.

Table 8. *The t-test scores for the three modes across the three data times for the MC test.*

Mode	Immediate posttest →	Immediate posttest →	One-week delay →
	One-week delay	Three-month delay	Three-month delay
Reading-only	0.17	1.87	2.01
Reading-while-listening	1.98	2.26	0.50
Listening-only	2.05*	3.69*	2.24*

Note. * $p < .05$.

It is noted that some of the mean scores in Table 6 appeared to increase over time without further exposure. This is not an uncommon phenomenon and has been shown in other studies (e.g., Waring & Takaki, 2003). Often this is because the *true means* vary by the size of the standard deviation and while it may appear that the mean scores went up, it is likely that no real increase in knowledge was gained over time. Another possible explanation for this is found in a recent study of the rate of learning collocation from graded reading (Waring, 2008). This study shows that certain subjects retain knowledge of partially known words learnt in their reading and associate that knowledge with other words in the lexicon as they continue to learn the language. It seems that the subjects' developing systemic knowledge of words over time has a facilitating effect on the entire lexicon, and thus has a knock-on effect on all partially known words (even substitute words), as has also been found in the present study.

Research Question 3: Are the subjects more likely to learn a word if they meet it more often?

The data for the effect on learning as influenced by a word's frequency of recurrence are presented in Table 9. These are the mean scores across the three books in each input mode. The table is read as follows: On the MC test for the reading-only mode, of the seven words that were met 15–20 times in each of the stories, 4.29 (2.0) of them were recognized; of the seven words met 10–13 times, 2.86 (2.3) were recognized; of the seven words met 7–9 times, 3.14 (1.4) were recognized; of the seven words met 2–3 times, 2.26 (1.2) were recognized; and so on.

Table 9. *Data by word frequency of recurrence at Posttest 1*

Test	Reading-only				Reading-while-listening				Listening-only			
	15–20	10–13	7–9	2–3	15–20	10–13	7–9	2–3	15–20	10–13	7–9	2–3
MC	4.29 (2.0)	2.86 (2.3)	3.14 (1.4)	2.26 (1.2)	4.43 (1.5)	4.03 (2.1)	3.23 (1.2)	1.63 (0.8)	2.06 (1.5)	2.54 (1.6)	2.23 (1.2)	1.37 (1.0)
Translation	1.97 (1.9)	1.39 (1.6)	0.70 (1.0)	0.04 (0.1)	1.86 (1.5)	1.44 (1.5)	1.01 (1.2)	0.07 (0.3)	0.19 (0.5)	0.11 (0.4)	0.14 (0.4)	0.11 (0.3)

Note. Standard deviations are in parentheses. Max = 7.

By and large, the data show that the more frequently an item is met, the more chance it has of

being learned. The data also show that the scores tend to decrease depending on the test type, with the meaning-translation test scores considerably lower than those on the MC test.

The frequency-of-recurrence data are valuable because they can indicate how frequently a word should be met in order to learn it in the three modes. The data in Table 9 show that the words met more frequently were more likely to be known at the immediate posttest in each mode. This finding was consistent across the two test types with mean scores dropping as recurrence frequency diminished.

ANOVA administrations were carried out to determine if there were any significant differences between the scores across the four frequency bands for the two tests for each mode. Here are the results: on the translation test for the reading-only mode, $F = 24.14$, $p < .01$, for reading-while-listening, $F = 20.80$, $p < .01$, and for listening-only, $F = 0.31$, $p = .82$; and on the MC test for the reading-only mode, $F = 24.63$, $p < .01$, for reading-while-listening, $F = 52.02$, $p < .01$, and for listening-only, $F = 4.67$, $p < .01$.

Table 10. *T-test scores for the translation test for each frequency band by input mode at Posttest 1*

Mode	15–20	15–20	15–20	10–13	10–13	7–9
	vs. 10–13	vs. 7–9	vs. 2–3	vs. 7–9	vs. 2–3	vs. 2–3
Reading-only	2.57*	5.56*	6.24*	3.99*	5.17*	4.18*
Reading-while-listening	1.37	2.92*	7.18*	2.00*	5.32*	4.33*
Listening-only	1.22	.49	.82	.33	.00	0.33

Note. * $p < .05$.

