The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge

Sarvenaz Hatami
California State University, Long Beach
United States

Abstract

This study compares the impact of second language (L2) reading and listening on the incidental acquisition and retention of five dimensions of vocabulary knowledge – spoken form, written form, part of speech, syntagmatic association, and form-meaning connection – at the level of recognition (form-meaning connection was measured also at the level of recall). The study also examines the relationship between frequency of word occurrence and vocabulary acquisition through reading versus listening. The participants were 139 pre-intermediate level English as a Foreign Language (EFL) learners with Farsi as their first language (L1), who were assigned to two experimental groups (i.e., reading and listening) and one control group. The experimental groups were exposed to the same text containing 16 target words (replaced by 16 non-words). The results on the immediate posttest revealed that readers scored higher than listeners on all five dimensions of word knowledge. Retention scores on a three-week delayed posttest (which due to the presence of testing effects could be measured for only one dimension of vocabulary knowledge, that is, form-meaning connection) were also higher for readers; however, listeners appeared to forget less within three weeks. Moreover, increase in the frequency of word occurrence in the text significantly benefitted incidental vocabulary acquisition through reading, but not through listening.

Keywords: incidental vocabulary acquisition, L2 reading, L2 listening, dimensions approach, depth of vocabulary knowledge, word recall, word recognition

Incidental vocabulary learning - the process of learning vocabulary without the intention of doing so or as a by-product of some other activity (Richards & Schmidt, 2002) - is known to play an important role in the development of L2 vocabulary (Huckin & Coady, 1999). In fact, beyond the first few thousand most common words in an L2, most vocabulary is acquired incidentally from context (Ellis, 1994; Huckin & Coady, 1999). Exposure to large amounts of written and spoken input enriches and consolidates knowledge of partially-known words and can also develop knowledge of new words (Brown, Waring, & Donkaewbua, 2008). Researchers therefore emphasize that incidental vocabulary acquisition is necessary for any well-balanced L2 vocabulary learning program (Nation, 2001; Schmitt, 2008).

A substantial amount of research has been conducted on the incidental acquisition of L2 vocabulary through reading (e.g., Elgort & Warren, 2014; Horst, Cobb, & Meara, 1998; Paribakht & Wesche, 1999; Pellicer-Sanchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2007). However, the number of research studies on L2 incidental vocabulary acquisition through listening is limited (Brown et al., 2008; van Zeeland & Schmitt, 2013a; Vidal, 2003). Even more scarce are studies that have investigated the comparative effects of reading and listening on L2 incidental vocabulary acquisition, which is an issue "of vital importance as it can help determine how much reading or listening (and what type) needs to be done in foreign language learning" (Brown et al., 2008, p. 139).

The limited number of studies that have investigated the differential effects of reading and listening on L2 incidental vocabulary acquisition (i.e., Brown et al., 2008; Vidal, 2011) have generally found that although both modes of language input result in vocabulary gains, reading is a more efficient source for incidental vocabulary acquisition than listening. However, in these studies, vocabulary gains have been measured in various ways. Brown et al. (2008) measured incidental vocabulary acquisition exclusively in terms of meaning. Although meaning is a crucial aspect of word knowledge, knowing a word involves much more. As the authors themselves acknowledge, through the tests used in their study they were not able to assess the full range of lexical knowledge gained from the language exposure; "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" (Brown et al., 2008, p. 158).

In a second study of the comparative effects of reading and listening on incidental vocabulary acquisition, in order to measure incidental vocabulary gains, Vidal (2011) used a depth of vocabulary knowledge measure, that is, a modified version of the Vocabulary Knowledge Scale (VKS) (see Wesche & Paribakht, 1996, for a description of the instrument). The VKS, a five-point scale starting at "I don't remember having seen this word before" and ending at "I can use this word in a sentence," exemplifies the *developmental* approach to conceptualizing and measuring depth of vocabulary knowledge. This approach, which represents the incremental nature of vocabulary learning, describes the acquisition of a lexical item along a continuum of mastery (Read, 1997). However, such developmental scales have long been the subject of criticism. For instance, as Schmitt (2010) explains, the beginning and ending points of such scales are inevitably only approximations, and the appropriate number of stages they should consist of is currently unknown (see pp. 217–221).

Another approach to measuring vocabulary depth is the *dimensions* or *components* approach (Read, 1997), which is known to be the most effective way to assess depth of vocabulary knowledge (Nation & Webb, 2011). In a seminal article that laid the foundation for the dimensions approach, Richards (1976) identified seven dimensions of vocabulary knowledge: word frequency, register, syntactic behaviour, form, semantic value, association, and conceptual meaning. Since then, several other scholars have attempted to describe what it means to know a word (e.g., Henriksen, 1999; Nation, 2001; Read, 2004). In the dimensions or components approach, the different dimensions of knowledge involved in knowing a lexical item are identified, and the degree to which each of these dimensions of word knowledge has been acquired is measured and quantified.

A number of incidental word learning studies have measured vocabulary gains using the dimensions approach. These studies have almost entirely focused on incidental vocabulary gains from reading (e.g., Chen & Truscott, 2010; Horst et al., 1998; Pellicer-Sanchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2007). van Zeeland and Schmitt's (2013a) study appears to be the only study that has used the dimensions approach to measure incidental vocabulary gains from listening. However, to date, no studies have been conducted to investigate the differential impact of reading and listening on various dimensions of word knowledge. Investigating this question through the dimensions approach is valuable because, although reading has been shown to be a more effective input mode for the incidental acquisition of word *meaning* (Brown et al., 2008), there may be other dimensions of word knowledge (e.g., written form, spoken form, part of speech, associations) that are more effectively acquired through listening. To this end, the current study compares the impact of the two input modes of reading and listening on the incidental acquisition and retention of different dimensions of vocabulary knowledge.

Literature Review

Incidental vocabulary acquisition from reading

As Horst et al. (1998) have reported in their review of the literature, early studies of incidental vocabulary acquisition from reading showed generally low vocabulary learning rates, that is, a rate of approximately one word correctly identified in every 12 words tested. However, these studies typically suffered from various methodological flaws (Horst et al., 1998), "including very small amounts of reading, insensitive measurement instruments, inadequate control of text difficulty, small numbers of target words, and no delayed posttests" (Schmitt, 2010, p. 29). Later studies that addressed some of these methodological issues have demonstrated higher vocabulary pick-up rates from reading (e.g., Elgort & Warren, 2014; Horst, 2005; Pellicer-Sanchez & Schmitt, 2010; Pigada & Schmitt, 2006). For instance, in a study by Horst (2005), 21 English as a Second Language (ESL) students with proficiency levels ranging from elementary to high-intermediate read an average of 10.5 graded readers of interest to each, over a six-week extensive reading program. At the end of the six weeks, individualized 100-item posttests (depending on the books each participant had chosen to read) were administered to the learners. The results indicated that participants gained new knowledge of more than half of the unknown words they encountered in the extensive reading materials that they had selected.

