LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

SYSTEMATIC REVIEW ARTICLE

The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies

Takumi Uchihara, Stuart Webb, and Akifumi Yanagisawa

University of Western Ontario

This meta-analysis aimed to clarify the complex relationship between repetition and second language (L2) incidental vocabulary learning by meta-analyzing primary studies reporting correlation coefficients between the number of encounters and vocabulary learning. We synthesized and quantitatively analyzed 45 effect sizes from 26 studies ($N = 1{,}918$) to calculate the mean effect size of the frequency–learning relationship and to explore the extent to which 10 empirically motivated variables moderate this relationship. Results showed that there was a medium effect ($r = .34$) of repetition on incidental vocabulary learning. Subsequent moderator analyses revealed that variability in the size of repetition effects across studies was explained by learner variables (age, vocabulary knowledge), treatment variables (spaced learning, visual support, engagement, range in number of encounters), and methodological differences (nonword use, forewarning of

This research received no specific grant from any funding agency. We are grateful to Judit Kormos, *Language Learning* reviewers, Luke Plonsky, Yo In'nami, Tatsuya Nakata, and Akira Murakami for their constructive feedback on data analysis and earlier versions of the manuscript. We also thank the following researchers who graciously provided information necessary for the current meta-analysis to be completed: Ana Pellicer-Sánchez, Marije Michel, Nina Daskalovska, Niousha Pavia, Sarvenaz Hatami, and Yanxue Feng.

This article has been awarded an Open Data badge. All data are publicly accessible via the Open Science Framework at https://osf.io/rmnk2. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Takumi Uchihara, University of Western Ontario, Faculty of Education, 1137 Western Road, London, ON N6G 1G7, Canada. E-mail: tuchihar@uwo.ca

DOI: 10.1111/lang.12343

an upcoming comprehension test, vocabulary test format). Based on the findings, we suggest future directions for L2 incidental vocabulary learning research.

Keywords frequency; vocabulary learning; incidental; second language; meta-analysis

Introduction

Research on incidental vocabulary learning has suggested that the number of times learners encounter an unknown word in context affects how likely they will learn that word successfully (e.g., Brown, Waring, & Donkaewbua, 2008; Horst, Cobb, & Meara, 1998; Peters & Webb, 2018; Webb, 2007). However, first language (L1) studies (e.g., Elley, 1989; Jenkins, Matlock, & Slocum, 1989; Saragi, Nation, & Meister, 1978) have varied considerably in the number of encounters that they have found learners need in order to learn words. Second language (L2) studies have observed even greater variability for the number of encounters necessary for learners to learn words, ranging from six (Rott, 1999), eight (Horst et al., 1998), 10 (Webb, 2007), 12 (Elgort & Warren, 2014), to more than 20 encounters (Waring & Takaki, 2003). This inconsistency might be attributed to learner-related variables, such as L2 proficiency (Zahar, Cobb, & Spada, 2001), and to methodological differences between studies, such as spaced versus massed learning conditions (Webb, 2014). Due to such complexities and such mixed findings, the extent to which frequency of encounters can be a reliable predictor of incidental vocabulary learning has remained unclear. Moreover, there has also been a lack of clarity about the extent to which learner and methodological variables affect the role that repetition plays in incidental vocabulary learning. With several decades having passed since Saragi et al. (1978) conducted their seminal study of the relationship between repeated encounters and incidental vocabulary learning, the time was ripe for a meta-analysis of this topic. The present study therefore aimed to clarify the complex landscape of frequency effects on L2 incidental vocabulary learning by conducting a meta-analysis with two goals: (a) to obtain an overall mean correlation between repeated encounters and L2 incidental vocabulary learning and (b) to determine how learner-related, treatment-related, and methodological variables account for variation related to the frequency effects.

Frequency and Vocabulary Learning

Incidental Vocabulary Learning

Defining the construct of incidental vocabulary learning is challenging (Gass, 1999) because vocabulary research has operationalized the construct in numerous ways (Bruton, López, & Mesa, 2011). Researchers have generally operationalized learning words incidentally in a methodological sense as learning that occurs when learners are not forewarned of an upcoming vocabulary test. The argument for this definition has been that, if learners know of an upcoming vocabulary test, they will pay special attention to target vocabulary and engage in intentional learning (Hulstijn, 2003; Peters, Hulstijn, Sercu, & Lutjeharms, 2009). A second prevalent definition of incidental vocabulary learning has been that it occurs as a byproduct of meaning-focused activities (Hulstijn, 2003, p. 362). In contrast to studies using word-focused tasks (e.g., flashcard learning) that have been viewed to be more intentional, incidental learning studies have often used meaning-focused comprehension tasks (e.g., reading a novel, listening to lectures). In these tasks, learners' attention is directed toward text comprehension and away from vocabulary learning (Swanborn & de Glopper, 1999). Nevertheless, neither the methodological operationalization (i.e., the absence of posttest announcement) nor task selection (i.e., using a reading comprehension task instead of a flashcard learning task) has eliminated the possibility that some learners may be involved in more intentional learning than others (Bruton et al., 2011; Pellicer-Sánchez & Schmitt, 2010). In the current study, we defined incidental vocabulary learning as the learning that emerges through a meaning-focused comprehension task in which learners are not told of an upcoming vocabulary test. We also acknowledge that there was no control over the variability of intentional learning across individuals, and we considered incidental vocabulary learning as a methodological rather than a theoretical construct (Hulstijn, 2003; Peters et al., 2009).

Frequency of Encounters

Despite the large body of research on incidental vocabulary learning through meaning-focused input, the number of encounters necessary for learning to occur has remained unclear. Saragi et al. (1978) conducted a landmark study revealing the important role of repetition for incidental vocabulary learning through reading. They found that native speakers of English needed to encounter target words at least 10 times before substantial learning occurred. Numerous researchers have since expanded on Saragi et al. to further explore the role that repetition plays in incidental learning with L2 learners (Brown et al., 2008; Horst et al., 1998; Malone, 2018; Pellicer-Sánchez & Schmitt, 2010; Peters & Webb, 2018; Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003; Webb, 2007; Webb & Chang, 2015; Webb, Newton, & Chang, 2013). These studies revealed that the number of encounters necessary for substantial learning to take place varied to a large degree. Such inconsistencies in results were an inevitable consequence of researchers' varied approaches to conceptualizing and measuring vocabulary learning. If word knowledge is defined as form recognition and a third of the words encountered (or 33%) being learned is considered reasonable (Laufer & Rozovski-Roitblat, 2015),[1] one encounter might suffice for learning to occur from reading (43% of the total words encountered in Chen & Truscott, 2010; 67% in Webb, 2007). Yet, acquiring meaning recognition probably requires more than one encounter. For example, researchers have suggested that two (45% of the words learned in Rott, 1999), two to four (33% of the words learned in Pellicer-Sánchez & Schmitt, 2010), eight to 10 (54% of the words learned in Waring & Takaki, 2003), and 10 or more encounters (36% of the words learned in Pigada & Schmitt, 2006) are required. Researchers have suggested that acquiring recall knowledge is even more difficult. For example, recalling forms of 27% of the target words required seven encounters in Chen and Truscott (2010), and recalling meanings of 29% of the target words required 10 encounters in Webb (2007). According to Pellicer-Sánchez and Schmitt (2010), at least 10 to 17 encounters might be necessary for word meanings to be recalled (48%). Vocabulary knowledge also involves multiple aspects other than knowledge of a form–meaning connection (Nation, 2013), and research has revealed that even more encounters might be necessary for learning such aspects of word knowledge as collocation, syntactic, word class, or association information (e.g., Pellicer-Sánchez & Schmitt, 2010; Webb, 2007; Webb et al., 2013).

To complicate matters, timing of administering vocabulary posttests also plays an integral role in determining the relationship between repetition and vocabulary learning. According to Rott (1999), six encounters may not be sufficient for word learning as evidenced by a loss of 12% of word knowledge (45% vs. 33% of the words considered acquired) on a meaning translation test after one month. In Waring and Takaki (2003), learners recalled 2.1 out of 5 word meanings immediately after reading, but they recalled only 0.3 words three months later (42% vs. 6%). Similarly, Webb and Chang (2015) reported that more than 25 encounters led to learners recognizing nearly half of word meanings after one week (47.63%), but they had retained only 7.2% of them after three months. In contrast to earlier studies measuring word knowledge as a product of meaning-focused input (i.e., explicit knowledge), recent studies have started exploring real-time processing involved in vocabulary learning as well as speed of word recognition (i.e., procedural and implicit knowledge) using sensitive measures, such as eye-tracking and priming techniques (Elgort, Brysbaert, Stevens, & Van Assche, 2018; Elgort & Warren, 2014; Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016). Sensitive measures are different from explicit measures in that the former capture learners' lexical knowledge during language processing under time pressure, whereas the latter provide an indication of learners' ability to recognize or recall word forms or meanings without time restrictions (e.g., multiple-choice, translation, gap-filling, or matching tests). The mounting evidence based on sensitive measures has provided support for earlier studies (e.g., Webb, 2007), suggesting that the role of repetition in vocabulary learning depends on the choice of measures used to assess learning (e.g., form recognition, meaning recall). As highlighted above, it is important to define the construct of vocabulary knowledge in incidental learning research. The focus of the current study was therefore on explicit word knowledge primarily as it relates to form and meaning, instead of procedural or implicit word knowledge.

Correlation Between Frequency of Encounters and Learning Gains

Correlations between repetition and vocabulary learning have given useful insights into the role of repetition in incidental vocabulary learning. Researchers have traditionally calculated a repetition–learning correlation by comparing the number of times each word occurred in a text with the number of participants who learned each word (Horst et al., 1998; Saragi et al., 1978). The strength of this research design relates to its high degree of ecological validity. Researchers have normally used existing books (Daskalovska, 2016) or commercially accessible television programs (Rodgers, 2013) without any significant manipulation of the number of encounters with target words. In this regard, learners' experience with the texts reflect real-life situations. The drawback of this approach, however, is a lack of experimental control over the number of encounters, which inevitably varies widely from one word to another (e.g., one to 209 in Saragi et al., 1978). Crucially, in such authentic materials, the number of encounters is likely to be correlated or confounded with other lexical or contextual variables, such as relevance of the items to text comprehension or intrinsic difficulty of the items (Hulme, Barsky, & Rodd, 2018). Certain researchers (Chen & Truscott, 2010; Webb, 2007; Webb et al., 2013) made attempts to overcome this problem by controlling for the number of encounters with each target word. Webb (2007), for instance, randomly assigned participants to one of the four treatment conditions spanning one, three, seven, and 10 encounters. Webb exposed each experimental group to the same set of target words a different number of times, which allowed for comparisons between groups. This between-participants design, as opposed to within-participants design (e.g., Saragi et al., 1978), has allowed researchers to attribute learning purely to frequency effects while they have controlled for other confounding variables.