Table 10 presents the *t*-test data for each input mode analyzed between each frequency band for the translation test, and Table 11 presents the same data for the MC test in order to show where the differences were.

Table 11. *T-test scores for the MC test at each frequency band by input at Posttest 1*

Mode	15–20	15–20	15–20	10–13	10–13	7–9
	vs. 10–13	vs. 7–9	vs. 2–3	vs. 7–9	vs. 2–3	vs. 2–3
Reading-only	3.50*	4.02*	6.42*	0.79	1.50	3.90*
Reading-while-listening	1.36	4.19*	9.49*	2.34*	6.34*	5.74*
Listening-only	1.79	.56	2.05*	1.01	3.43*	3.43*

Note. * $p < .05$.

As one would expect, the more frequently met words were better learnt than the less frequently met words. Both tests showed significant decay between each frequency band. This did not happen for the listening-only mode probably because of floor effects.

Table 9 also confirms differences in the acquisition rates by frequency of recurrence by input mode. The MC tests for reading-only and reading-while-listening modes yielded the following rates for the 7–9 frequency band: 45% (3.14/7) and 46% (3.23/7) respectively. However, the meaning-translation test rates for the 7–9 band were far lower: 10% for reading-only and 14% for reading-while-listening.

In the listening-only mode, according to the MC test results, even having met a word 10–13

times, there is a less than 36% (2.54/7) chance that the word can be recognized. Furthermore, meaning-translation test results indicate that 10–13 meetings of a word will yield only a 1.5% (0.11/7) chance that its meaning will be understood when encountered again. Moreover, only 3% (0.19) of the 7 words met 15–20 times in the texts were acquired. The data suggest that the acquisition of words through listening is considerably slower than from reading, and as such more recurrences of words are needed for acquisition (as defined by a correct score on the meaning-translation test) to take place.

Ultimately, this suggests that there is little or no chance a new word will be picked up from listening unless the word is met considerably more than 20 times. Extrapolation of these data shows that maybe 50 or even 100 meetings may not be enough to acquire a word's meaning from listening-only. As has recently been shown, even partial knowledge such as the ability to recognise a word's form is hard to pick up from listening alone (Donkaewbua, 2008). It also suggests that far more listening than reading needs to be done for vocabulary learning through extensive exposure. It should also be noted that in this study more uptake of vocabulary might have been possible if the listening treatment had been in shorter, more manageable sessions.

The reading-only mode data in this study replicate the Waring and Takaki (2003) findings, which showed that (a) unless words are met a sufficient number of times and (b) are met again soon after reading, then the word knowledge gained will decay. Recent research indicates that a sufficient number is likely to be much higher than 7–9 times for long term retention, and in fact may be closer to 30–50 times or higher (Waring, 2008) for new words met through graded reading.

Research Question 4: Are there significant differences in acquisition rates depending on whether the test is a multiple-choice test or a meaning-translation test?

The aim here is to determine if there are significant differences between the test types, which in turn can tell us if one type of test is more difficult than others, or to put it another way, do the tests measure different levels of word knowledge? This has considerable implications for the type of test used in this kind of research. There were significant differences between each test within each input mode as shown by the data in Table 4 and Table 12 and the ANOVA scores. For the reading-only mode there was a significant difference between the two test types, $F = 57.17, p < .01$, for the reading-while-listening mode, $F = 68.14, p < .01$, and for the listening-only mode, $F = 208.49, p < .01$.

The t -test results (based on adjusted alpha) in Table 12 show that the scores differed significantly depending on which test was taken. These data show that the test types employed by researchers that aim to assess gains from incidental vocabulary acquisition matter greatly.

Table 12. *T-test results between test types at Posttest 1*

Mode	MC test vs. translation test
Reading-only	20.4**
Reading-while-listening	23.1**
Listening-only	17.8**

Note. ** $p < .01$.