In a case study of vocabulary learning through extensive reading by Pigada and Schmitt (2006), a pre-intermediate level learner of French read four graded readers (a total of approximately 30,000 words) consisting of 133 target words over a period of one month. The findings revealed a pick-up rate of about one word in every 1.5 words tested. This impressive pick-up rate might have been due to the highly-capable participant in the study, as well as giving credit to partial knowledge of words through the informative one-on-one interview procedure. However, many of the gains reported were in orthography. It should also be noted that since this was a case study, the findings are not generalizable.

Pellicer-Sanchez and Schmitt (2010) investigated incidental word learning from an authentic

novel by 20 relatively advanced EFL learners. The English novel used in the study (consisting of approximately 67,000 words) included some African lexical items, 34 of which were chosen as target words. For all target words and word knowledge dimensions (i.e., meaning recognition and recall, spelling recognition, and word class recall), the results revealed measurable learning in 9.4 (close to one third) of the target words.

On the other hand, Waring and Takaki (2003) examined the rate at which vocabulary was learned by 15 intermediate level (and above) Japanese EFL learners from reading a graded reader (consisting of 5,872 words). On average, the meaning of only one of the 25 target words was remembered after three months, which meant that the learners acquired only one new word from one hour of reading. These results led the authors to conclude that while graded reading helps to deepen knowledge of already known words, it does not lead to the acquisition of many new words.

As evident from the brief review above, the reported vocabulary pick-up rates from reading vary considerably. In addition to learner-related factors, this variation could be attributed to differences in the nature and length of texts, target words, diversity and sensitivity of the measures, and the overall designs used in these studies.

Incidental vocabulary acquisition from listening

A number of L2 studies have investigated incidental vocabulary learning through listening (Brown et al., 2008; Chang, 2012; R. Ellis, 1995; van Zeeland & Schmitt, 2013a; Vidal, 2003, 2011). For instance, in a study by Vidal (2003), 116 Spanish EFL learners viewed three video-taped academic lectures on the topic of tourism (each lecture consisted of approximately 1,800 running words). Thirty-six target words (12 in each lecture) were chosen for the study. For the pre-test, immediate posttest, and delayed posttest (administered four weeks later), a modified version of the VKS was used to measure vocabulary knowledge. The results revealed a significant difference in vocabulary gains between the pre-test and the immediate posttest, indicating that listening (while watching) to academic lectures in EFL results in vocabulary growth.

van Zeeland and Schmitt (2013a) used the dimensions approach to investigate L2 vocabulary learning and retention through listening by 30 high-intermediate to advanced ESL learners. Learners listened to four passages with a total of 24 target words (the passages ranged from 883 to 1,322 words in length). The participants' gains in the three dimensions of vocabulary knowledge measured in the study were as follows: form (45.8% of the target words) > grammar (33.7%) > meaning (8.5%) immediately after listening, and form (25%) = grammar (24.6%) > meaning (7.5%) two weeks later. While knowledge of meaning was more difficult than form and grammar to develop, once developed it appeared to be retained for longer. Overall, learners gained knowledge of 29% of the target words immediately after listening and retained knowledge of 19% of the words two weeks later.

While it appears that L2 listening can lead to vocabulary gains, this is a neglected area in Second Language Acquisition (SLA) research and the studies conducted are too few in number to allow any general conclusions. Moreover, these studies have, for the most part, attempted to measure

only one or two dimensions of word knowledge. As van Zeeland and Schmitt (2013a) state:

> This lack of sensitive vocabulary knowledge assessment in listening studies is surprising. As learning gains from listening have [been] found to be small, even significantly smaller than those from reading, the dimensions approach should serve particularly well in revealing the smallest increments in learning. (p. 611)

Incidental vocabulary acquisition from reading versus listening

Studies that have compared the effects of reading and listening on L2 incidental word learning are very few in number (i.e., Brown et al., 2008; Vidal, 2011). Brown et al. (2008), in a study of 35 Japanese EFL learners with pre-intermediate or intermediate level proficiency, compared the impact of three input modes – reading, reading-while-listening, and listening – on incidental vocabulary acquisition and retention. For each mode of input, participants were exposed to a graded reader consisting of approximately 5,500 words. One posttest and two delayed posttests (one week later and three months later) measuring meaning recall and recognition of the target words were administered. The results on the immediate meaning recognition test revealed considerable gains of 48% and 45% (of the 28 target words) for the reading-while-listening group and reading-only group, respectively, and 29% for the listening-only group. The results on the immediate meaning recall test were as follows: 16% and 15% for the reading-while-listening group and the reading-only group, respectively, and only 2% for the listening-only group. However, after three months, on average, when learners were tested for meaning recall, the meaning of only one of the 28 target words was retained by the reading-only and the reading-while-listening groups, and the meaning of none of the target words by the listening-only group.

Vidal (2011) also compared the effects of reading and listening on the incidental vocabulary acquisition and retention of 230 Spanish university EFL learners at four different levels of language proficiency. The participants were assigned to one of three groups: they either (a) read three academic passages, (b) watched three academic lectures, or (c) received no input (i.e., control group). The texts ranged from 1,516 to 1,837 words in length. Thirty-six target words were chosen (12 in each text). All three groups received pretests, posttests, and delayed posttests of their knowledge of the target words. For this purpose, a modified version of the VKS was used with a maximum score of 5 for each target word. When considering the maximum possible score on the test (i.e., 180), average acquisition rates for the lowest and highest proficiency readers were 19.38% and 37.69% respectively. On the other hand, average acquisition rates for the lowest and highest proficiency listeners were 7.08% and 28.35% respectively. The difference in gains decreased as the proficiency level of the learners increased. Similar trends were also observed for the retention of the target words. As the author notes, "very low-proficiency listeners, as represented by the 25th percentile, had serious difficulties with the processing of speech and had to struggle for meaning in real time" (p. 244). Hence, when compared to the readers of the same proficiency level, their initial gains were very small and their loss of gains was larger.

Frequency of occurrence of target words

Studies that have focused on incidental word learning through reading have generally found that

the more frequently a word occurs in a text, the more likely it will be learned. For instance, Horst et al. (1998) found that the target words in their study needed to occur at least eight times for sizable learning gains to take place. In Waring and Takaki's (2003) study, the results appeared to suggest that in order for a learner to have a 50% chance of recognizing the word form or its meaning three months later, a word needed to be encountered at least eight times. However, there was only a 10% to 15% chance of recalling a word's meaning after three months, even if the word was met more than 18 times. The results of Pigada and Schmitt's (2006) study revealed that although there is no specific point at which the acquisition of meaning is guaranteed, by about 10 or more exposures there appears to be substantial learning gains. Yet, only when words were encountered 20 or more times was there a good possibility for all three dimensions of word knowledge (i.e., spelling, meaning, grammatical features) to be acquired. In a study with 121 Japanese EFL learners, Webb (2007) investigated the impact of one, three, seven, and 10 encounters on five different dimensions of word knowledge (i.e., orthography, meaning and form, paradigmatic association, syntagmatic association, and grammatical functions). The results revealed that each time the repetitions increased, at least one dimension of word knowledge was enhanced. Webb also found that unknown words needed to be met a minimum of 10 times in context for considerable vocabulary growth to occur; but to gain full mastery of a word, more than 10 repetitions might be necessary. Pellicer-Sanchez and Schmitt (2010) also found that with even a single exposure, there was sizeable learning in the recognition of word form and meaning, but very little gain in the recall of word class or meaning. According to the authors, noticeable increase in gains began with 5-8 exposures and accelerated with 10-17 exposures.