Because Saragi et al. (1978) found a moderate correlation between frequency and gains in meaning recognition ($r = .34$), many researchers have since followed this practice in order to explore the roles of frequency in L2 domains. Focusing on low-intermediate English foreign language learners, Horst et al. (1998) found a relatively high correlation ($r = .49$) between frequency and meaning recognition, providing evidence in line with Saragi et al. A follow-up study with high-intermediate learners (Horst, 2000), however, did not find significant frequency effects ($r = .01$). Despite using the same text (i.e., *The Golden Fleece*), mode of input (i.e., reading with audio support), and test format (i.e., meaning recognition through a multiple-choice test), Zahar et al. (2001) and Tekmen and Daloğlu (2006) reported varying sizes of correlations ($r =$ .21–.40 and $r =$ .45–.56, respectively). Webb (2007) measured different aspects of receptive and productive word knowledge and found high correlations when learning gains were measured through recall test formats, such as a dictation task ($r = .50$) and a meaning translation task ($r = .43$). In studies focusing on spoken input and using a developmental scale (Wesche & Paribakht, 1996) to measure learning gains, Vidal (2003) reported a moderate correlation ($r = .34$) when English foreign language university students listened to videotaped academic lectures. A follow-up study (Vidal, 2011) found a relatively high correlation for a listening group ($r = .49$), which was nonetheless smaller than the correlation found in a reading group ($r = .69$). In a longitudinal study that involved viewing a series of episodes of an American TV show, Rodgers (2013) found a significant but small correlation on a demanding test format of meaning recognition ($r = .30$) but no significant correlation on a less demanding test format ($r = .18$).[2] Webb and Chang (2015) found a negligible size of correlation between frequency of encounters and meaning recognition ($r = -.03$) when learners engaged in an extensive reading program with audio support over 13 weeks. According to the effect-size benchmarks developed by Plonsky and Oswald (2014), where $r = .25$ (small effect), $r = .40$ (medium effect), and $r = .60$ (large effect), the effects of frequency of encounters on incidental vocabulary learning have ranged from none (e.g., –.03, .01), to small (e.g., .18, .29), to medium (e.g., .34, .49), and up to large (e.g., .56, .69).

In summary, researchers have agreed that repetition is a crucial variable promoting L2 incidental vocabulary acquisition and that considerable variability exists in the size of frequency effects on learning across studies. The view with which researchers also currently agree is that incidental vocabulary learning research should shift its focus away from trying to determine a threshold number of encounters necessary for learning to occur. Instead, researchers' focus should be on trying to understand the complexity of how frequency relates to other variables in determining vocabulary learning. Therefore, it is crucial to explore whether and to what extent different variables moderate the relationship between frequency of encounters and incidental vocabulary learning.

Review of Moderator Variables

There are many potential moderator variables influencing the relationship between repeated encounters and incidental vocabulary learning. Besides the ways for researchers to operationalize learning gains (e.g., test format, test timing), numerous other variables are believed to contribute to or determine the role that frequency plays in vocabulary learning. The potential moderators include learner variables (e.g., age, gender, proficiency, motivation, working memory, or background knowledge), word characteristics (e.g., imageability, concreteness, cognateness, number of letters and syllables, or parts of speech), text characteristics (e.g., genre, text length, richness of contextual clues, or keyness of the word for comprehension), and methodological or treatment variations (e.g., spaced vs. massed treatment conditions, or use of nonwords vs. real words) (Elgort et al., 2018; Elgort & Warren, 2014; Malone, 2018; Paribakht & Wesche, 1999; Peters, Heynen, & Puimège, 2016; Peters & Webb, 2018; Webb, 2008, 2014; Webb & Chang, 2015). Due to empirical motivations based on previous studies as well as to logistic reasons related to feasibility and reliability of variable identification and coding (e.g., general L2 proficiency), we limited our moderators to a total of 10 variables.[3] We used the literature review as the main rationale for choosing these variables—variables that either pertain to conflicting results or that remain unexplored in research. Also, the variables of interest have been considered difficult to investigate in primary research (e.g., nonword use), but they lent themselves to the meta-analysis approach that we employed in this study. Therefore, we used 10 moderator variables that we categorized as learner-related, treatment-related, and methodological variables for our investigation.

Learner Variables

Age

A recent meta-analysis on incidental vocabulary learning through spoken input found a positive effect of age (de Vos, Schriefers, Nivard, & Lemhöfer, 2018). In this study, university students outperformed children in kindergarten and elementary school. Because older learners tend to have more years of experience with the L2 and with text comprehension than younger learners, these findings indicated that older learners may use strategies more effectively to derive the meanings of unknown words. To the best of our knowledge, virtually no research has explored age effects on the relationship between frequency of encounters and vocabulary learning.

Vocabulary Knowledge

Zahar et al. (2001) found that frequency of encounters was more beneficial for learners with a smaller vocabulary size than for learners with a larger vocabulary size. Lexically proficient learners presumably have better comprehension of L2 words and sentences surrounding unknown words; thus, they can capitalize on rich contextual clues to attempt successful inferences of L2 words in fewer encounters than novice learners. However, follow-up studies (Daskalovska, 2014a; Tekmen & Daloğlu, 2006) did not support this hypothesis because the researchers did not find higher correlations between frequency and learning for lower-proficiency learners. Similarly, Elgort and Warren (2014) found a significant interaction between frequency and proficiency in predicting meaning recall, indicating that more proficient learners tend to benefit from increasing the number of encounters with unknown words. To explore these conflicting results, in the present study, we identified L2 vocabulary knowledge as an important potential moderator in incidental vocabulary learning.

Treatment Variables

Spacing

Higher correlations have often been found in massed treatment conditions where participants were given one-day treatment exposure (e.g., $r = .54$ in Tekmen & Daloğlu, 2006) as opposed to spaced treatment conditions where participants were given multiple treatment exposures over an extended period of time (e.g., $r = .18$ in Rodgers, 2013). One reason for this difference was that the interval between treatments and posttests varied greatly in spaced conditions, which might have masked frequency effects on learning (Webb & Chang, 2015). For example, Webb and Chang reported a negligible frequency effect ($r = -.03$) with the treatment–testing interval ranging from one to 13 weeks, whereas Zahar et al. (2001) found a medium effect ($r = .36$) with a two-day interval. Thus, the evidence has suggested a hypothesis that a smaller frequency effect would be found for learners in spaced learning conditions compared to those in massed learning conditions.

Mode of Input

The effect of repetition may be more salient in learning through written input than spoken input (Brown et al., 2008; Hatami, 2017; Vidal, 2011). Due to the transitory nature of spoken language and the inherent difficulty with speech segmentation, L2 listeners tend to lack the capacity for processing L2 speech efficiently and for collecting adequate contextual information so as to work out the meanings of unknown spoken words (van Zeeland & Schmitt, 2013). In reading while listening, on the other hand, repetition effects are expected to come more into play because the audio input prevents readers from skipping over or skimming through sentences where target words are embedded (Horst et al., 1998; Malone, 2018). Similarly to listening, however, reading-while-listening conditions may not be as favorable as self-paced reading because the audio track can make it difficult for readers with audio support to pause or go back to earlier sentences when they have problems with text comprehension. Emerging studies on viewing have also pointed to the importance of repetition for word learning (Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013). Peters and Webb (2018) suggested that frequency of encounters plays a different role in watching L2 television than in listening due to the presence of visual support. Research comparing the effects of frequency in different modes of input is lacking. Therefore, the effect of frequency on incidental vocabulary learning in different input modes deserves investigation.

Visual Aid

Research has shown that the presence of visual imagery occurring with written or aural input improves vocabulary learning. Horst et al. (1998) suggested that the presence of pictures in a book improved vocabulary learning. Elley (1989) also suggested that illustrations supported vocabulary learning when children listened to stories. Neuman and Koskinen (1992) found vocabulary learning was significantly greater when learners viewed television with captions than when they read. Visual support may make words more salient, and this could perhaps reduce the effects of frequency (Horst et al., 1998). However, how visual support affects the role of repeated encounters in learning remains to be examined (Peters et al., 2016; Peters & Webb, 2018).

Engagement

Researchers have agreed that "the more a learner engages with a new word, the more likely they are to learn it" (Schmitt, 2008, p. 338). This notion was embodied in Hulstijn and Laufer's (2001) involvement load hypothesis that proposed that the amount of attention and amount of elaboration with L2 words determine vocabulary learning (Laufer & Rozovski-Roitblat, 2015). In the context of incidental vocabulary learning, learners in controlled conditions do not have any external support or interaction with peers (Webb, 2007), whereas learners in ecologically valid conditions (e.g., extensive reading) likely engage with L2 words using dictionaries or having discussions (Webb & Chang, 2015).[4] Thus far, no attempts have been made to directly explore how engagement influences the role of repeated encounters in vocabulary learning. However, earlier studies indicated that the effect of engagement might surpass that of repetition on vocabulary learning, reducing the frequency effects that have been observed in engagement-free meaning-focused input (Laufer & Rozovski-Roitblat, 2011; Webb & Chang, 2015).