Table 13 presents the *t*-test data for differences between each of the various input modes for each test. While there was no significant difference between reading-only and reading-while-listening modes for the two tests, four of the *t* tests showed significant differences. There were significant differences between the listening-only and reading-only scores, and listening-only and reading-while-listening scores on each test type.

Table 13. *T-test results between the input modes at Posttest 1*

Test	Reading-only vs. reading-while-listening	Listening-only vs. reading-only	Listening-only vs. reading-while-listening
MC	0.86	5.70*	7.23*
Translation	0.41	5.67*	5.52*

Note. * $p < .05$.

In terms of test type, the MC test showed significant differences in the listening-only versus reading-only modes, $t = 5.70$, and the listening-only versus reading-while-listening modes, $t = 7.23$, but not in the reading-only versus reading-while-listening modes, $t = 0.86$. Similarly, the meaning-translation test showed significant differences in the listening-only versus reading-only modes, $t = 5.67$, and in the listening-only versus reading-while-listening modes, $t = 5.52$, but not in the reading-only versus reading-while-listening modes, $t = 0.41$.

In sum, the data show that the subjects picked up some words from their reading and listening experiences in this study, but far fewer words were picked up in the listening-only mode compared with the other two modes. The data for the reading-only mode replicate that of Waring and Takaki (2003), which found that on the unprompted translation test few words were picked up and retained, but if measured by an MC test, some words were known. This suggests that the recognition of words from reading is acquired before a meaning can be produced on a translation test.

Research Question 5: Do the subjects prefer to read only, read while listening, or listen only to stories?

Table 14 presents the data from the questionnaire that was administered immediately after the reading and listening sessions for each of the three stories. It is evident that the subjects were most comfortable with the story met in the reading-while-listening mode. They were also quite comfortable with the story met in the reading-only mode. However, the story they met in the listening-only mode was clearly the least favored with almost all scores below the median of 2.5.

Table 14. *Mean scores from the questionnaire*

Mode	Was it easy to read or listen to?	Did you know most of the words?	Did you understand the story?	Was the story interesting?
Reading-only	2.86	2.97	3.17	3.40
Reading-while-listening	3.34	2.86	3.34	3.54
Listening-only	1.91	2.34	2.06	2.60

Note. Max = 5.

Table 15 presents the data from the written comments extracted from the subjects' brief essays.

These essays were written at the conclusion of the reading and listening sessions, and on completion of Posttest 2 in Week 7. *The Elephant Man* was generally perceived to be both the most interesting and the easiest book.

The essay data revealed that the great majority of subjects were inclined towards the reading-while-listening mode (72%). In addition, while a sizeable minority was in favor of the reading-only mode (28%), no subjects indicated unequivocally that they preferred the listening-only mode. These data are supported by their actual performance in each mode. The all-texts scores for the meaning-translation test at Posttest 1 (see Table 4) has the reading-while-listening mode ranked first with 16% of the words learned, the reading-only mode ranked second with 15% of the words learned, and listening-only lies in third place with 2% words learned. The data in Table 14 also point to listening-only being the most difficult, the least pleasurable, and the most difficult to understand. This would most likely have rendered the story also less interesting. The reading-only and reading-while-listening mode ratings though fared considerably better with all the scores above the median of 2.5.

Table 15. Data for the written comments in essays

	<i>The Elephant Man</i>	<i>One-Way Ticket</i>	<i>The Witches of Pendle</i>
Q1. Which book did you like the most?	26 (74%)	5 (14%)	4 (12%)
Q2. Which book was easiest?	21 (60%)	8 (23%)	6 (17%)
	Reading-only	Reading-while-listening	Listening-only
Q3. Which mode did you prefer?	10 (28%)	25 (72%)	0 (0%)

Note. $n = 35$.