In studies of L2 incidental vocabulary acquisition from listening, frequency of occurrence has also been shown to positively affect L2 vocabulary gains; however, the effect does not appear to be strong. In Vidal's (2003) study, the target words occurred 1, 2, 3, 4, 5, and 6 times in each lecture. Vocabulary gains from listening generally increased as the word was repeated more times; but overall, the effect of frequency of occurrence was not strong. van Zeeland and Schmitt (2013a) also investigated the impact of frequency of occurrence (i.e., 3, 7, 11, or 15 times) on incidental vocabulary gains from listening. Frequency of occurrence did not appear to have an effect on the acquisition or retention (two weeks later) of any of the three vocabulary dimensions, except for the increase from 3 to 7 occurrences (and this effect was found for only form and grammar, not meaning, and on only the immediate posttest, not the delayed posttest).

In studies that have compared the impact of reading and listening on L2 incidental word learning, the effect of repetition on vocabulary gains in reading has been shown to be stronger than in listening. In Brown et al. (2008), four frequency bands were selected: 2-3, 7-9, 10-13, and 15-20 times. Brown and colleagues found that items with higher frequency of occurrence in the text were more likely to be learned and retained; however, gains were considerably smaller through listening than through reading. Based on the results, the authors concluded that it is very unlikely for a new word to be acquired through listening unless it occurs considerably more than 20 times.

Vidal (2011) also attempted to investigate how the relationship between frequency of occurrence (1, 2, 3, 4, 5, and 6 times) and vocabulary learning compares between reading and listening. In both modes, vocabulary learning increased as word repetition increased; however, the effect of repetition was considerably stronger in reading. Moreover, in the reading condition, the greatest increase in learning occurred between two and three repetitions; while in the listening condition,

the greatest increase occurred between five and six repetitions.

As the above-mentioned studies clearly show and as Nation and Wang (1999) assert, "there is no set number of repetitions that will ensure learning" since "so many factors influence vocabulary learning from written [and spoken] text[s]" (p. 363). These factors can be related to the word, the text, the task, or the learner (Paribakht & Wesche, 1999).

From the review above, it appears that reading is a more effective source of input than listening for L2 incidental vocabulary acquisition. However, to date, no studies have measured various dimensions of word knowledge when comparing the effects of reading and listening on incidental word learning. If other dimensions of word knowledge are measured, different results regarding the effectiveness of these two input modes might emerge. The current study, therefore, seeks to answer the following research questions:

1. What is the differential impact of reading and listening on the L2 incidental acquisition of five dimensions of word knowledge (i.e., spoken form, written form, part of speech, syntagmatic association, and form-meaning connection) as measured by an immediate posttest?
2. What is the differential impact of reading and listening on the L2 retention of these five dimensions of word knowledge as measured by a delayed posttest?
3. How does the relationship between frequency of word occurrence in the text and L2 incidental acquisition of these five dimensions of word knowledge combined compare across reading and listening?

Method

Participants

The participants for this study were 139 undergraduate students from various majors studying at a high ranking university in Iran. They were 56 females and 83 males with an average age of 19.63 years (*SD* = 1.39; range = 18-25 years). The participants all shared the same L1, Farsi. None of the participants had ever resided in an English-speaking country. These participants had formally studied EFL for approximately seven years at middle and high school and were at pre-intermediate levels of English language proficiency. On the Vocabulary Levels Test (VLT; Schmitt, Schmitt, & Clapham, 2001), their mean scores (out of 30) for the 2,000, 3,000, and 5,000 word levels were 22.91, 14.74, and 7.44, respectively. The VLT was administered to ensure that participants were at a proficiency level that allowed them to read or listen to the selected texts with little or no difficulty. Mastery of at least 50% of the 2,000 word level was chosen as the minimum cut-off point for inclusion of participants. All participants received cash incentives (equivalent to $10 CAD) for their participation in the study.

The participants were then randomly assigned to two experimental groups, reading (*n* = 39) and listening (*n* = 51), and one control group (*n* = 49). The reading group's mean score (out of 90) on the combination of the 2,000, 3,000, and 5,000 word levels of the VLT was 43.21 (*SD* = 15.73), the listening group's mean was 45.14 (*SD* = 10.82), and the control group's mean was 46.73 (*SD*

= 14.29). There was no significant difference between the three groups, $F(2, 138) = .73$, $p = .48$.

This study began with 211 participants; however, 72 participants were excluded because of low scores on the VLT, absence, or their non-random assignment (this was intentionally done in order to answer a research question to be addressed in another paper). For these reasons, the groups in the present study, despite their random assignment, were not quite equal in size.

Materials

Target words. Sixteen words in the text were chosen as target words. The target words were selected primarily on the basis of their part of speech and frequency of occurrence in the text. These target words, which were most likely already known to the learners, were then replaced by 16 non-words that follow English phonological and orthographic patterns. Many previous studies have used non-words to replace already known concepts (e.g., Brown et al., 2008; Waring & Takaki, 2003; Webb, 2007). This is considered the simplest level of learning a new word because the non-word is representing a familiar concept and only a new label for this familiar concept needs to be acquired (Nation, 2001). However, as Nation and Webb (2011) assert, "much of second or foreign language learning, certainly in the initial stages, does not immediately involve the development of a lot of new concepts. Typically, L2 word forms are connected to already existing L1 meanings" (p. 267).

To select these non-words, initially 46 non-words from Meara's (2013) list of imaginary words were selected. From these 46, 16 were excluded based on the judgments of three professors of Teaching English as a Second Language (TESL), for the following reasons: the non-words had irregular and confusing pronunciation or spelling, had real English words embedded in them, were very common English first names or surnames, or looked French. The remaining 30 non-words were then embedded in a questionnaire and administered to five native speakers of English (1 male, 4 females, mean age = 38 years) and five non-native speakers of English with Farsi as their L1 (4 males, 1 female, mean age = 29.8 years). The questionnaire required participants to judge the plausibility of each non-word as an English word (yes-no question), as well as the pronunciation and spelling difficulty of each non-word (using a 5-point scale; 1 = very easy, 5 = very difficult). Based on the responses, 16 non-words were chosen for this study, all of which shared the following characteristics: they were two syllables and five to six letters in length; they were judged as plausible English words by at least eight of the 10 judges; and the average spelling difficulty and pronunciation difficulty ratings for each were lower than 3. When considering the responses provided by the non-native speakers only, the 16 chosen non-words were judged as plausible English words by at least four of the five judges; and the average spelling difficulty and pronunciation difficulty ratings for each were lower than 2.7. These steps were taken to ensure that the target words were largely equivalent in terms of learning difficulty.