Range in Encounters

The range in number of encounters with target words has varied from study to study: 1–3 (Hulstijn, Hollander, & Greidanus, 1996), 4–8 (Pellicer-Sánchez, 2017), 1–7 (Chen & Truscott, 2010), 2–17 (Horst et al., 1998), 5–54 (Rodgers, 2013), 1–70 (Webb & Chang, 2015), and 1–209 (Saragi et al., 1978). The absence of significant correlations between frequency of encounters and learning may be attributed to a restricted range in the number of encounters (Pellicer-Sánchez, 2017). Crucially, a limited range of encounters (e.g., one to three) tends to underestimate the resulting correlation because the data stem from a mere fragment of the full-scale, "true" relationship that would be detected with a wider range of encounters (e.g., one to 209) (see Thorndike, 1949, for discussion of this issue in greater detail). Therefore, it was reasonable for us to hypothesize that the wider the range in the number of encounters, the higher the correlation would be.

Methodological Variables

Nonword Use

A main methodological concern in incidental vocabulary learning research is how to control for learners' preexisting knowledge of target words and how to ensure that learning gains are purely attributed to treatments. One way to deal with this issue is to use nonwords (Nation & Webb, 2011; Reynolds, 2018). Researchers have normally substituted nonwords for high-frequency real words appearing in a text to make sure that participants do not engage in learning new concepts (Hatami, 2017; van Zeeland & Schmitt, 2013; Waring & Takaki, 2003; Webb, 2007). The downside of using nonwords is that the learning may not reflect actual L2 learning in a real-life situation. Building upon Webb's (2007) study using nonwords, Chen and Truscott (2010) conducted a study using real words and suggested that results based on nonwords likely lead to an overestimate of the learning that would occur in more natural situations.

However, the extent to which such an overestimate of learning affects the role of frequency has remained unanswered; therefore, it deserves further exploration.

Comprehension Test Announcement

Learners who are informed that there will be a comprehension test after completing meaning-focused tasks (e.g., reading a text) may learn more vocabulary incidentally. Forewarned participants know in advance that they may be asked about both content and text features and may direct their attention to different aspects of texts, including topic-related words and grammatical connections in the passage (Swanborn & De Glopper, 2002). This might eventually bring about higher learning gains as a result of repeated encounters, given that topic-related words are often recurrent and likely chosen as target words if they are considered unfamiliar to participants. Forewarned participants may engage in meaning-focused activities with greater mental effort (Craik & Lockhart, 1972; Paribakht & Wesche, 1999), making them more attentive to every occurrence of target words in context and consequently more amenable to frequency effects.

Test Format

Previous research has suggested that recall test formats tend to yield higher correlations between repetition and learning than recognition test formats (Chen & Truscott, 2010; Peters & Webb, 2018; Webb, 2007). The reason may be that other noise variables, such as wild guessing, can be a confound in recognition measures. Thus, recall tests may provide a more accurate indication of knowledge of form–meaning connection (Chen & Truscott, 2010; Webb, 2007).

Research Questions

As our review above has shown, the relationship between frequency of encounters and vocabulary learning appears to be influenced by many variables pertaining to individual differences and methodological variations across studies. In order to clarify the complexity of the frequency–learning relationship, we conducted a meta-analysis of correlations between frequency of encounters and incidental vocabulary learning. The following research questions guided our study:

1. What is the overall relationship between frequency of encounters and L2 incidental vocabulary learning?
2. Which empirically motivated variables moderate the relationship between frequency of encounters and L2 incidental vocabulary learning?

Method

Literature Search

As literature search guidelines have suggested (In'nami & Koizumi, 2010; Plonsky & Oswald, 2015), we employed the following databases to identify resources to include in the meta-analysis: the Education Resources Information Center (ERIC), Linguistics and Language Behavior Abstracts (LLBA), ProQuest Dissertations and Theses, PsycINFO, Google, Google Scholar, and VARGA, which is an online repository of literature on vocabulary studies available at Paul Meara's Lognostics website (http://www.lognostics.co.uk/varga). We also searched 19 journals of applied linguistics, language teaching, and second language acquisition using the search functions available in the journal websites. We searched abstracts published from May 1978 to October 2018 using various combinations of the following keywords: frequency, repetition, repeated encounters/exposure(s), incidental word/vocabulary/lexical learning/acquisition, second language acquisition/learning, foreign language learning, reading, listening, reading while listening, viewing, correlation, Pearson, and Spearman. In addition, we conducted ancestry searches by examining the references of relevant review articles (e.g., Webb, 2014) and doctoral dissertations (e.g., Horst, 2000). As a result, we screened 2,336 reports that appeared initially eligible for the meta-analysis with reference to the following selection criteria.

Inclusion Criteria

We set the following 10 criteria for including the retrieved studies:

1. The study must have measured vocabulary gains resulting from incidental learning conditions in which the target words were not directly taught or studied but were expected to be picked up through completing meaning-focused activities. Moreover, no explicit mention must have been made to participants regarding upcoming vocabulary posttests after the treatment although it was possible that participants have been told of upcoming comprehension tests.
2. The study must have concerned incidental vocabulary learning through various meaning-focused modes of input including reading, listening, reading-while-listening, and viewing activities.
3. The study must have focused on learning target words through L2 input. Studies that involved learning target words encountered in L1 input (e.g., Saragi et al., 1978) would not be included.

4. The study must have measured explicit knowledge of form–meaning connection, not procedural word knowledge using an eye-tracking technique (Godfroid et al., 2018) or implicit knowledge using response-time measures (Elgort & Warren, 2014).
5. The study must have focused on learning single-word items, not multiword items (e.g., Pellicer-Sánchez, 2017).
6. The article reporting on the study must have been written in English.
7. The study must have reported a correlation coefficient between frequency of encounters and learning gains or have provided sufficient information to calculate correlation coefficient.
8. The study must have ensured that vocabulary test scores that were subsequently compared to frequency of encounters were not confounded by preexisting knowledge of target words. Common approaches to controlling for preexisting knowledge must have included conducting pretests, using nonwords, and pilot testing of target words with another population.
9. In studies using multiple vocabulary tests or production tests (e.g., form recall), these tests must have been administered before recognition tests (e.g., multiple-choice recognition). This order of test administration was crucial, given that in the opposite order, presentation of a target word as an option would provide the answers for items in any subsequent production tests.
10. For within-participants designs, all participants must have been exposed to the same text where the number of encounters varied across target words (e.g., Horst et al., 1998). Conversely, for between-participants designs, participants must have been assigned to experimental groups in which the number of encounters varied across participants (e.g., Webb, 2007). Initially, we made attempts to include studies adopting both within- and between-participants designs. However, the number of studies identified for the between-participants design was small ($k = 3$): Al-Shehri (2015), Chen and Truscott (2010), and Webb (2007). Therefore, we included only studies adopting within-participants design.

We included studies meeting all 10 criteria in our meta-analysis. Due to the limited number of studies reporting correlations based on delayed vocabulary posttests (Ellis, 1995; Webb & Chang, 2015), our data were mainly based on immediate vocabulary posttests (mean interval between treatment and testing = 2.4 days, *range* = 0–14 days).

To minimize publication bias (the fact that studies reporting significant findings and large effect sizes tend to be submitted for publication and published), this meta-analysis included both published and unpublished works (e.g., five master's theses and doctoral dissertations, one conference proceeding, and one conference presentation). The drawback of this comprehensive approach was the possibility that the reliability of the results from the meta-analysis might be compromised. One way to deal with this issue is to examine the quality of research design. An important methodological consideration in incidental vocabulary learning studies is to control for preexisting knowledge of target words (Nation & Webb, 2011). The eighth inclusion criterion served as a safeguard against this potential confounding variable. In our final dataset, most studies had conducted pretests (39/45, 87%), and the remainder had used nonwords (6/45, 13%). In the vast majority of studies that had used pretests, attempts had been made to avoid test effects (e.g., alerting participants to target words) by administering pretests a day or more prior to treatment sessions (37/39, 95%, $M = 9.9$ days), and this finding was also true for the majority of unpublished studies (12/13, 92%, $M = 7.7$ days). The information about the interval between pretest and treatment was not accessible in the remaining two studies. Additionally, we conducted sensitivity analysis by examining the extent to which our meta-analysis results could be replicated with published data only (i.e., with unpublished data excluded). The analysis confirmed that our results were not likely influenced by unpublished sources of information (see Appendix S1 in the Supporting Information online). Last, information about internal consistency of measurement (e.g., Cronbach's alpha) can be used to account for the attenuating effects of measurement unreliability in weighting effect sizes (Jeon & Yamashita, 2014). However, because only six studies reported reliability values, we could not compute a weighted effect size based on the information provided.

We contacted authors and gratefully received from six authors additional information that was needed to complete our meta-analysis (A. Pellicer-Sánchez, M. Michel, N. Daskalovska, N. Pavia, S. Hatami, and Y. Feng). Although the data from a study conducted by Ferris (1988) were no longer available (D. Ferris, personal communication, March 6, 2018), information from other papers that have cited her work (Daskalovska, 2014a; Horst, 2000; Krashen, 1989) was sufficient. Therefore, we also included Ferris's (1988) study in our dataset. In all, we identified and selected for this meta-analysis 26 studies ($N = 1{,}918$) providing a total of 58 effect sizes. Although we adopted a comprehensive approach, several studies cited in the literature review of this article (e.g., Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Waring & Takaki, 2003) did not meet our selection criteria, and we did not include them in the data analysis.

Coding

We coded the selected studies using the coding scheme table presented in Appendix S2 in the Supporting Information online. The table illustrates 17 variables, including 10 moderator variables related to learner, treatment, and methodology, as well as information regarding identification (publication author, title, year, type), sample sizes, and effect sizes. To deal with the issue of multiple effect sizes (because meta-analysis based on multiple effect sizes produced from the same participants violates the requirement of independence of observations) and to select a single effect size, we averaged multiple effect sizes prior to the meta-analysis (Plonsky & Oswald, 2015). This averaging method enables meta-analytic outcomes to be comprehensive without any unnecessary loss of valuable data. As a result, we reduced the 58 effect sizes to a sample of 45 independent effect sizes, 13 of which were from four replication studies (Daskalovska, 2010, 2014b; Reynolds, 2018; Tekmen & Daloğlu, 2006) and one original study conducted by Zahar et al. (2001).