Although not a research question in this study, it is nevertheless interesting to look at the subjects' responses to Items 1 and 2 in their short essays (i.e., the story they liked the most) and the story they thought the easiest (Table 15). It is clear that *The Elephant Man* was the most favored story by far (74%), followed by *One-Way Ticket* (14%), and then by *The Witches of Pendle* (12%). This pattern is repeated in the subjects' responses to which book they thought the easiest. By examining more closely the subjects' written comments regarding their favorite book, and which they considered the easiest, a broader picture begins to emerge of the type of material that students may readily engage with at an intellectual or emotional level.

From the data, it emerged that there was a good degree of intellectual and emotional involvement due to the stories being interesting, thought provoking, moving, funny or sad. It would seem that, as Elley (1989) argued, "attention levels are greatest when students are aroused by... such variables as novelty, humor, conflict, suspense, incongruity, vividness, and the like" (p. 185). All three stories possessed these variables to a greater or lesser extent.

Finally, on reviewing the subjects' reasons as to why they found a particular story the easiest, 75% reported that it was because the story was in their preferred mode, which, as we have seen,

was predominantly reading-while-listening, followed by reading-only, reflecting corresponding success rates on the tests. It would seem, therefore, that such over-riding preferences for mode would be worthy of teachers' consideration when planning lessons.

General Discussion

The results of the meaning-translation test at the immediate posttest show that the subjects were able to learn new words from context and that they learned most words in the reading-while-listening mode (4.39 of 28 words), followed by the reading-only mode (4.10 of 28) and then the listening-only mode (0.56 of 28). Moreover, the results from the meaning-translation and MC tests indicated that relatively little decay occurred over 3 months. However, the meaning-translation test scores dropped more considerably, albeit from a much lower starting point.

In terms of 3 months' retention of unprompted meaning, on average the subjects learned one new word from reading while listening to a graded reader, one new word from reading-only, and effectively no words from listening-only. In terms of the acquisition of new (previously unknown before exposure) vocabulary, this was quite a disappointing rate of return considering the effort involved. More encouragingly, however, the data from the MC test indicated higher learning and retention rates. This, in turn, suggests that some partial knowledge not accessed by the insensitive meaning-translation test was found to be known via the more sensitive (i.e., their knowledge was prompted) MC test.

The data also indicated that the more frequently a word is met, the more chance it has of being learned. It also suggests that unless the words are met a sufficient number of times and are met again soon after in subsequent reading or listening experiences, then the word knowledge gained will decay. A sufficient number is likely to be considerably higher than seven to nine times for long-term retention (Waring, 2008).

It was found that the type of instrument used to assess vocabulary gains in learning-from-context research had a great bearing on the degree of success deemed to have occurred. In this study, Table 4 shows that the lowest mean rate of uptake of new vocabulary as measured by the MC test was 29% (8.20 of 28 words in the listening-only mode). This was almost double the highest mean rate of uptake as measured by the meaning-translation test, which was found to be 16% (4.39 words in the reading-while-listening mode). Therefore, as Waring and Takaki (2003) pointed out, great care must be taken when selecting test types in studies of a similar design to the one undertaken here.

In terms of preferred input mode, reading-while-listening was considered the most comfortable by the majority of subjects; a sizeable minority favored reading-only, while no one explicitly favored listening-only. The vocabulary gains shown in the data mirrored these preferences. It would seem that for the majority of subjects in this study, reading while listening to a 400-headword-level graded reader narrated at 93 wpm promoted good understanding. Informal interviews with some of the subjects after the study revealed that a key reason for favoring the reading-while-listening mode was that the necessity of having to segment or chunk the text of the story as they read it was done for them by the narrator on the cassette. Consequently, it would

appear they had enough spare working-memory space to access the content more effectively, and in turn make better deductions of the meanings of the target words. This coincides with what Amer (1997) and Dhaif (1990) found in their studies.

However, as Goh (2002) pointed out, “in the case of advanced listeners, the bottom-up processes [of word recognition] are largely automatized...they do not need to spend time on matching sequences of sounds with written words in their mental lexicon” (p. 7). Accordingly, they would tend to direct their attention to making higher-level inferences (i.e., engaging in the utilization of already perceived or segmented information). In the present study, it was found that whereas the majority of subjects were comfortable with the reading-while-listening mode, more proficient subjects were not always inclined towards this mode.