Reading and listening material. The Monkey's Paw, an elementary level graded reader selected from the Oxford Bookworms series, was used for both the reading and the listening groups. Using the BNC-COCA-25 VocabProfile (available at www.lextutor.ca/vp/), any words beyond the first 1,000 word level were simplified. In addition, all proper nouns judged to be unfamiliar to the participants by the researcher (whose L1 is also Farsi) were changed to more familiar ones; for example, *Herbert* was changed to *Jack*. Four frequency bands (FB) were chosen (a frequency

band refers to a range of frequencies of target word occurrence in the text): 2-5 (i.e., FB1), 7-10 (i.e., FB2), 12-15 (i.e., FB3), and 17-20 (i.e., FB4). There were four target words in each frequency band (see Appendix A for details): two nouns, one adjective, and one verb (only used in the past tense throughout the story). The text contained 4,231 words, and a lexical coverage of 95.84% was achieved. Previous studies have shown that a lexical coverage of 95% to 98% provides adequate comprehension of written and spoken texts (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Stæhr, 2009; van Zeeland & Schmitt, 2013b).

For the listening experiment, the text was read aloud by a native speaker of Canadian English (a TESL professor), recorded on a CD, and later played for the listening group. The narration was produced at an average speech rate of 117.5 words per minute and had a duration time of 36 minutes. Following Vidal (2011), the reading group was given the same amount of time as the listening group to read the text (i.e., 36 minutes). This approach is based on Hirai's (1999) study in which he compared the listening rates and reading rates of Japanese EFL learners with varying proficiency levels. Hirai found that the optimal listening rates and reading rates are similar for L2 learners.

Instruments

Language background questionnaire (LBQ). An LBQ (translated into Farsi) was administered to participants to collect demographic information (e.g., sex, age, native country, native language, other languages spoken and proficiency levels). Participants also reported whether or not they had lived in an English-speaking country, and how long they had studied English outside of school and university.

Vocabulary posttest. One of the most comprehensive frameworks explaining different dimensions of word knowledge has been proposed by Nation (2001). In order to assess L2 incidental vocabulary acquisition and retention in this study, five dimensions of word knowledge were chosen from Nation's framework and were measured using five recognition tests and one recall test (see Appendix B for examples of these tests). These six tests assessed recognition of spoken form (Test1-SF), recognition of written form (Test2-WF), recall of meaning (Test3-Mg-Recall), recognition of part of speech (Test4-PS), recognition of syntagmatic association[1] (Test5-SA), and recognition of meaning (Test6-Mg-Recog). Hence, all five dimensions were measured at the level of recognition, while form-meaning connection was measured also at the level of recall. The tests were based on the work of Webb (2005), Chen and Truscott (2010), and van Zeeland and Schmitt (2013a), with some additions and adjustments.

The vocabulary posttest was the same for both the experimental groups and the control group in the form of a 12-page booklet; each of the six tests appeared on two consecutive pages, with eight (of the 16) target words on one page and another eight on the next. The tests were sequenced so that any possible learning effect was avoided (Webb, 2005). Learners were instructed to avoid going back to revise answers, and supervision was provided to ensure this. All test instructions appeared in both English and Farsi. In scoring, each correct response in the

---

[1] Syntagmatic association exists when two words have a sequential relationship to one another (e.g., verb-noun pairs like eat-food), while paradigmatic association exists when two words belong to the same word class (e.g., verb-verb pairs like eat-drink) (Schmitt, 2010).

multiple-choice tests was awarded one point. For the meaning recall test, correct answers were given one point, and answers with a similar meaning were given half a point. For example, when the correct response was *tea*, if the participant responded *tea*, one point was given, and if they responded *a drink*, half a point was given. Two raters scored the meaning recall test and an inter-rater reliability of 98% was achieved.

Procedures

Before the study was conducted, the materials and instruments were piloted with four Iranian EFL learners with characteristics similar to those of the target population. As a result, changes were made to some of the instructions and Farsi translations. For the actual experiment, data were collected over the course of three sessions as outlined below:

Session 1. Following an explanation of the study, participants were invited to sign a consent form. If in agreement, they were then asked to complete the LBQ and the VLT. This session took approximately 50 minutes.

Session 2. In this session, which was held about two weeks after Session 1, participants were not informed of the vocabulary acquisition focus of the study, and thus the vocabulary posttest remained unannounced. They were told that the main purpose of this session was to either read or listen to a classic English novel and to try to understand it. The participants were then presented with the written or spoken texts. Immediately after all the readings or listenings were completed, the unannounced vocabulary posttest was administered. Participants were first provided with two practice examples for each of the six tests. They were then given as much time as they needed to complete the main tests. This second session took approximately 75 minutes.

Session 3. Three weeks after Session 2, the delayed posttest was administered. According to Schmitt (2010), "a delayed posttest of three weeks should be indicative of learning which is stable and durable" (p. 157). The delayed posttest was administered to measure retention of different dimensions of word knowledge, and it consisted of the same series of vocabulary tests as the immediate posttest. After the participants completed the test, they were debriefed regarding the non-words not being real English words. This final session took approximately 30 minutes.

The control group completed all the above-mentioned procedures (i.e., consent form, LBQ, VLT, immediate posttest, and delayed posttest) at the same intervals; however, they were not exposed to the spoken or written texts.

Results

Before conducting the analyses for each research question, a one-way repeated measures MANOVA was run to see if the scores of the control group had significantly improved from the immediate posttest (Time_1) to the delayed posttest (Time_2). As the control group did not receive any treatment, a significant increase in their scores from Time_1 to Time_2 indicated

that testing effects were present. In Table 1, the estimated means for the control group on the six tests at Time_1 and Time_2 are indicated. The results of the repeated measures MANOVA yielded a significant multivariate effect for Time, Wilks' $\lambda = .41$, $F(6, 41) = 10.01$, $p < .001$, partial $\eta^2 = .59$, power = 1.0. The results of the univariate ANOVAs indicated a significant increase in scores for the control group from Time_1 to Time_2 on Test1-SF, $F(1, 46) = 9.45$, $p = .004$, partial $\eta^2 = .17$, power = .85; Test2-WF, $F(1, 46) = 50.13$, $p < .001$, partial $\eta^2 = .52$, power = 1.0; Test4-PS, $F(1, 46) = 4.62$, $p = .037$, partial $\eta^2 = .09$, power = .56; and Test5-SA, $F(1, 46) = 8.21$, $p = .006$, partial $\eta^2 = .15$, power = .80. However, univariate results were non-significant for Test3-Mg-Recall, $F(1, 46) = 2.89$, $p = .096$, partial $\eta^2 = .06$, power = .38; and Test6-Mg-Recog, $F(1, 46) = .20$, $p = .660$, partial $\eta^2 = .004$, power = .07. Because testing effects appeared to be present for Test1-SF, Test2-WF, Test4-PS, and Test5-SA, these four tests were eliminated from the data analysis conducted for the second research question in this study.