To establish the reliability of the coding procedures, we randomly selected 10 studies (35%) and two of the authors independently coded them. Following Boulton and Cobb's (2017) approach, we calculated the number of discrepancies between the two researchers' codings and found the agreement to be 98%. After we had resolved all disagreements through discussion, the first author coded the remaining studies.

Moderator Variables

We coded a total of 10 moderator variables using the following criteria.

Age

Due to the limited number of studies reporting the age of participants (19/45, 40%), following de Vos et al. (2018), we used grade levels to indicate this variable: primary school, secondary school, and university levels.

Basic Vocabulary

Due to inconsistencies in the information of vocabulary test scores reported by researchers, we defined vocabulary knowledge in a narrow sense as basic vocabulary. This refers to the knowledge of the most frequent 2,000 word families (Nation, 2013) measured by one of the frequency levels in the Vocabulary Levels Test (Nation, 1990; Schmitt, Schmitt, & Clapham, 2001; Webb, Sasao, & Ballance, 2017). We included the reported mean score as a percentage score (0 to 100%). A total of 22 studies were available for this variable ($M = 74$, $SD = 14$, $range = 50–94$).

Table 1 Exposure categories coded for spaced and massed learning conditions

Category	Description	Example study	Code
Single exposure	Exposed to a single text within a day	Hatami (2017)	Massed
Self-paced multiple exposures	Exposed to long texts (e.g., an entire novel) in their spare time for more than a day	Daskalovska (2016), Ferris (1988)	Spaced
Controlled multiple exposures within a day	Exposed to multiple short texts in class within a day	Van Zeeland and Schmitt (2013)	Massed
Controlled multiple exposures over time	Exposed to multiple texts in class for more than a day	Horst et al. (1998), Rodgers (2013)	Spaced
Repeated exposures within a day	Exposed to the same text multiple times within a day	Tekmen and Daloğlu (2006), Zahar et al. (2001)	Massed
Repeated exposures over time	Exposed to the same text multiple times for more than a day	Neuman and Koskinen (1992)	Spaced

Spacing

The majority of vocabulary studies comparing spaced and massed learning have been conducted in a paired-associate paradigm (e.g., Nakata, 2015). Spacing has often been operationalized within a strictly controlled laboratory setting in which participants study individual L2 items in isolation at different time intervals. Given that a simple application of such operationalization to our meta-analysis might not have been appropriate because our focus was on incidental learning, we initially identified and categorized treatment procedures commonly employed in this area, as summarized in Table 1. Table 1 suggests that the number of text exposures participants received varied greatly across studies (e.g., an entire novel, multiple graded readers, multiple TV episodes). Therefore, regardless of the number of texts or treatments, when treatment sessions were completed within a single day, we coded the study as a massed condition; when treatment sessions lasted for more than two days, we coded the study as a spaced condition. For example, we coded two studies (Hatami, 2017; van Zeeland & Schmitt, 2013) as massed conditions despite the difference in the number of text exposures. In the former, participants listened to a single text

(4,231 words) within a single day, whereas, in the latter, participants listened to four short texts ($M = 1{,}110$ words) within a single day.

Mode of Input

This variable consisted of five categories: reading, listening, reading-while-listening, viewing, and other (i.e., combination of two or more modes). We initially further categorized viewing into subgroups according to the presence/absence of captions; however, due to small sample size for the viewing-without-caption category ($k = 2$), we combined all relevant studies under the category of viewing ($k = 6$).

Visual Aid

We coded this variable based on whether any visual information related to treatment texts was available to participants. Visual support was present in viewing conditions (Rodgers, 2013), and it was also present in written texts (e.g., graded readers) containing pictures (Horst et al., 1998; Webb & Chang, 2015).

Engagement

The majority of effect size samples ($k = 31$) were based on highly controlled conditions where participants were not allowed to have any external support (e.g., dictionary use) or communicative activities (e.g., post-reading activities). However, 11 samples came from studies with a high degree of ecological validity ($k = 11$) in which students had opportunities to engage with L2 words. Engagement included using dictionaries (Webb & Chang, 2015), asking questions (Ellis, 1995), taking notes (Vidal, 2003), and having discussions (Horst, 2000).

Range in Encounters

We calculated this variable by subtracting the minimum number of encounters from the maximum number of encounters ($M = 20.81$, $SD = 19.72$, *range* = 5–111). For example, in Horst et al. (1998), the minimum was 1 and the maximum was 17; hence, the range of encounters was calculated as 16.

Nonword Use

Although most studies used real words as target items ($k = 39$), some studies ($k = 6$) replaced real L2 words in a treatment text with nonwords that are phonologically and orthographically plausible in the target language (e.g., *vinse*, *grike*, *droil* in van Zeeland & Schmitt, 2013).

Comprehension Test Announcement

Following a previous meta-analytic study on L1 incidental vocabulary learning (Swanborn & de Glopper, 1999), we coded this variable as present when studies explicitly announced that comprehension tests would come after treatments. Conversely, we coded the variable for studies that did not include an announcement of a subsequent comprehension test as not present.

Test Format

We coded the format of vocabulary measurement as recognition, recall, and Vocabulary Knowledge Scale (Wesche & Paribakht, 1996). Recognition knowledge was typically measured through multiple-choice and yes/no checklist tests, whereas recall knowledge was measured through translation and gap-filling tests. In addition, we assigned a separate category to studies using Wesche and Paribakht's developmental scale, the Vocabulary Knowledge Scale, to measure vocabulary gains. Although acknowledging longstanding issues with this test format (see Schmitt, 2010), we included the data in the current meta-analysis for several reasons. First, the Vocabulary Knowledge Scale has been extensively used in incidental vocabulary learning research, and we believe examining the results based on the test in comparison to other test formats will further inform researchers' understanding of the relationship between frequency and vocabulary learning. Also, the quality of the data from those studies using the Vocabulary Knowledge Scale was sound, given that they met our inclusion criteria and that the researchers had made efforts to increase the validity of the test scores (Vidal, 2003, 2011; Yang & Sun, 2013).[5]

Data Analysis

We used the Comprehensive Meta-Analysis (Version 3.3) software to calculate the mean effect size and conduct moderator analysis for 10 moderator variables. Prior to the effect size aggregation and moderator analysis, we conducted three analyses to assess the extent to which publication bias influenced the current data set: (a) fail-safe N, (b) Orwin's fail-safe N, and (c) the trim-and-fill method (Borenstein, Hedges, Higgins, & Rothstein, 2009). All three measures consistently indicated that there was little concern regarding the influence of publication bias on our meta-analysis findings (see Appendix S3 in the Supporting Information online for detailed information regarding the results of these three measures and the funnel plot).

The basic unit of analysis was the Fisher's z score transformed from the correlation coefficients (r) retrieved from each included study. We then converted Fisher's z back to r in reporting the results for the sake of interpretability.

Following earlier meta-analytic studies based on correlations (Jeon & Yamashita, 2014; Li, 2016), we used Fisher's z instead of r because of its better statistical properties, including normal distribution and stable variance. We employed a random-effects model to compute the inverse-variance weighted mean correlation and a mixed-effects model for subsequent moderator analysis. We included only categories with three or more effect sizes for effect size aggregation and moderator analysis (Li, 2016). In effect-size aggregation, we conducted the homogeneity test using a within-group Q statistic to examine whether there would be a significant variation in true effect sizes across studies. For moderator analysis, we computed a between-group Q value for categorical variables, and we conducted meta-regression analysis for continuous variables. To increase the interpretability of the results regarding continuous variables, we conducted Pearson r and Spearman *rho* correlation analyses after we had applied both between and within inverse-variance weighting to the effect sizes.

Results

Effect Size Aggregation

To answer to the first research question that asked about the overall relationship between repetition and incidental vocabulary learning, we aggregated 45 effect sizes to produce a weighted mean effect size and a 95% confidence interval (Figure 1). Results showed that the overall relationship between repeated encounters and vocabulary learning was significant, $r = .34$, 95% CI [.27, .40], $p < .001$, a medium effect according to Plonsky and Oswald's (2014) effect size benchmarks. The homogeneity test was statistically significant, $Q(44) = 95.15, p < .001$, indicating that variability in true effect across studies as well as sampling error could have created this difference. We then conducted a series of moderator analyses to examine the extent to which empirically motivated variables could account for this variability.

Moderator Analysis

To answer the second research question that asked which and to what extent empirically motivated variables could moderate the relationship between repetition and incidental vocabulary learning, we performed moderator analyses with 10 variables related to learners (age and basic vocabulary), treatment (spacing, mode of input, visual aid, engagement, and range in encounters), and methodology (nonword use, comprehension test announcement, and test format). Basic vocabulary and range in encounters were continuous variables; all the other variables were categorical variables. Table 2 presents the