Finally, while the familiar reading-only mode allowed subjects to keep to their own pace, and if necessary to back track without interruption, the subjects encountered considerable obstacles when trying to comprehend the story and substitute words they met in listening-only mode. Clearly, the inaccurate perception of the pronunciation of words and phrases is potentially a greater barrier in listening than in reading.

Implications for Teaching and Learning

This study has shown that relatively minimal growth and retention of new vocabulary occurs when reading a single graded reader, and thus points to the need for repeated encounters with words in a collection of graded texts at regular intervals. To ensure exposure to great amounts of written text, graded readers should form part of an extensive reading program and learners should endeavor to read approximately a book a week at coverage rates of 95% or more (Day & Bamford, 1998; Nation & Wang, 1999).

Learning vocabulary from listening. The results of this study also confirm learners’ potential difficulty with the listening-only mode. Although some of the contributing factors were outlined earlier, further research will have to be done to determine whether poor performance on the listening-only tests is a linguistic, testing, or language-processing problem. It is certainly clear, however, that teachers of Japanese learners of English should not assume that learners can listen at the same headword level at which they can read. This probably also applies to learners of English from other language backgrounds whose L1 phonological systems are markedly dissimilar to that of English.

Moreover, we could say, at least for these subjects, that because their reading level was substantially higher than their listening level, it would be wise for them to practice extensive listening at either (a) an easier graded-reader level than that at which they can read comfortably, or (b) at a slower speed of narration. The data also suggest that teachers should create extensive-listening tests to determine at what level students can listen comfortably rather than rely on tests based on reading ability. Lastly, if learners want to improve their aural perception of streams of speech, one bridge to proficiency in listening-only may be to do extended practice in the reading-while-listening mode first. Alternatively, learners could read the book first, then read-while-listening to it, and finally listen only. In this way, learners would be primed for the words when they listen to them.

Inferring meaning from context. As was done with the 35 subjects in this study, foreign-language learners should be provided with opportunities and guidance on how to capitalize on the incidental learning of vocabulary from their extensive reading and listening. As Nation (2001) pointed out, “inferring vocabulary meaning from context...is an essential strategy for developing reading comprehension and promoting lexical acquisition” (p. 240). Thus, if learners do a lot of reading and listening, there will be considerable cumulative enrichment of partially known words as well as the establishment of certain new words in their lexicons. Inferring the meanings of unknown words from context is therefore important both for coping with and learning unfamiliar words.

Limitations of the Study

This study examined data from only 35 subjects. Thirty-three other subjects had taken part at an earlier stage of the experiment, but for various reasons were not able to submit all the data. This suggests that in order to collect more reliable data, it is important to ensure that there is a larger cohort of subjects. A second limitation was that this study examined only Japanese learners. Therefore, learners from other language backgrounds should be investigated as well. A replication of this experiment would be welcomed. Thirdly, subjects were exposed to a mean of only 5,567 words in each input mode. Therefore, to gather more data on the effectiveness of learning vocabulary from reading and listening to stories in a foreign language, it would be better to devise studies that include multiple or longer texts in each mode. Lastly, the study assumed that the use of a 400-headword-level graded reader would provide no significant hindrance for the necessary conditions for inferring new words from context. As this was not precisely determined beforehand, it may have been a factor in the low learning and retention rates, especially in the listening-only mode.

Conclusion

This study has shown that relatively few new words are learnt from reading a graded reader as measured by a meaning-translation test. However, more vocabulary knowledge was acquired from the reading if we take the MC test as a measure of vocabulary knowledge. These two tests together suggest that the nature of vocabulary learning from extensive reading or listening is more complex than can be determined from this study. Indeed, it suggests that a considerable amount of vocabulary knowledge was gained from the exposure, but was not assessed. Such knowledge might include the noticing of lexical phrases, collocational and colligational patterns, new nuances of meanings, improved lexical access speed, and so on. It is probably here that the true benefit of reading and listening extensively occurs.