Table 1. *Estimated means for the control group at Time_1 and Time_2*

Test	Time	M	SD Error	95% Confidence Interval Lower Bound	Upper Bound
Test1-SF	1	4.36	.33	3.69	5.03
	2	5.43	.38	4.67	6.18
Test2-WF	1	4.55	.42	3.71	5.39
	2	7.75	.39	6.96	8.53
Test3-Mg-Recall	1	.00	.00	.00	.00
	2	.11	.06	-.02	.23
Test4-PS	1	3.32	.38	2.55	4.09
	2	4.23	.40	3.43	5.04
Test5-SA	1	1.62	.27	1.08	2.15
	2	2.32	.31	1.69	2.95
Test6-Mg-Recog	1	2.13	.26	1.61	2.65
	2	2.23	.30	1.64	2.83

Note. Time_1 = immediate posttest; Time_2 = delayed posttest. The maximum possible score is 16.

Research question 1

In Table 2, the data from the immediate posttest (Time_1) for tests 1 to 6 for the three groups (i.e., reading, listening, and control) are summarized. A one-way between-subjects MANOVA was run with Group (reading vs. listening vs. control) as the independent variable and the scores on tests 1 to 6 at Time_1 as the six dependent variables. The results revealed a significant multivariate effect for Group, Wilks' $\lambda = .33$, $F(12, 262) = 15.99$, $p < .001$, partial $\eta^2 = .42$,

power = 1.0. Univariate tests also indicated a significant effect for Group on each of the six tests, as follows: Test1-SF, $F(2, 136) = 40.11$, $p < .001$; Test2-WF, $F(2, 136) = 75.27$, $p < .001$; Test3-Mg-Recall, $F(2, 136) = 27.21$, $p < .001$; Test4-PS, $F(2, 136) = 33.34$, $p < .001$; Test5-SA, $F(2, 136) = 50.13$, $p < .001$; and Test6-Mg-Recog, $F(2, 136) = 40.85$, $p < .001$. Follow-up post-hoc tests were performed; Tukey HSD was used in cases where Levene's values were greater than .05 and Games-Howell in all other cases (Gamst, Meyers, & Guarino, 2008). The results revealed that the control group was significantly different ($p < .001$) from the treatment groups (i.e., reading and listening) on all six tests at Time_1. The difference between the reading group and the listening group was non-significant only on Test1-SF ($p = .67$) and significant on all the other tests: Test2-WF ($p < .001$), Test3-Mg-Recall ($p = .005$), Test4-PS ($p < .001$), Test5-SA ($p < .001$), and Test6-Mg-Recog ($p = .007$). The largest mean difference between the reading group and listening group was found on Test2-WF (MD = 4.80), followed by Test5-SA (MD = 3.74), Test4-PS (MD = 2.84), Test6-Mg-Recog (MD = 2.70), Test3-Mg-Recall (MD = 1.26), and finally, Test1-SF (MD = .45) (see Figure 1).

Table 2. *Descriptive statistics for tests 1 to 6 at Time_1*

Test	Group	*n*	*M*	*SD*
Test1-SF	Control	49	4.33	2.30
	Listening	51	8.04	2.76
	Reading	39	8.49	2.26
Test2-WF	Control	49	4.51	2.90
	Listening	51	7.20	2.91
	Reading	39	12.00	2.74
Test3-Mg-Recall	Control	49	.00	.00
	Listening	51	0.95	1.41
	Reading	39	2.21	2.08
Test4-PS	Control	49	3.35	2.63
	Listening	51	5.96	2.99
	Reading	39	8.80	3.76
Test5-SA	Control	49	1.69	1.83
	Listening	51	4.06	2.94
	Reading	39	7.80	3.66
Test6-Mg-Recog	Control	49	2.14	1.80
	Listening	51	5.92	3.80
	Reading	39	8.62	4.23

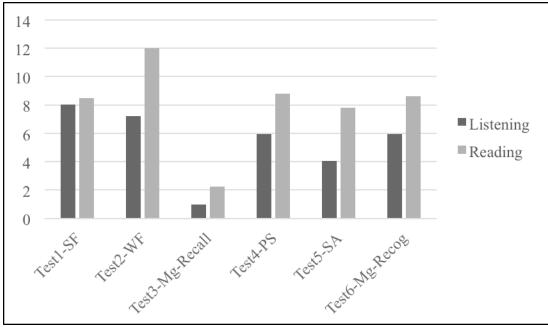*Note.* The maximum possible score is 16.

Figure 1. Differences between the reading and listening groups on tests 1 to 6 at Time_1.

Research question 2

As previously mentioned, because of the presence of testing effects for Test1-SF, Test2-WF, Test4-PS, and Test5-SA, these four tests were excluded from the analysis for this question. Therefore, retention was only examined for Test3-Mg-Recall and Test6-Mg-Recog. In other words, while for the first research question on *acquisition,* the differential impact of reading and listening was investigated on five dimensions of word knowledge, for the second research question on *retention* it was only possible to examine this impact on one dimension of word knowledge (i.e., meaning), but at two different levels (i.e., recognition and recall). In Table 3, the data at Time_1 and Time_2 for Test3-Mg-Recall and Test6-Mg-Recog are summarized. These data are also presented graphically in Figures 2 and 3.

Table 3. *Descriptive statistics for Test3-Mg-Recall and Test6-Mg-Recog at Time_1 and Time_2*

Test	Time	Group	*n*	*M*	*SD*
		Control	47	.00	.00
	1	Listening	47	.84	1.32
Test3-Mg-Recall		Reading	37	2.14	2.03
		Control	47	.11	.43
	2	Listening	47	.84	1.21
		Reading	37	1.53	1.47

		Control	47	2.13	1.77
	1	Listening	47	5.66	3.83
Test6-Mg-Recog		Reading	37	8.46	4.29
		Control	47	2.23	2.04
	2	Listening	47	5.64	3.35
		Reading	37	7.27	4.36

Note. Time_1 = immediate posttest; Time_2 = delayed posttest.
The maximum possible score is 16. Eight missing cases (the scores
of two participants were outliers in this analysis, and therefore
excluded, and six participants did not complete the delayed posttest).