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Pellicer-Sanchez & Serrano (2017) sample 1	-0.086	-0.490	0.348	-0.376	0.707
Webb & Chang (2015)	-0.030	-0.280	0.224	-0.229	0.819
Horst (2000)	0.010	-0.593	0.606	0.028	0.977
Feng (2017) sample 3	0.014	-0.502	0.523	0.049	0.961
Pellicer-Sanchez & Serrano (2017) sample 2	0.023	-0.266	0.308	0.153	0.879
Daskalovska (2014a)	0.060	-0.419	0.513	0.233	0.816
Brown (1993)	0.063	-0.152	0.272	0.571	0.568
Feng (2017) sample 4	0.066	-0.377	0.484	0.278	0.781
Daskalovska (2014b) sample 3	0.110	-0.254	0.447	0.584	0.559
Neuman & Koskinen (1992)	0.120	-0.239	0.450	0.649	0.516
Van Zeeland & Schmitt (2013) sample 2	0.204	-0.489	0.739	0.546	0.585
Feng (2017) sample 1	0.206	-0.248	0.585	0.884	0.376
Feng (2017) sample 2	0.210	-0.156	0.525	1.128	0.259
Zahar et al. (2001) sample 3	0.210	-0.138	0.512	1.187	0.235
Zahar et al. (2001) sample 2	0.220	-0.175	0.554	1.096	0.273
Yang & Sun (2013)	0.221	-0.024	0.441	1.769	0.077
Zahar et al. (2001) sample 4	0.240	-0.119	0.543	1.318	0.187
Daskalovska (2016) sample 1	0.250	-0.102	0.546	1.399	0.162
Van Zeeland & Schmitt (2013) sample 1	0.285	-0.181	0.646	1.206	0.228
Daskalovska (2014b) sample 1	0.290	-0.072	0.584	1.580	0.114
Ellis (1995) sample 1	0.290	-0.101	0.604	1.463	0.144
Daskalovska (2014b) sample 2	0.300	-0.054	0.587	1.667	0.096
Rodgers (2013) sample 1	0.300	0.164	0.425	4.199	0.000
Alahirsh (2014)	0.310	-0.183	0.679	1.241	0.214
Ellis (1995) sample 2	0.310	-0.107	0.634	1.469	0.142
Ferris (1988)	0.320	-0.046	0.610	1.723	0.085
Hatami (2017) sample 1	0.337	0.024	0.590	2.104	0.035
Daskalovska (2010)	0.340	0.134	0.518	3.167	0.002
Daskalovska (2016) sample 2	0.340	-0.023	0.624	1.840	0.066
Vidal (2003)	0.343	0.171	0.494	3.800	0.000
Rodgers (2013) sample 2	0.392	0.092	0.627	2.519	0.012
Zahar et al. (2001) sample 1	0.400	-0.026	0.703	1.847	0.065
Pavia et al. (in press) sample 2	0.424	0.003	0.717	1.973	0.049
Tekmen & Daloglu (2006) sample 1	0.477	0.160	0.705	2.843	0.004
Vidal (2011) sample 2	0.488	0.301	0.639	4.681	0.000
Horst et al. (1998)	0.490	0.182	0.710	2.985	0.003
Hatami (2017) sample 2	0.497	0.256	0.679	3.775	0.000
Reynolds (2018) sample 1	0.511	0.327	0.658	4.918	0.000
Mohamed (2009)	0.515	-0.123	0.852	1.611	0.107
Tekmen & Daloglu (2006) sample 3	0.539	0.246	0.742	3.356	0.001
Pavia et al. (in press) sample 1	0.545	0.045	0.827	2.117	0.034
Tekmen & Daloglu (2006) sample 2	0.563	0.267	0.762	3.432	0.001
Reynolds (2018) sample 2	0.633	0.488	0.744	6.882	0.000
Wenzhong & Feng (2009)	0.634	0.366	0.805	4.029	0.000
Vidal (2011) sample 1	0.687	0.575	0.774	8.793	0.000
	0.335	0.268	0.398	9.320	0.000

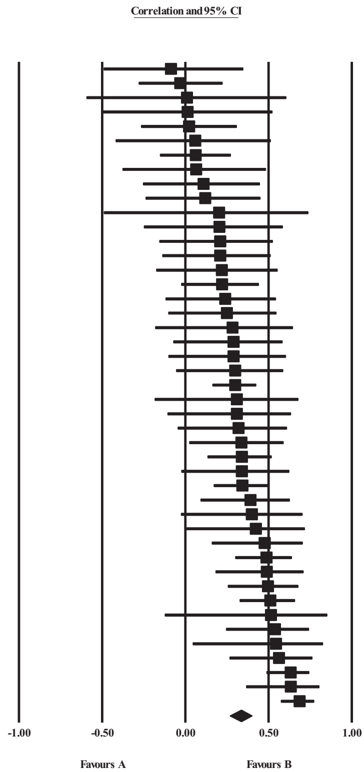

Figure 1 Overall average correlation between frequency of encounters and learning gains (indicated by a diamond) and correlations with confidence intervals for each study.

analyses for the categorical variables, and Tables 3 and 4 present those for the two continuous variables.

Learner Moderators

Table 2 shows that the Q test for age group revealed a significant difference among the three age groups: primary school, secondary school, and university. A post hoc Q test showed a significantly larger effect for university students than for secondary school students, $Q(40) = 5.59, p = .01$, indicating that university students benefitted more from repeated encounters with L2 words compared to secondary school students. The differences between primary school and university students and between primary and secondary school students did not reach

Table 2 Results of analysis of categorical moderator variables

Moderator variables	*k*	*r*	95% CI	*p*	*Q*(*df*)	*p*
Learner variables						
Age					6.19(44)	.040
Primary school	4	.20	[–.09, .47]	.180		
Secondary school	11	.23	[.13, .32]	.001		
University	30	.38	[.30, .45]	.001		
Treatment variables						
Spacing					4.91(44)	.020
Spaced learning	14	.23	[.12, .34]	.001		
Massed learning	31	.38	[.31, .45]	.001		
Mode of input					7.17(44)	.060
Reading	11	.41	[.29, .52]	.001		
Listening	11	.39	[.24, .52]	.001		
Reading while listening	16	.28	[.17, .37]	.001		
Viewing	6	.22	[.11, .32]	.001		
Visual aid					7.92(44)	.001
Yes	13	.21	[.10, .31]	.001		
No	32	.38	[.31, .45]	.001		
Engagement					9.95(43)	.001
Yes	11	.17	[.06, .29]	.001		
No	33	.39	[.32, .45]	.001		
Methodology variables						
Nonword use					6.09(45)	.001
Yes	6	.50	[.36, .62]	.001		
No	39	.30	[.24, .37]	.001		
Comprehension test announcement					4.81(42)	.020
Yes	6	.51	[.33, .66]	.001		
No	37	.30	[.23, .36]	.001		
Test format					4.55(57)	.100
Recognition	43	.29	[.22, .36]	.001		
Recall	11	.43	[.29, .56]	.001		
Vocabulary Knowledge Scale	4	.45	[.22, .64]	.001		

Table 3 Results of regression analysis of continuous moderator variables

Variable	*k*	*Q*(*df*)	*b*	95% CI	*p*
Basic vocabulary	22	2.29(21)	–0.0046	[–0.0106, 0.0014]	.120
Range in encounters	42	7.62(41)	–0.0048	[–0.0081, –0.0014]	.001

Table 4 Results of weighted correlation analysis of continuous moderator variables

Variable	*k*	*r* (*rho*)	95% CI	*p*
Basic vocabulary	22	–.43 (–.44)	[–.85, –01]	.001
Range in encounters	42	–.44 (–.50)	[–.72, –.15]	.001

significance, $Q(33) = 1.45$, $p = .22$, and $Q(14) = .03$, $p = .85$, respectively. Considering the small sample size of the primary school group ($k = 4$) and the fact that the Q test tends to be underpowered with limited samples (Borenstein et al., 2009), the results should be interpreted with caution. As for the basic vocabulary analysis reported in Table 3, mixed-effects meta-regression using the unrestricted maximum likelihood method showed a negative relationship between basic vocabulary and frequency effects (i.e., learners with a smaller vocabulary size tended to benefit more from repeated encounters with L2 words), but it did not reach statistical significance, $Q(21) = 2.29$, $p > .05$. The absence of statistical significance might relate again to the fact that Q tests are likely underpowered, particularly when analyzing small samples (i.e., for basic vocabulary, only 22 out of 45 effect sizes were available) (Borenstein et al., 2009). The weighted correlation analyses (shown in Table 4) revealed a significant negative relationship between vocabulary knowledge and frequency effects.

A closer inspection of the data (see Figure 2 for the scatterplot) suggested that this negative relationship was prominent around the samples including lexically proficient participants. Using the cutoff point of 80% for mastery of a vocabulary level suggested by Schmitt (as cited in Xing & Fulcher, 2007, p. 184), we split the data into two groups: 2,000 level mastered and 2,000 level unmastered. The further analysis revealed that the relationship was negligible but its direction was positive in the group of unmastered samples ($r = .046$), whereas the direction was negative in the group of mastered samples ($r = –.245$) (see Appendix S4 in the Supporting Information online for the scatterplots). The results confirmed that the overall negative relationship between frequency and basic vocabulary stemmed from the group of lexically proficient learners.

Notably, the opposite direction of the relationship between frequency and the two variables—age and basic vocabulary—was counterintuitive (i.e., frequency was more beneficial for older learners but less beneficial for more proficient learners). We had expected a linear relationship between age and basic vocabulary (i.e., older learners being more lexically proficient than younger learners); therefore, we anticipated a positive correlation between

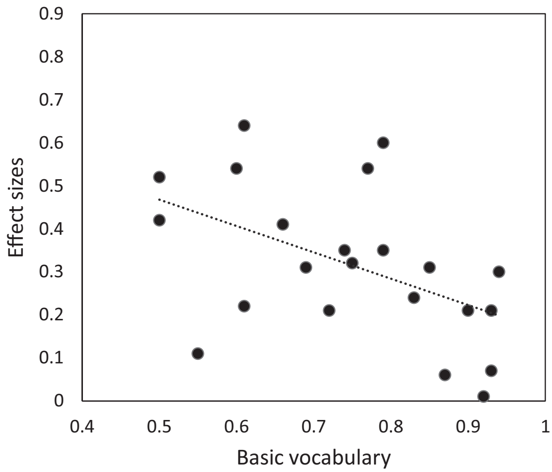

Figure 2 Relationship between basic vocabulary and effect sizes (frequency–learning association).

frequency and basic vocabulary. However, further analysis showed no significant difference in vocabulary knowledge between the two age groups[6] (indicated by the large overlap between the two confidence intervals): secondary school ($k = 5$, $M = 69\%$, 95% CI [52, 85]) and university ($k = 17$, $M = 76\%$, 95% CI [69, 83]). These findings meant that basic vocabulary knowledge and age were not related in our data. Regardless of vocabulary knowledge, older learners tended to benefit more from repeated encounters than younger learners.

Treatment Moderators

First, the Q test for the frequency effect (see Table 2) indicated that the effect was significantly larger for massed learning conditions compared to the spaced learning conditions. A further examination of the exposure type in massed and spaced conditions was conducted on single exposure, self-paced multiple exposures, controlled multiple exposures over time, and repeated exposures (the other two variables were not included in the analysis due to the small samples, $k < 3$). As Table 5 shows, learners who were repeatedly exposed to the same text within a single day benefited most from frequency. This was followed by learners who received a single exposure, and, to a much lesser extent, self-paced and controlled multiple-exposures conditions.