Investigating how much collocation, lexical pattern knowledge and so forth is learnt from extensive reading and listening is probably where the future lies with this type of research, because numerous studies including this one have now determined how much learners can pick up from word-focused experiments, as opposed to word knowledge at the supra-word level (i.e., collocation and lexical patterns). We feel it is now time for researchers to look beyond the word level and research the more complex nature of vocabulary learning as measured by collocational

knowledge, lexical pattern knowledge and so forth.

References

- Akinyemi, R. (1994). *The witches of pendle*. Oxford, England: Oxford University Press.
- Amer, A. (1997). The effect of the teacher's reading aloud on the reading comprehension of EFL students. *ELT Journal*, 51, 43–47.
- Bassett, J. (1991). *One-way ticket*. Oxford, England: Oxford University Press.
- Boettcher, J. (1980). *Fluent readers' strategies for assigning meaning to unfamiliar words in context*. Unpublished doctoral dissertation, University of Minnesota, USA.
- Bright, J., & McGregor, G. (1970). *Teaching English as a second language*. London: Longman.
- Brown, R. (2000). Extensive reading in action. *Studies in English Language and Literature*, 41, 79–123.
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Carey, S. (1982). Semantic development: The state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 345–389). Cambridge, England: Cambridge University Press.
- Cho, K., & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37, 662–667.
- Clark, E. (1973). What's in a word: On the child's acquisition of semantics in his first language. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York, NY: Academic Press.
- Dale, E., O'Rourke, J., & Bamman, H. (1971). *Techniques in teaching vocabulary*. Palo Alto: Field Educational Publications.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge, England: Cambridge University Press.
- Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7, 541–551.
- Deighton, L. (1959). *Vocabulary development in the classroom*. New York: Columbia University Press.
- Dhaif, H. (1990). Reading aloud for comprehension: A neglected teaching aid. *Reading in a Foreign Language*, 7, 457–464.
- Donkaewbua, S. (2008). *The effects of previous partial word knowledge on vocabulary learning through listening*. Manuscript in preparation. Victoria University of Wellington, New Zealand.
- Dupuy, B., & Krashen S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning*, 4, 55–63.
- Eichholz, G., & Barbe, R. (1961). An experiment in vocabulary development. *Educational Research Bulletin*, 28, 1–7.
- Eller, R., Pappas, C., & Brown, E. (1988). The lexical development of kindergartners: Learning from written context. *Journal of Reading Behavior*, 20, 5–24.
- Elley, W. (1985). *What do children learn from being read to?* Wellington, New Zealand: New Zealand Council of Educational Research & Institute of Education.
- Elley, W. (1988). *New vocabulary: How do children learn new words?* Wellington, New