A mixed-design MANOVA was performed with Group (reading vs. listening vs. control) as the between-subjects factor, Time (Time_1 vs. Time_2) as the within-subjects factor, and the scores on Test3-Mg-Recall and Test6-Mg-Recog as the two dependent variables. The results yielded significant multivariate effects for Group, Wilks' λ = .63, $F(4, 254)$ = 16.77, $p < .001$, partial η^2 = .21, power = 1.0; Time, Wilks' λ = .93, $F(2, 127)$ = 4.56, $p < .05$, partial η^2 = .07, power = .77; and the interaction between Group and Time, Wilks' λ = .83, $F(4, 254)$ = 6.13, $p < .001$, partial η^2 = .09, power = .99. Univariate testing also found significant effects for Group ($p < .001$), Time ($p < .05$), and the interaction between Group and Time ($p < .05$) on each of the two tests. Simple effects analysis showed that for Test3-Mg-Recall and Test6-Mg-Recog, at both Time_1 and Time_2, the control group was significantly different from the treatment groups ($p < .05$). For Test3-Mg-Recall, the difference between the reading group and the listening group was statistically significant at both Time_1 ($p < .001$) and Time_2 ($p < .05$). For Test6-Mg-Recog, the difference between the reading group and listening group was significant at Time_1 (p = .001), but non-significant at Time_2 (p = .08).
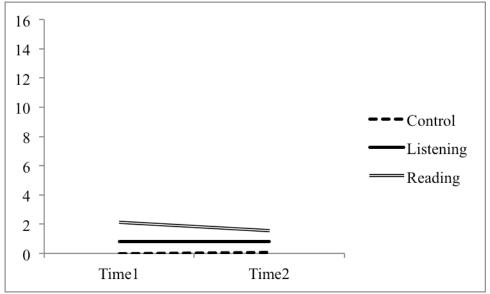

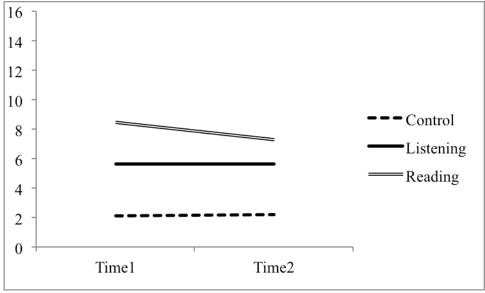**Figure 2.** Mean scores on Test3-Mg-Recall at Time_1 and Time_2.

Figure 3. Mean scores on Test6-Mg-Recog at Time_1 and Time_2.

Research question 3

In order to respond to this question, for each frequency band, the mean scores on the five recognition tests, that is Test1-SF, Test2-WF, Test4-PS, Test5-SA, and Test6-Mg-Recog at Time_1 were calculated, subsequently added together, and averaged. Therefore, one set of mean scores was obtained for each frequency band. Test3-Mg-Recall (i.e., the only recall test in this study) was excluded. In Table 4, the data for the four frequency bands at Time_1 are summarized.

Table 4. *Descriptive statistics for the four frequency bands at Time_1*

Frequency band	Group	*M*	*SD*
FB1	Control	.24	.15
(2-5)	Listening	.36	.17
	Reading	.50	.17
FB2	Control	.19	.12
(7-10)	Listening	.42	.19
	Reading	.58	.24
FB3	Control	.18	.12
(12-15)	Listening	.40	.19
	Reading	.69	.17
FB4	Control	.18	.08
(17-20)	Listening	.39	.22
	Reading	.60	.23

A mixed-design ANOVA was conducted with Group (reading vs. listening vs. control) as the between-subjects factor, and Frequency Band (FB1 vs. FB2 vs. FB3 vs. FB4) as the within-subjects factor, and the combined (and later averaged) mean scores on Test1-SF, Test2-WF, Test4-PS, Test5-SA, and Test6-Mg-Recog as the dependent variable. The results of the ANOVA (Huynh-Feldt correction applied) yielded a significant main effect for Group, $F(2, 129) = 80.51$, $p < .001$, partial $\eta^2 = .56$, power = 1.0; Frequency Band, $F(2.93, 378.17) = 6.33$, $p < .001$, partial $\eta^2 = .05$, power = .96; and a significant interaction effect between Group and Frequency Band, $F(5.86, 378.17) = 11.64$, $p < .001$, partial $\eta^2 = .15$, power = 1.0. Simple effects analysis indicated that for all four frequency bands, the control group was significantly different from the treatment groups ($p \leq .001$). For all four frequency bands, the reading group was also found to be significantly different from the listening group ($p < .001$). Moreover, in the listening group no significant differences were found between any of the four frequency bands ($p > .05$); while in the reading group, significant differences ($p < .05$) were found between all frequency bands, except between FB2 and FB4 ($p > .05$) (see Figure 4).
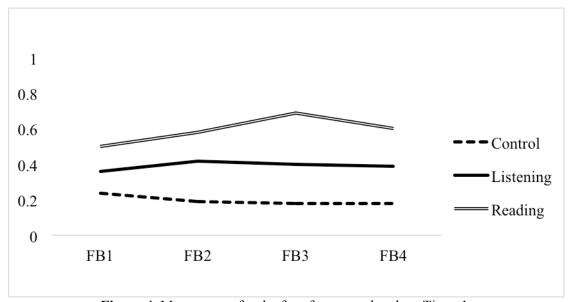

Figure 4. Mean scores for the four frequency bands at Time_1.

Discussion

Vocabulary acquisition

This study has shown that both L2 reading and listening are sources of incidental vocabulary learning. However, reading was found to be a more effective input mode than listening for the incidental acquisition of the different dimensions of word knowledge measured in this study. In other words, except on the test of recognition of spoken form where the difference between the two groups was non-significant, the reading group scored significantly higher than the listening group on all the other tests measuring vocabulary acquisition. The largest difference between readers and listeners was in the incidental acquisition of written form, followed by syntagmatic

association, part of speech, meaning recognition, and meaning recall. It therefore appears that, not surprisingly, the written form is significantly easier to acquire from written input than spoken input while, interestingly, the spoken form is acquired with similar ease from written and spoken input. This could be because written and spoken input both gain access to the phonological storage in the working memory and can be stored as phonological code, with this difference that spoken input gains direct access to the phonological store, while non-phonological input (e.g., written words) needs to be recoded into its phonological form through a process of subvocal rehearsal to become registered in the phonological store (Baddeley, 2007).

The finding that reading was in general superior to listening for L2 incidental word learning is in line with findings from previous research (i.e., Brown et al., 2008; Vidal, 2011). This finding could be largely due to the fact that in reading, "there is opportunity for the reader to study the context, to form hypotheses at leisure and cross-validate them ... The word is frozen in time on the page, whereas in speech it passes ephemerally" (N. C. Ellis, 1995, pp. 105-106). Moreover, the participants in this study, because of their EFL context and educational background, had received more exposure to and practice with written texts than spoken texts in English. They might, therefore, have failed to some extent to keep up with the flow of speech and to recognize the boundaries between spoken forms in connected speech (Brown et al., 2008).

For the reading group, the gains and order of acquisition of the different dimensions and levels (recognition vs. recall) of word knowledge were as follows: written form (75% of the target words) > part of speech (55%) > meaning recognition (54%) > spoken form (53%) > syntagmatic association (49%) > meaning recall (14%). It should be noted that the large and impressive recognition percentages are not equivalent to complete knowledge of a new word, "but rather reflect that [the recognition of] one or more aspects of knowledge had started to develop" (van Zeeland & Schmitt, 2013a, p. 615). The gains for the reading group in this study support previous findings. For example, the readers' gains in meaning recognition and meaning recall were 45% and 15% in Brown et al. (2008), 42% and 18% in Waring and Takaki (2003), and 43% and 14% in Pellicer-Sanchez and Schmitt (2010), respectively (compared with 54% and 14% in this study). In Waring and Takaki, the written form recognition gains were 61% (compared with 75% here). One reason for the lower recognition rates in these studies (when compared to the present study) could be the presence of an *I don't know* option in the meaning recognition tests. An *I don't know* option often allows learners to simply opt out, without encouraging them to draw on their sub-conscious knowledge to make an informed guess (Nation, 2012); in the present study, an *I don't know* option was not provided in order to make the tests more sensitive.