Second, the Q test for mode of input (see Table 2) approached statistical significance, and we found relatively large frequency effects for reading and

Table 5 Results of moderator analysis for exposure type

Variable	*k*	*r*	95% CI	*p*	*Q*(*df*)	*p*
Exposure type					14.17(40)	.001
Single exposure	18	.33	[.23, .43]	.001		
Self-paced multiple exposures	5	.19	[.05, .33]	.001		
Controlled multiple exposures over time	7	.19	[.02, .35]	.020		
Repeated exposures within a day	11	.46	[.37, .55]	.001		

listening compared to reading while listening and viewing. Third, the Q test showed visual aid to be a significant moderator (see Table 2). Learners presented with visual information during meaning-focused tasks benefited less from repeated encounters than those who had no access to the information. Fourth, regarding engagement with target words (see Table 2), we found a significantly larger effect for treatments involving no engagement compared to treatment involving engagement. Last, with regard to range in encounters, mixed-effects meta-regression analysis using the unrestricted maximum likelihood method (see Table 3) showed a negative and significant relationship between range in encounters and frequency effects, indicating that the wider the range, the smaller the frequency effects. This finding was also confirmed by subsequent weighted correlation analyses for range of encounters that Table 4 shows. The scatterplot (see Figure 3) illustrates that the size of the frequency effect declined after a range of approximately 20 encounters. We confirmed this trend by the significant difference (indicated by no confidence interval overlap) between two range groups: less than 20 encounters ($k = 32$, $r = .37$, 95% CI [.31, .44]) versus 20 or more encounters ($k = 12$, $r = .15$, 95% CI [.04, .25]). These results did not suggest any negative relationship between learning and frequency per se; rather, the results indicated that the frequency–learning correlation did not seem to increase beyond a range of around 20 encounters.

Methodological Moderators

We found a significantly larger frequency effect for nonwords than for real words (see Table 2), indicating that the effect of repeated encounters became much stronger when learners were exposed to nonwords than when they were exposed to real words. Similarly, Table 2 shows that a large effect for studies forewarning learners of an upcoming comprehension test. The Q test showed that this effect

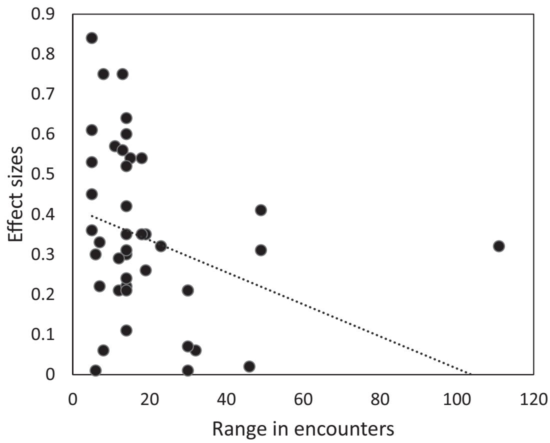

Figure 3 Relationship between range in encounters and effect sizes (frequency–learning association).

was significantly larger than it was for studies with non-forewarned learners. Finally, the results in Table 2 for test format showed larger frequency effects for the recall test format and the Vocabulary Knowledge Scale in comparison to the recognition test format, but the *Q* test did not find that the difference was significant. A post hoc *Q* test suggested a tendency for recall test formats to contribute to larger frequency effects than for recognition test formats, $Q(53) = 3.16, p = .07$. We found no significant differences found between the Vocabulary Knowledge Scale and recognition or between the Vocabulary Knowledge Scale and recall, $Q(46) = 1.85, p = .17$, and $Q(14) = 0.02, p = .88$, respectively.

Discussion

In answer to the first research question regarding the overall relationship between frequency of encounters and incidental vocabulary learning, the results showed that there was an average correlation of .34, a medium effect according to Plonsky and Oswald's (2014) effect size benchmarks. In other words, about 11% of the variance in word learning through meaning-focused input was explained by frequency of encounters. Although this result is revealing, a large effect size might be expected, considering the great theoretical value and substantial attention that many researchers have given to the role of frequency in vocabulary acquisition over the past four decades (Elley, 1989; Godfroid et al., 2018; Horst et al., 1998; Jenkins et al., 1989; Pellicer-Sánchez & Schmitt, 2010; Peters & Webb, 2018; Saragi et al., 1978; van Zeeland &

Schmitt, 2013; Vidal, 2011; Waring & Takaki, 2003; Webb, 2007; Webb & Chang, 2015; Webb et al., 2013). Our meta-analysis indicated that, although frequency of encounters is important, it should perhaps be considered as one of many variables affecting vocabulary learning (Webb & Nation, 2017). In answer to the second research question that asked whether and to what extent moderator variables would account for the variability in the effects of repetition, the results revealed that frequency–learning associations were significantly moderated by a range of variables pertaining to learner, treatment, and methodology.

Age and Basic Vocabulary

The results showed that a larger frequency effect was found for older learners (university level, $r = .38$) than for younger learners (primary and secondary school levels, $r = .23$ and .20, respectively) although their basic vocabulary knowledge (i.e., the most frequent 2,000 word families) was constant. The reason for the absence of a significant difference between university and primary school students might relate to the small sample sizes for the latter group ($k = 4$). These findings indicated that, with maturity and accumulated experience in engaging with L2 texts, learners may gradually develop literacy and strategic skills, enabling better text comprehension in terms of accuracy and automaticity. This might free up attentional resources for older learners and allow them to accumulate word knowledge efficiently over repeated encounters.

Interestingly, the result for basic vocabulary and frequency showed that vocabulary learning of lexically proficient learners was less likely explained by frequency of encounters compared to the vocabulary learning of low-proficiency learners. A further examination of the data showed that the diminishing effect of frequency was present in lexical learning by participants who had mastered the most frequent 2,000 level, whereas it might not be the case for those who had not mastered that level. These findings suggest a nonlinear relationship between vocabulary knowledge and frequency effects on incidental word learning. Beyond a certain point in vocabulary growth, learners may be able to acquire L2 words in fewer encounters and need not receive as many encounters as learners with smaller vocabulary sizes (Zahar et al., 2001). However, it is important to note that in this meta-analysis the definition of vocabulary knowledge was limited to the most frequent 2,000 word families. Although knowledge of these high-frequency words is crucial for L2 general proficiency (Nation, 2013; Webb & Nation, 2017), looking at this level alone might not reveal a full picture of the relationship between frequency effects and vocabulary knowledge of advanced learners.

Last, comparing the primary study results and our meta-analysis results is not simple. In our meta-analysis, the negative correlation between frequency effects and basic vocabulary was based on the variance in vocabulary test scores (i.e., Vocabulary Levels Test 2,000 word level) between studies (not between learners as in primary studies). In the studies that we included, the researchers made attempts to ensure that the difficulty of the treatment texts was suitable for the participants who had received the treatment. The researchers often examined vocabulary demands by using a lexical profiler (at least 16 out of 26 studies), pilot-testing the texts (Vidal, 2003), consulting experts for evaluation (Neuman & Koskinen, 1992), or using simplified versions of texts (Hatami, 2017). Thus, the results of the diminishing role of frequency for lexically proficient learners (Vocabulary Levels Test 2,000 word level > 80%) as well as its negligible effect ($r = .046$) for less proficient learners (Vocabulary Levels Test 2,000 word level < 80%) should be interpreted within the context where text difficulty was controlled for.

Spacing and Exposure Type

A smaller frequency effect was found when studies were conducted in spaced learning conditions ($r = .23$) than in massed learning conditions ($r = .38$). This result supports our prediction based on Webb (2014, p. 2) positing that "[repetition] effects may be greatest when repeated encounters occur within a short span." A further examination of the exposure type revealed that two massed conditions showed medium effect sizes (repeated exposures $r = .46$, and single exposure $r = .33$), whereas two spaced conditions showed small effect sizes (controlled and self-paced multiple exposures $r = .19$ for both). A relatively high correlation between frequency of encounters and vocabulary learning for repeated exposures indicates that reading or listening to the same text repeatedly might make unknown words more salient during the subsequent exposure. This is chiefly because by the second or third time of exposure, learners are familiar with the text content and need less effort for comprehension, thereby directing their attention to information that they might still need to work out, including unknown words.

Mode of Input and Visual Aid

Although we found no significant differences across any of the four types of input mode conditions, frequency appeared to benefit learning through reading ($r = .41$) and listening ($r = .39$) more than reading while listening ($r = .28$) or viewing ($r = .22$). We did not expect a relatively high correlation for listening and frequency because previous studies have suggested a small effect of frequency on learning through listening (Brown et al., 2008; Hatami, 2017; Vidal, 2011). The relatively large effect for listening might relate to the way that we categorized this variable, a general category of listening that encompassed a variety of spoken activities spanning listening to simplified input, engaging in spoken interaction, and listening to songs repeatedly (Ellis, 1995; Pavia, Webb, & Faez, in press). These activities were different in genre and text length from the ones that earlier studies had used to compare the two modes: graded readers (Brown et al., 2008; Hatami, 2017) and academic lectures (Vidal, 2011). Similarly, the results did not support our prediction that the controlled nature of a reading-while-listening condition would offer a favorable situation where frequency effects would figure significantly because readers' attention is focused on every word and sentence. A possible reason for this might be that dual modes of input could leave a strong trace of memory for L2 words, which might help in learning the words in fewer encounters. A recent study (Malone, 2018) suggested that aural support while reading induced deeper cognitive processing to the extent that the effect of simultaneous aural input overrode that of repeated encounters upon meaning recognition learning. Last, as for viewing, its small effect size ($r = .22$) compared to the other modes of input merits further exploration. A possible reason for this might be that other variables, such as visual images in combination with L1 and L2 subtitles (Peters & Webb, 2018; Rodgers, 2013), may attenuate the frequency effects that result from viewing L2 television. Our finding that visual aid was a significant moderator supports this explanation. Learners presented with visual information ($r = .21$) during a meaning-focused task benefited less from repeated encounters than those who had no access to the information ($r = .38$). This finding accords with the view that visual imagery heightens the salience of unknown words and that it has an attenuating impact on frequency of encounters (Horst et al., 1998).