- Zealand: New Zealand Council of Educational Research & Institute of Education.
- Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24, 174–187.
- Elley, W. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning*, 41, 375–411.
- Elley, W., & Mangubhai, F. (1981). The long-term effects of a book flood on children's language growth. *Directions*, 7, 15–24.
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57, 325–334.
- Gentner, D. (1975). Evidence for the psychological reality of semantic components: The verbs of possession. In D. Norman & D. Rumelhart (Eds.), *Explorations in cognition* (pp. 211–246). San Francisco, CA: WH Freeman and Co.
- Goh, C. (2002). *Teaching listening in the language classroom*. Singapore: SEAMEO Regional Language Centre.
- Grabe, W., & Stoller, F. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 98–122). Cambridge, England: Cambridge University Press.
- Hafiz, F., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System*, 18, 31–42.
- Harmer, J. (2003). *The practice of English language teaching*. Essex: Longman.
- Hayashi, K. (1999). Reading strategies and extensive reading in EFL classes. *RELC Journal*, 30, 114–132.
- Higa, M. (1965). The psycholinguistic concept of difficulty and the teaching of foreign language vocabulary. *Language Learning*, 15, 167–179.
- Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal*, 83, 367–384.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61, 355–382.
- 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.
- Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. *TESOL Quarterly*, 20, 763–768.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21, 767–787.
- Joe, A. (1994). *The effects of text-based tasks on incidental vocabulary learning*. Unpublished MA thesis, Victoria University of Wellington, New Zealand.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19, 357–377.
- Joe, A., Nation, P., & Newton, J. (1996). *Sensitive vocabulary tests*. Unpublished paper, Victoria University of Wellington, New Zealand.
- Krashen, S. (1993). *The power of reading*. Colorado: Eaglewood Libraries Unlimited.
- Krashen, S. (1994). The pleasure hypothesis. In J. Alatis (Ed.), *Georgetown University round table on languages and linguistics* (pp. 299–322). Washington, DC: Georgetown University Press.
- Krashen, S. (2003). *Explorations in language acquisition and use*. Portsmouth: 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, acquisition and pedagogy* (pp. 140–155). Cambridge, England: Cambridge University Press.
- Laufer, B., & Sim, D. (1985). An attempt to measure the threshold of competence for reading comprehension. *Foreign Language Annals*, 18, 405–411.
- Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25, 91–102.
- McArthur, T. (2003). English as an Asian language. *English Today*, 19, 19–22.
- Moody, H. (1974). Technique and art in reading aloud. *ELT Journal*, 28, 315–324.
- Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233–253.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Nation, P., & Wang, M. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12, 355–380.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1–28.
- Pinker, S. (1994). *The language instinct*. London: Penguin Books.
- 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.
- Prowse, P. (2005). Success with extensive listening. Retrieved May 15, 2005, from www.cambridge.org/elt/readers/prowse2.htm
- Richards, J., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics*. Malaysia: Pearson Education.
- Robb, T., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5, 239–251.
- Rodgers, T. (1969). On measuring vocabulary difficulty: An analysis of item variables in learning Russian-English vocabulary pairs. *IRAL*, 7, 327–343.
- Saragi, T., Nation, P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6, 72–78.
- Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88.
- Senechal, M., & Cornell, E. (1993) Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, 28, 361–374.
- Smith, R. (1997). Transforming a non-reading culture. In G. Jacobs, C. Davis, & W. Renandya (Eds.), *Successful strategies for extensive reading* (pp. 30–43). Singapore: SEAMEO Regional Language Centre.
- Vicary, T. (1989). *The elephant man*. Oxford, England: Oxford University Press.
- Waring, R. (1997). Graded and extensive reading: Questions and answers. *The Language Teacher*, 21, 7–13.
- Waring, R. (2008). *Issues when researching the incidental learning of collocations from reading*. Manuscript in preparation. Notre Dame Seishin University, Japan
- Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-Speaking World*, 4, 96–110.
- 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 Acquisition*, 27, 33–52.

West, M. (1953). The technique of reading aloud to a class. *ELT Journal*, 8, 21–24.

Widdowson, H. (1979). *Explorations in applied linguistics*. Oxford, England: Oxford University Press.