For the listening group, the gains and relative order of acquisition were as follows: spoken form (50% of the target words) > written form (45%) > part of speech (37.3%) > meaning recognition (37%) > syntagmatic association (25%) > meaning recall (6%). These gains also match findings from previous research to a great extent. In Brown et al.'s (2008) study, the listening group gains for meaning recognition and for meaning recall were 29% and 2%, respectively (compared with 37% and 6% in this study). In van Zeeland and Schmitt (2013a), spoken form was recognized for 45.8% of the target words (compared with 50% here), grammar was recognized for 33.7% (compared with 37.3% here), and meaning was recalled for 8.5% (compared with 6% here). Again, one reason for the slightly lower recognition rates in these studies when compared with the current study could be the presence of an *I don't know* option in their recognition tests. In

general, differences in gains (whether through reading or listening) among these studies could be due to a number of factors, including but certainly not limited to differences in text length and lexical coverage rates, number of target words and their learning difficulty, tests used for measuring vocabulary gains, speech rates, as well as learner-related factors, such as L1 and proficiency level. Hence, any comparison between different studies (as is done above in terms of percentages of vocabulary gains) should be considered with caution.

As can be seen in both groups, the largest gains were in word form. On the other hand, the smallest gains were in meaning recall; after 36 minutes of exposure to a simplified text containing as many as 20 repetitions of the target words, meaning was recalled for only two words (out of 16) by the readers and for only one word by the listeners. Therefore, meaning recall appears to be the most difficult to acquire incidentally. This order of acquisition, that is, form (here, written form for the reading group and spoken form for the listening group) at the high end, and meaning (here, syntagmatic association and form-meaning connection) at the low end has been documented in previous research on reading (Chen & Truscott, 2010; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007) and listening (van Zeeland & Schmitt, 2013a). Thus, whether the input mode is written or spoken, the incidental acquisition of form tends to precede the incidental acquisition of meaning.

Vocabulary retention

As previously mentioned, it was possible to measure retention for only form-meaning connection. After three weeks, the reading group scored significantly higher than the listening group on the test of meaning recall; however, on the test of meaning recognition, the difference between the two groups was no longer significant. This is because readers and listeners were found to differ in terms of retention rates. The results showed that after three weeks the listeners had retained almost all the knowledge of meaning (both recognition and recall) that they had initially acquired. This finding is congruent with van Zeeland and Schmitt (2013a), who also found that, contrary to knowledge of form and grammar (much of which was lost), knowledge of meaning was likely to be retained by listeners two weeks later. Overall, in this study, immediately after listening, listeners were able to recognize the meaning of 35.4% of the target words and recall the meaning of 5.3% of the target words and they were able to retain this knowledge three weeks later. The readers in this study, on the other hand, lost some of their initial gains. On the meaning recognition test, readers demonstrated knowledge of 52.9% of the target words immediately after reading, but retained knowledge of 45.4% of the words three weeks later. On the meaning recall test, readers demonstrated knowledge of 13.4% of the target words immediately after reading and retained knowledge of 9.6% of the words three weeks later. Readers, therefore, lost 14.2% of their initial gains in meaning recognition, and 28.4% of their gains in meaning recall.

This interesting finding, that weeks after exposure, loss of vocabulary knowledge was greater for readers than listeners, has also been shown in previous research (i.e., Brown et al., 2008; Vidal, 2011). As Vidal (2011) pointed out, this might be due to the role of phonological memory in vocabulary learning. While spoken input gains direct access to the phonological storage in the working memory, written input needs to be recoded into phonological code to become stored in the phonological storage (Baddeley, 2007). "It seems reasonable to suppose that direct ... access to the phonological storage might result in more stable, distinct, and durable memory traces"

(Vidal, 2011, p. 244). Hence, it appears that this is where the value of listening for the incidental learning of word meaning lies: what is gained, although limited, is well retained.

Frequency of occurrence

Similar to previous studies, the data in this study also show that the more frequently words are met, the more likely they are to be acquired. This is especially true of words met in written texts. While in the reading group there was a significant increase in gains from FB1 to FB2 and from FB2 to FB3 (but no increase in gains from FB3 to FB4), in the listening group the difference in gains between the four frequency bands was not significant. It appears that for frequency of occurrence to have a significant impact on vocabulary gains through listening, much more than 20 occurrences are necessary. In Brown et al. (2008) also, no significant differences were found between the frequency bands for the listening group, which the authors attribute to floor effects. Thus, in this study, as documented in previous studies (i.e., Brown et al., 2008; van Zeeland & Schmitt, 2013a; Vidal, 2011), the effect of frequency of occurrence is smaller in listening than reading. This is supported by Vidal (2011), who found that while frequency of occurrence is the most important predictor of vocabulary gains through reading (when compared with three other factors, i.e., type of word, type of elaboration, and predictability from word form and parts), it is the least important predictor of gains through listening.

In this study, the greatest increase in gains for the listening group occurred between FB1 (2-5 occurrences) and FB2 (7-10 occurrences). Beyond FB2, frequency of occurrence appeared to have no impact on the immediate acquisition of word knowledge through listening. Similar results were found in van Zeeland and Schmitt (2013a); gains of word form and grammar occurred between 3 and 7 occurrences, yet beyond 7 and up to 15 occurrences, no further gains were observed. On the other hand, for the reading group in this study the greatest increase in gains occurred between FB2 (7-10 occurrences) and FB3 (12-15 occurrences). This frequency threshold of 10+, which leads to substantially better learning of words through reading, has also been shown in Pellicer-Sanchez and Schmitt (2010), Pigada and Schmitt (2006), and Webb (2007).

Limitations and Suggestions for Future Research

There are a number of limitations to this study. Only one proficiency level was examined; however, the differential impact of reading and listening on L2 incidental vocabulary acquisition can vary with proficiency level (Vidal, 2011). Thus, it is worth exploring the research questions raised in this study at other proficiency levels. Moreover, due to practical constraints, listeners in this study received aural input, but their vocabulary gains were measured using written tests. The test of spoken form was the only test administered aurally and the test of written form inevitably had to be written. However, the tests of part of speech, syntagmatic association, and form-meaning connection were also in the written format; while this allowed the participants to complete the tests at their own pace, the mismatch between the mode of input and mode of measurement for listeners might have decreased their scores to some extent (Alali & Schmitt, 2012). Furthermore, the research design and type of vocabulary posttest (i.e., mainly multiple-choice with numerous repetitions of the target words throughout the test) used in this study did

not allow retention rates to be measured for all dimensions of word knowledge. Future research could benefit from employing a different research design (e.g., see the research design used in van Zeeland & Schmitt, 2013a) in order to more fully capture retention rates for different dimensions of word knowledge. Finally, since listening appears to be less effective than reading for L2 incidental word learning and has been reported by L2 learners to be their least preferred input mode when compared to reading and reading-while-listening (Brown et al., 2008), attempts at providing further understanding of the complex process of L2 incidental word learning from listening and identifying factors that might play a role in facilitating this process would indeed be worthwhile.