Engagement

Studies involving engagement produced a smaller frequency effect on vocabulary learning ($r = .17$) compared to engagement-free studies ($r = .39$). Earlier studies alluded to the possibility that engagement (e.g., dictionary use, discussion of the stories, or the use of learning journals) might reduce the frequency effect on vocabulary learning (Laufer & Rozovski-Roitblat, 2011; Webb & Chang, 2015). A possible reason for this attenuating effect is that engagement with target words might override the effect of repeated encounters. For instance, according to the involvement load hypothesis (Hulstijn & Laufer, 2001), looking up the meaning of an unknown L2 word in a dictionary could induce a certain level of motivational and cognitive involvement (e.g., need to know the meaning of the word, search for the word meaning, and evaluation for judging which one of multiple meanings listed in the dictionary entry fits the context where it appears). As few as one dictionary lookup of a word could lead to larger gains than multiple encounters with the word (Laufer & Rozovski-Roitblat, 2015). Alternatively, it is possible to argue that such engagement attempts might result in intentional learning and increased attentional processing of target word forms.

Range in Encounters

The rationale for including this variable was based on the hypothesis that a restricted range of repetitions might misleadingly disguise the size of true frequency effects on incidental vocabulary learning. However, frequency effects were not found in the expected direction. The results indicated that a wider range in number of encounters led to a weaker correlation between frequency and learning. Further examination of the relationship showed that frequency effects remained prominent up to the range of around 20 encounters, after which the effects appeared to start declining. The result indicates that, although frequency affects learning positively, its effect may not remain constant for initial and later encounters (Bisson, van Heuven, Conklin, & Tunney, 2014; Godfroid et al., 2018). This is likely because, although encounters may contribute to incidental learning, there are diminishing learning gains as the number of repetitions increases beyond a certain point (Webb & Nation, 2017). Recent eye-tracking research has lent support to this claim. For example, Elgort et al. (2018) observed a plateau effect on processing of target novel words after a certain number of encounters (e.g., 8 to 10 encounters). Thus, the general view that more is better for learning may not always be true (Elgort et al., 2018; Pellicer-Sánchez, 2016). However, it should be noted that this meta-analysis focused exclusively on knowledge of form and meaning, and it is likely that different numbers of encounters enhance other aspects of word knowledge (e.g., derivation, grammatical functions) to different degrees (Webb, 2007).

Nonword Use

The difference between studies using nonwords ($r = .50$) and real words ($r = .30$) was substantial. This finding revealed that frequency effects became much more salient when participants learned words that they had never seen or heard before. By extension, this leads to the conclusion that the more fully developed knowledge of a word becomes, the weaker is the frequency effect on learning that word (Bisson et al., 2014). Another possible reason might relate to special attention elicited by nonwords (Reynolds, 2018). In reality, some of the target (real) words that are supposed to be unfamiliar to participants could be partially known to them (e.g., forms of the words might be at least recognizable). Compared to a partially known (real) word, a nonword could draw much more attention from learners because the presence of high-frequency words surrounding it might enhance the salience of the nonword. Therefore, use of nonwords in L2 research might not only overestimate the amount of learning (Chen & Truscott, 2010; Reynolds, 2018) but also inflate the size of correlations between learning gains and frequency of encounters.

Comprehension Test Announcement

As we had predicted, learners who were forewarned of an upcoming comprehension test ($r = .51$) benefited more from repeated encounters than those who were not forewarned ($r = .30$). However, the size of the effect was larger than expected ($r = .51$), which approached a large effect (Plonsky & Oswald, 2014). Forewarned learners might have been more attentive to a range of text features (e.g., topic-related words) that would be otherwise overlooked (Swanborn & de Glopper, 2002). Announcement of a comprehension test may encourage a wider scope of learners' attention to textual information that could increase the likelihood that target words in a text receive attention.

Test Format

Although the difference between recognition and recall did not reach statistical significance ($p = .07$), frequency effects appeared to be larger when vocabulary learning was measured using a recall format ($r = .43$) than a recognition format ($r = .29$). These results were in line with those of earlier studies suggesting that more demanding measures are likely to be associated with frequency than are less demanding measures (Chen & Truscott, 2010; Peters & Webb, 2018; Webb, 2007). Recall measures might be less subject to construct-irrelevant strategies (e.g., guessing) leading such test scores to better reflect learners' knowledge of form–meaning connections than recognition measures such as multiple-choice and matching tests (Kremmel & Schmitt, 2016; Webb, 2007). Another possible reason is that a greater number of encounters might not be necessary to score correctly on a recognition test, whereas it might still continue to contribute to scoring successfully on a recall test. This is mainly because receptive knowledge is easier to acquire than productive knowledge given that form recognition has been found to develop faster than form or meaning recall (Godfroid et al., 2018; Webb, 2007). Consequently, the relationship between frequency and recognition might be less pronounced than that between frequency and recall. We found a medium frequency effect for the Vocabulary Knowledge Scale format ($r = .45$).

Despite long-standing issues with the construct validity of the scale (Schmitt, 2010), it seems that the test serves the purpose of capturing incidental learning and is sufficiently sensitive to reflect accumulated information from every encounter (Wesche & Paribakht, 1996).

Conclusion

By systematically reviewing the body of research focusing on L2 incidental vocabulary learning during the last four decades since Saragi et al. (1978), this meta-analysis sought to clarify complexities involved in the relationship between frequency and incidental vocabulary learning. The findings demonstrated that although frequency of encounters is an important predictor of incidental vocabulary learning, it is not a single determiner but one of many variables affecting vocabulary learning through meaning-focused input. The results also revealed significant variation in the size of frequency effects across studies and identified a number of important variables accounting for such variability, spanning learner-related, treatment-related, and methodological variables.

We need to acknowledge several points in order to interpret our results accurately without overinterpretation. First, the fact that only six out of 25 studies (24%) reported reliability coefficients (e.g., Cronbach's alpha) for vocabulary tests is alarming. Compared to 13 earlier meta-analyses reporting reliability in different second language acquisition fields ($M = 35.4\%$, $range = 6–64\%$), our study (24%) would be ranked fifth from the bottom (Larson-Hall & Plonsky, 2015). Vocabulary researchers should be encouraged to report reliability coefficients in future studies.

Second, we found a medium correlation ($r = .34$) between frequency of encounters and vocabulary learning. However, this finding resulted from an aggregation of primary studies employing within-participants designs. In this situation, there is little to no control over the number of exposures to the target words, which might cause a confounding influence on learning, such as the influence of word properties. In fact, the mean correlation ($r = .38$, 95% CI [.23, .51]) based on three between-participants studies (Al-Shehri, 2015; Chen & Truscott, 2010; Webb, 2007) is larger than the correlation based on the within-participants studies ($r = .34$, 95% CI [.27, .40]). Although the highly controlled nature of between-participants studies (e.g., Webb, 2007) might overestimate the strength of the relationship, it is also reasonable that our meta-analysis might have underestimated it. As we stressed throughout this article, the role of frequency in lexical learning is greatly complicated by the presence of numerous other variables including individual differences and word characteristics. Therefore, a future direction for incidental vocabulary learning research should be to explore how frequency relates to other variables in the process (e.g., through eye-movement) or product (e.g., through posttest scores) of vocabulary learning rather than to determine a frequency threshold necessary for learning. With a mixed-effects modelling approach, researchers can examine altogether the effects of both learner-level (e.g., proficiency, age, motivation, working memory) and item-level (e.g., frequency, cognateness, concreteness, word class, word length) variables on learning gains with a view to obtaining deeper insights into the complex relationship between frequency and learning (Elgort et al., 2018; Elgort & Warren, 2014; Godfroid et al., 2018; Peters & Webb, 2018).

An additional point in need of acknowledgment is that our meta-analysis focused on knowledge of single-word items pertaining to form–meaning connection. The size of frequency effects or the way frequency functions in learning single-word items might be different from the way that it functions in learning multiword items. Durrant's (2014) meta-analytic study on corpus frequency and collocation learning indicated a moderate correlation between learner knowledge and frequency of two words occurring together in various types of corpora. Given the growing interest in incidental learning of collocations (Pellicer-Sánchez, 2017; Webb et al., 2013), the relationship between frequency of encounters and collocation learning also merits future meta-analysis. Other aspects of word knowledge also need more rigorous attention in incidental vocabulary learning research. Due to a lack of studies exploring aspects other than form–meaning connection (six out of 26 studies measured collocation, association, and grammatical function), we did not include them in our investigation. Future studies should measure aspects of vocabulary knowledge in addition to or apart from form–meaning connection. This would shed further light on the complex relationship between frequency and vocabulary learning.

Final revised version accepted 12 January 2019

Notes

1 In incidental vocabulary learning research, it has been common to report the rate of learning based on the proportion of words learned to the total number of target words (Swanborn & de Glopper, 1999). Laufer and Rozovski-Roitblat (2015) suggested that learning gains of a third of all target words (i.e., 33%) should be considered sufficient given that learners' attention is not explicitly directed to target vocabulary.

2 A demanding test format includes distractors among multiple-choice options that share aspects of form or meaning with the correct answer, whereas a less

demanding test format includes distractors that do not share aspects of form or meaning with the correct answer.

3 We initially coded and analyzed characteristics of target words as a moderator variable (number of letters per word, number of syllables per word, ratio of nouns to the total number of target words, word frequency, concreteness, familiarity, imageability, meaningfulness, age of acquisition). Given that we found the effects of these lexical features to be marginal, we have not reported these results due to space limitations.

4 One might argue that comprehension activities involving dictionary use or classroom discussion likely draw learners' attention to target vocabulary. However, we considered learning through such activities as incidental when learners' word engagement attempts were optional and spontaneous rather than forced by teachers or researchers.

5 One of the validity issues with the Vocabulary Knowledge Scale lies in the first two items in a checklist format: (a) "I don't remember having seen this word before," and (b) "I have seen this word before, but I don't know what it means." The fact that these two questions rely entirely on learners' self-report increases the likelihood of their making wild guesses. To control for the overestimation of word knowledge, Vidal (2003, 2011) and Yang and Sun (2013) included nonwords in the pool of test items, and they used the number of incorrectly identified words (i.e., false alarm) to adjust the final scores downwards.