Appendix A

The List of Test Items for the 3 Stories

Group	<i>The Elephant Man</i>			<i>One-Way Ticket</i>			<i>The Witches of Pendle</i>		
	Word	Substitute	Recurrences	Word	Substitute	Recurrences	Word	Substitute	Recurrences
15–20 group	Happy	Mird	20	Guard	Loncher	20	Floor	Gaffle	15
	Day/s	Sall/s	17	Police	Dant	20	Family	Blunk	18
	Book/s	Hoult/s	19	Money	Shunk	17	Big	Rait	16
	All	Lert	17	Woman/men	Bandom/s	18	True/truth	Wathe	20
	Creature	Dront	15	Tall/er	Nagent/er	16	Judge	Heaft	20
	Shop/keeper	Plirty/keeper	15	Long	Boke	15	Afraid	Clomb	15
	Lady/ies	Smole/s	15	Diamond	Mong	15	Prison	Wessant	15
10–13 group	Cab	Tander	13	Slow/ly	Mald/ly	12	Friend	Fandle	13
	Door/s	Plitch	13	Story/ies	Preat/s	12	Eye	Florp	13
	Bed	Crost	12	Knife	Flotter	11	Pedlar	Pline	11
	Skin	Labin	12	Restaurant	Onder	11	Food	Chorm	10
	Theatre	Weat	11	Newspaper	Nivel	11	Warm	Thift	10
	Nurse	Koon	11	Station	Whiffle	11	Fire	Gorgan	10
	Body	Bletch	10	Holiday	Trank	10	Hair	Gurt	10
7–9 group	Home	Alart	9	Voice	Blamp	9	Dark	Poken	9
	Letter	Hine	9	Loud/ly	Dage/ly	9	Picture	Fent	8
	Little	Pusy	9	Quick/ly	Roth/ly	8	Horse	Brask	8
	Mouth	Reak	7	Great	Gline	9	Table	Chutter	7
	Leg	Grift	6	Seat	Shuft	8	Bad	Lood	7
	Bag	Slape	7	Drink	Mastime	6	Kind/ly	Spollen/ly	6
	Old	Throst	7	Husband	Mollet	7	Noise/y	Drint/y	6
2–3 group	Hole	Kisp	3	Expensive	Dasp	3	Difficult	Aspute	2
	Famous	Frime	2	Boat	Elver	3	Summer	Starp	3
	Blind	Creach	2	Corridor	Scront	3	Ugly	Lorky	3
	Cigarette	Queffle	2	Top	Pib	2	Hat	Jerth	3
	Trousers	Spullers	2	Village	Bawn	2	Bottle	Keem	2
	Heavy	Sweth	2	Blood	Chonter	2	Boy	Platt	2
	Nose	Culb	2	Hot	Teft	2	Chair	Slone	2
Total		272			272			264	
Running words		5415			5522			5765	
Coverage		95.0%			95.0%			95.4%	

Appendix B

Samples of the test

Test 1. Meaning-translation Test

What do these words mean? Write the meaning in Japanese.

(以下の単語の意味は何だと思えますか？日本語で答えて下さい。もし考えが複数ある場合は、自信のある順に書いて下さい。)

- | | | | |
|----|--------------|--------|--------|
| 1) | mird | 1..... | 2..... |
| 2) | sall | 1..... | 2..... |
| 3) | hault | 1..... | 2..... |

Test 2. Multiple Choice Recognition Test

Circle the word with the nearest meaning.

(1～28の単語の意味に近いと思う単語を4つの中からそれぞれ選んで下さい。分からない場合は分からないを選んで下さい。)

- | | | | | | | |
|----|--------------|------|--------|----|----|-------|
| 1) | mird | ハッピー | ワクワクする | 寒い | 臭い | 分からない |
| 2) | sall | 日 | 誕生 | 腕 | 願い | 分からない |
| 3) | hault | 人 | 本 | 家 | 鳥 | 分からない |

English translations of the correct choices and distracters in this sample test for *The Elephant Man*.

- | | | | | | | |
|----|--------------|--------------|-------------|-------|--------|--------------|
| 1) | mird | <u>happy</u> | exciting | cold | smelly | I don't know |
| 2) | sall | <u>day</u> | birth | arm | wish | I don't know |
| 3) | hault | Person | <u>book</u> | house | bird | I don't know |

About the Authors

Ronan Brown, MA MEd, has taught English as a foreign language in Saudi Arabia, UAE, China, and Japan. He is a professor of English at Seinan Gakuin University in Fukuoka, Japan. His research interests include extensive reading, vocabulary acquisition, and the teaching of literature in the language classroom. E-mail: ronan@seinan-gu.ac.jp

Dr. Rob Waring researches extensive reading and second language vocabulary acquisition. He has presented and published widely on these topics. He is an associate professor at Notre Dame Seishin University in Okayama, Japan. Professor Waring is a board member of the Extensive Reading Foundation. E-mail: waring_robert@yahoo.com

Dr. Sangrawee Donkaewbua is currently a lecturer at Rajabhat Mahasarakham University, Thailand. In 2008, she was awarded her PhD in Applied Linguistics from Victoria University of Wellington, New Zealand. E-mail: sangraweed@yahoo.com