The findings of this study clearly revealed the significance of reading and listening as sources of incidental word learning, particularly reading, which can lead to considerable gains in many dimensions of word knowledge. However, while reading results in greater vocabulary gains, listening leads to lower forgetting rates and thus, ideally, learners should be exposed to both modes of input. This research, therefore, strongly supports the value of extensive reading and listening for L2 word learning. Furthermore, the results of this study showed that recall of word meaning is particularly difficult to acquire incidentally, which emphasizes the need to complement incidental vocabulary acquisition with intentional vocabulary teaching and learning (Schmitt, 2008). On the other hand, acquisition of word form (at least to a recognition level) seems to benefit from incidental learning to a great extent and thus might require less deliberate attention during class time. Moreover, frequency of word occurrence was shown to impact incidental vocabulary gains through reading, but not through listening. Frequency of occurrence therefore is an important factor to consider when developing graded written texts; but for the development of graded spoken texts, it seems to be more practical and worthwhile to focus on variables other than frequency, such as informativeness of context. This calls for future studies to investigate different factors that might facilitate incidental vocabulary acquisition through listening, which when compared to reading, has received less attention from L2 researchers and appears to be more challenging for L2 learners.

References

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal, 3*, 153–180. doi: 10.1002/tesj.13

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press.

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

Chang, L. (2012). *Investigating the relationships between Chinese university EFL learners' metacognitive listening strategies and their comprehension and incidental vocabulary acquisition from listening tasks* (Doctoral dissertation). University of Auckland, New Zealand.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*, 693–713. doi: 10.1093/applin/amq031

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414. doi: 10.1111/lang.12052

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, *8*, 103–128. doi: 10.1080/0958822940080202

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning*, *5*, 1–32.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, *16*, 409–441. doi: 10.1093/applin/16.4.409

Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). *A conceptual and computational approach with SPSS and SAS: Analysis of variance designs*. Cambridge, UK: Cambridge University Press.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*, 303–317. doi: 10.1017/S0272263199002089

Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal*, *83*, 367–384. doi: 10.1111/0026-7902.00028

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, *61*, 355–382. doi: 10.1353/cml.2005.0018

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.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, *21*, 181–193.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Meara, P. (2013). Imaginary words. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–4). Malden, MA: Wiley-Blackwell.

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

Nation, I. S. P. (2012). *The Vocabulary Size Test.* Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, I. S. P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*, 355–380.

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

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition*, *21*, 195–224.

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

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

Read, J. (1997). Vocabulary and testing. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 303–320). Cambridge, UK: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–227). Amsterdam: John Benjamins.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, *10*, 77–89. doi: 10.2307/3585941

Richards, J. C., & Schmidt, R. W. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). London, UK: Longman.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. doi: 10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88. doi: 10.1177/026553220101800103

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*, 577–607. doi: 10.1017/S0272263109990039

van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624. doi:10.1016/j.system.2013.07.012

van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*, 457–479. doi: 10.1093/applin/ams074

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*, 56–89. doi: 10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. doi: 10.1111/j.1467-9922.2010.00593.x

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. doi: 10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65. doi: 10.1093/applin/aml048

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge:

Depth versus breadth. *Canadian Modern Language Review, 53*, 13–40.

Appendix A

Frequency Bands, Target Words, and Non-words

Frequency Band	Target Word	Number of Occurrences	Part of Speech	Non-word
2-5	chair	2	N.	bartle
	big	3	Adj.	scally
	tea	4	N.	lorey
	smiled	5	V.	kemble → kembled
7-10	watched	7	V.	bamber → bambered
	warm	8	Adj.	turley
	noise	9	N.	gamage
	window(s)	10	N.	mollet(s)
12-15	laughed	12	V.	gummer → gummered
	living-room	13	N.	palote
	afraid	14	Adj.	alden
	bed	15	N.	hislop
17-20	old	17	Adj.	galpin
	asked	18	V.	mundy → mundied
	husband	19	N.	pegler
	hand	20	N.	lomax

Appendix B

Tests 1 to 6 Used in the Vocabulary Posttest: Details and Examples

Test1-SF: Recognition of spoken form

[This measure had an aural multiple choice format; participants heard twice the target word and three distracters from a recording and had 5 seconds to check the box corresponding to the correct spoken form of the target word.]

Example:

Participants heard:

Which pronunciation is correct? Please check the box.

Number one [2sec]   A bartle [2sec]   B bertel [2sec]   C burdle [2sec]   D bardel [2sec.]
Number one [2sec]   A bartle [2sec]   B bertel [2sec]   C burdle [2sec]   D bardel [5sec.]

At the same time, the participants saw on the test page:

Which pronunciation is correct? Please check (✔) the box.

    1. ☐ A   ☐ B   ☐ C   ☐ D

Test2-WF: Recognition of written form

[This multiple-choice test consisted of the target word and three distracters. The same distracters used for the test of spoken form were used for this test.]

Example:

Which spelling is correct? Please check (✔) the box.

    1. ☐ bartle   ☐ bertel   ☐ burdle   ☐ bardel

Test3-Mg-Recall: Recall of form-meaning connection

[Meaning recall was measured using a translation test.]

Example:

Translate into Farsi.

    1. bartle..

Test4-PS: Recognition of part of speech

[For this test, the target word was presented in three different sentences. Each sentence used the target word as a different part of speech. Only one of the sentences was correct, and the other two were distracters. In order to avoid any learning effects on the tests that followed, sentences were created in such a way that no clues to the meaning of the target words were provided.]

Example:

Which sentence is correct? Please check (✔) the box.

    1. bartle ☐ It is a bartle. (Noun)
                ☐ He is very bartle. (Adjective)
                ☐ She bartled. (Verb)

Test5-SA: Recognition of syntagmatic association

[In this test, the target word was presented followed by four choices: one choice was in a sequential relationship with the target word and the other three choices were distracters. All choices were in the same word class. Because the correct option was a target word in the passage, all the distracters were chosen from the passage as well.]

Example:

1. Which word is more likely to be used with **bartle** in a sentence? Please check (✔) the box.

   ☐ sit   ☐ go   ☐ open   ☐ stop

Test6-Mg-Recog: Recognition of form-meaning connection

[In this final test, the target word was presented followed by four options: the original real English word which it had replaced in the text and three distracters. The distracters belonged to the same word class. Because the correct option had not been read or listened to in the passage, all the distracters were chosen from outside the passage as well.]

Example:

1. Which is the correct meaning for **bartle**? Please check (✔) the box.

   ☐ book   ☐ chair   ☐ food   ☐ head

About the Author

Sarvenaz Hatami is an assistant professor of TESOL in the Department of Linguistics at California State University, Long Beach. Her areas of research interest include second language vocabulary learning and teaching and teacher education. E-mail: sarvenaz.hatami@csulb.edu