6 We did not include the primary school category in the analysis because no studies focusing on primary school students reported test scores indicating basic vocabulary as we defined it in our study.

References

The full reference list of the studies included in the meta-analysis is available in Appendix S5 in the Supporting Information online.

Al-Shehri, M. A. A. (2015). *Incidental EFL vocabulary learning from listening to stories: The effects of frequency of exposure, elaboration and learner factors*. (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.

Bisson, M. J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*, 855–877. https://doi.org/10.1111/lang.12085

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*, 348–393. https://doi.org/10.1111/lang.12224

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.

Bruton, A., López, M. G., & Mesa, R. E. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly*, *45*, 759–768. https://doi.org/10.5054/tq.2011.268061

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

Comprehensive Meta-Analysis (Version 3.3). [Computer software]. Englewood, NJ: Biostat.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Daskalovska, N. (2010). Acquiring vocabulary through reading: A replication of the Golden Fleece study. In D. Köksal, I. H. Erten, E. Z. Topkaya, & A. Yavuz, A. (Eds.), *Current trends in ESL research and language teaching* (pp. 314–319). Selcuk-Izmir, Turkey: Tübitak.

Daskalovska, N. (2014a). Incidental vocabulary acquisition from reading an authentic text. *The Reading Matrix*, *14*, 201–216. http://eprints.ugd.edu.mk/11570

Daskalovska, N. (2014b). Incidental vocabulary learning through reading. *Journal of Linguistics and Language Teaching*, *5*, 57–65. http://eprints.ugd.edu.mk/11106

Daskalovska, N. (2016). Acquisition of three word knowledge aspects through reading. *Journal of Educational Research*, *109*, 68–80. https://doi.org/10.1080/00220671.2014.918530

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*, 906–941. https://doi.org/10.1111/lang.12296

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*, 443–477. https://doi.org/10.1075/ijcl.19.4.01dur

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*, 341–366. https://doi.org/10.1017/S0272263117000109

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. https://doi.org/10.1111/lang.12052

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*, 174–187. https://doi.org/10.2307/747863

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

Ferris, D. (1988). *Reading and second language vocabulary acquisition*. (Unpublished manuscript), University of Southern California, Los Angeles, CA.

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*, 319–333. https://doi.org/10.2307/329439

Godfroid, A., Ahn, J., Cho, I., Ballard, L., Cui, Y., Johnson, S., . . . Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*, 563–584. https://doi.org/10.1017/S1366728917000219

Hatami, S. (2017). The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge. *Reading in a Foreign Language*, *29*, 61–85.

Horst, M. (2000). *Text encounters of the frequent kind: Learning L2 vocabulary through reading*. (Unpublished doctoral dissertation). University of Wales, Swansea, UK.

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.

Hulme, R. C., Barsky, D., & Rodd, J. M. (2018). Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures. *Language Learning*. Published online 27 August 2018. https://doi.org/10.1111/lang.12313

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford, UK: Blackwell.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*, 327–339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558. https://doi.org/10.1111/0023-8333.00164

In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, *44*, 169–184. https://doi.org/10.5054/tq.2010.215253

Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly*, *24*, 215–235. https://doi.org/10.2307/747865

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*, 160–212. https://doi.org/10.1111/lang.12034

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, *73*, 440–464. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*, 377–392. https://doi.org/10.1080/15434303.2016.1237516

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, *65*, 127–159. https://doi.org/10.1111/lang.12115

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, *15*, 391–411. https://doi.org/10.1177/1362168811412019

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*, 687–711. https://doi.org/10.1177/1362168814559797

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*, 801–842. https://doi.org/10.1017/S027226311500042X

Malone, J. (2018). Incidental vocabulary learning in SLA. *Studies in Second Language Acquisition*, *40*, 651–675. https://doi.org/10.1017/S0272263117000341

Mohamed, A. A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*, 269–293. https://doi.org/10.1017/S0272263117000092

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition*, *37*, 677–711. https://doi.org/10.1017/S0272263114000825

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

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

Neuman, S. B., & Koskinen, P. (1992). Captioned television as "comprehensible input": Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*, 95–106. https://doi.org/10.2307/747835

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Second Language Acquisition*, *21*, 195–224. https://doi.org/10.1017/S027226319900203X

Pavia, N., Webb, S., & Faez, F. (in press). Incidental vocabulary learning from listening to L2 songs. *Studies in Second Language Acquisition*.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, *38*, 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*, 381–402. https://doi.org/10.1177/1362168815618428

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do “Things Fall Apart”? *Reading in a Foreign Language*, *22*, 31–55.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*, 113–151. https://doi.org/10.1111/j.1467-9922.2009.00502.x

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/S0272263117000407

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*, 1–28. https://doi.org/10.1613/jair.301

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106–178). New York, NY: Routledge.

Reynolds, B. L. (2018). The effects of nonce words, frequency, contextual richness, and L2 vocabulary knowledge on the incidental acquisition of vocabulary through reading: More than a replication of Zahar et al. (2001) & Tekmen and Daloğlu (2006). *International Review of Applied Linguistics*. Published online 22 June 2018. https://doi.org/10.1515/iral-2015-0115

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished doctoral dissertation). Victoria University, Wellington, NZ.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners’ incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, *21*, 589–619. https://doi.org/10.1017/S0272263199004039

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*, 72–78. https://doi.org/10.1016/0346-251X(78)90027-1

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. https://doi.org/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. https://doi.org/10.1177/026553220101800103

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*, 261–285. https://doi.org/10.3102/00346543069003261

Swanborn, M. S. L., & De Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, *52*, 95–117. https://doi.org/10.1111/1467-9922.00178

Tekmen, E. A. F., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, *39*, 220–243. https://doi.org/10.1111/j.1944-9720.2006.tb02263.x

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: Wiley.

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

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*, 56–89. https://doi.org/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. https://doi.org/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. https://doi.org/10.1177/003368828501600214

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

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

Webb, S. (2014). Repetition in incidental vocabulary learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781405198431.wbeal1425

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*, 667–686. https://doi.org/10.1177/1362168814559800

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University Press.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL–International Journal of Applied Linguistics*, *168*, 33–69. https://doi.org/10.1075/itl.168.1.02web

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*, 13–40.

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, *35*, 182–191. https://doi.org/10.1016/j.system.2006.12.009

Yang, H. C., & Sun, Y. C. (2013). It is more than knowledge seeking: Examining the effects of OpenCourseWare lectures on vocabulary acquisition in English as a foreign language (EFL) context. *Computer Assisted Language Learning*, *26*, 1–20. https://doi.org/10.1080/09588221.2011.624523

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*, 541–572. https://doi.org/10.3138/cmlr.57.4.541

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Sensitivity Analysis.
Appendix S2. Coding Scheme.
Appendix S3. Results of Publication Bias Measures and Funnel Plot.
Appendix S4. Relationship Between Frequency and Basic Vocabulary for 2,000 Frequency Level Mastered and Unmastered Samples.
Appendix S5. References of Included Studies.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

How Useful Is Repeated Exposure to Words for Vocabulary Learning?

What This Research Was About and Why It Is Important

Research on the learning of second language words through exposure to reading or listening materials (that is, without explicit teaching of the words) suggests that the number of encounters with words affects the likelihood of them being learned. However, the number of encounters needed for successful learning varies considerably, and estimations have ranged from between six to twenty or more. This inconsistency might be attributed to different learner characteristics (e.g., learners' proficiency) and methodological differences between studies (e.g., how words were presented—in back-to-back repetitions or interspersed among other words). Thus, it is unclear to what extent the number of encounters with words really is important for vocabulary learning while reading or listening. This study is a meta-analysis on this topic. That is, the researchers brought together the findings of 26 prior research studies to clarify the relationship between repeated exposure to words and second language vocabulary learning.

What the Researchers Did

- Following a thorough search of published studies on the learning of second language vocabulary from reading and/or listening, the researchers identified 26 studies that fit their selection criteria.
- The researchers analyzed 45 separate findings across the selected 26 studies, computing a mean finding for the relationship between the number of word repetitions and learners' knowledge of these words.
- The researchers also explored how a number of variables influenced this relationship: learner variables (e.g., learners' age, learners' prior vocabulary knowledge), experimental/learning variables (e.g., the availability of visual support during listening, learners' engagement), and methodological variables (e.g., whether or not learners are forewarned of an upcoming comprehension test, type of task by which vocabulary knowledge is tested).

What the Researchers Found

- The researchers found a medium-strength relationship between the number of exposures to a word and learners' knowledge of these words, called "the repetition effect."
- The researchers showed that this repetition effect was influenced by factors related to: learners (e.g., age, prior vocabulary knowledge), experimental/learning variables (e.g., visual support, engagement with L2 words), and methodological differences (forewarning of an upcoming test, vocabulary test format).
- The repetition effect appeared to be especially prominent at the earliest stage of vocabulary learning, for learners with less prior vocabulary knowledge and also for learning absolutely new words.
- The repetition effect appeared to be sensitive to contextual and pedagogical factors, such as the use of visual support (e.g., pictures) and learners' engagement with words, such as occasional dictionary use. These factors reduced the effects of increasing the number of repeated exposures to words. That is, these pedagogical factors resulted in fewer repetitions being needed for successful learning.

Things to Consider

- This synthesis only included studies that had tested learners' knowledge of the spelling and meaning of words. However, knowing a word involves other dimensions, such as its pronunciation and grammatical function, and so it is still unclear how repeated exposures are useful for learning these other dimensions of a word.
- Simply increasing the number of encounters with words does not always seem to help vocabulary learning in all circumstances. That is, a general, overarching principle of "the more repeated exposures, the better" does not always seem to hold true for learning the spelling and meaning of vocabulary while reading or listening in a second language.

How to cite this summary: Uchihara, T., Webb, S., & Yanagisawa, A. (2019). How useful is repeated exposure to words for vocabulary learning? *OASIS Summary* of Uchihara et al. in *Language Learning*. https://oasis-database.org