CAMBRIDGE
UNIVERSITY PRESS

STUDY

How effective is second language incidental vocabulary learning? A meta-analysis

Stuart Webb[1]* iD, Takumi Uchihara[2] and Akifumi Yanagisawa[3]

[1]University of Western Ontario, Ontario, Canada, [2]Waseda University, Tokyo, Japan and [3]University of Tsukuba, Ibaraki, Japan
*Corresponding author. Email: swebb27@uwo.ca

Abstract

There is a great deal of variation in gains found between studies of second language (L2) incidental vocabulary learning, as well as many factors that affect learning. This meta-analysis investigated the effects of exposure to L2 meaning-focused input on incidental vocabulary learning with an aim to clarify the proportional gains that occur through meaning-focused learning. Twenty-four primary studies were retrieved providing 29 different effect sizes and a total sample size of 2,771 participants (1,517 in experimental groups vs. 1,254 in control groups). Results showed large overall effects for incidental vocabulary learning on first and follow-up posttests. Mean proportions of target words learned ranged from 9–18% on immediate posttests, and 6–17% on delayed posttests. Incidental L2 vocabulary learning gains were similar across reading (17%, 15%), listening (15%, 13%), and reading while listening (13%, 17%) conditions on immediate and delayed posttest. In contrast, the proportion of words learned in viewing conditions on immediate posttests was smaller (7%, 5%). Findings also revealed that the amount of incidental learning varies according to a range of moderator variables including learner characteristics (L2 proficiency, institutional levels), materials (text type and audience), learning activities (spacing, mode of input), and methodological features (approaches to controlling prior word knowledge).

1. Introduction

There have been many studies investigating the extent to which words might be learned through exposure to second language (L2) input. Initially, most studies of L2 incidental vocabulary learning focused on reading (e.g., Cho & Krashen, 1994; Horst, 2005; Pitts et al., 1989; Zahar et al., 2001). However, in recent years there has been an increasing number of studies investigating L2 vocabulary learning through listening to aural input and viewing audiovisual input (e.g., Peters & Webb, 2018; Van Zeeland & Schmitt, 2013a). The degree to which words are learned incidentally has varied considerably across studies. For example, studies of incidental vocabulary learning through reading revealed gains of less than 10% (e.g., Pitts et al., 1989; Zahar et al., 2001) and more than 40% (e.g., Cho & Krashen, 1994; Horst, 2005). Studies of incidental vocabulary learning through listening have reported gains from 3.29% (Pavia et al., 2019) to 29% (Van Zeeland & Schmitt, 2013a), while studies of incidental vocabulary learning through viewing television have reported gains as low as 8% (Peters & Webb, 2018) and as high as 30% (Rodgers & Webb, 2020).

The variation in gains across studies of incidental vocabulary learning is likely owing to the many differences in participant, methodological, and treatment variables. For example, participants in studies include primary (e.g., Pavia et al., 2019), secondary (e.g., Szudarski & Carter, 2016), and university students (e.g., Rott, 1999), with varying proficiency levels. Treatments involve learning from expository

(e.g., Vidal, 2011) and narrative (e.g., Day et al., 1991) text through different modes of input such as reading (e.g., Day et al., 1991), listening (Van Zeeland & Schmitt, 2013a), reading while listening (e.g., Webb & Chang, 2015a), and viewing (e.g., Peters & Webb, 2018) in a single session (e.g., Peters & Webb, 2018), as well as through exposure to the input across multiple study sessions (e.g., Rodgers & Webb, 2020). In addition, prior knowledge of target items has been controlled across studies through the use of pseudowords as target items (e.g., Webb, 2007), pretests (e.g., Vidal, 2011), and pilot studies (e.g., Miyasako, 2002), while gains in vocabulary knowledge have been assessed through different test formats including form recognition (e.g., Peters & Webb, 2018), meaning recognition (e.g., Day et al., 1991), and meaning recall (Rott, 1999).

The present study aimed at examining the overall effects of meaning-focused learning on incidental L2 vocabulary. A meta-analysis of 24 primary studies that provided 29 different effect sizes and a total sample size of 2,771 participants was used to determine overall rates of incidental vocabulary learning through reading, listening, reading while listening, and viewing. A secondary aim was to investigate which variables moderate the effects of meaning-focused learning on incidental vocabulary learning. The research should help to clarify the proportional gains that occur through meaning-focused learning, as well as the degree to which these gains vary across different modes of input. The results should also help to guide teachers and learners to optimize vocabulary learning by identifying variables that affect incidental vocabulary gains through meaning-focused learning.

2. Defining incidental vocabulary learning

Incidental learning may commonly be perceived to be learning without intention. This may be owing in part to the fact that incidental learning is frequently contrasted with intentional learning (e.g., Hulstijn, 2001; Laufer, 2003; Webb & Nation, 2017). Defining incidental learning as learning without intention in research is problematic, however, because intention to learn is likely to vary among learners as well as within an individual from moment to moment (Webb, 2020). For example, during meaning-focused learning our attention may be oriented solely towards comprehension or move between understanding and learning unfamiliar language features when they are of interest or necessary for comprehension. Therefore, in research, incidental learning is commonly defined as either: (a) learning as a byproduct of a meaning-focused task (e.g., Chen & Truscott, 2010; Ellis, 1999), or (b) learning without knowledge of a forthcoming test (e.g., Hulstijn, 2001).

3. Incidental vocabulary learning gains

There have been many studies demonstrating that L2 words can be learned incidentally through reading (e.g., Day et al., 1991; Pitts et al., 1989; Waring & Takaki, 2003). However, research also reveals that vocabulary is learned incidentally through listening (e.g., Jin & Webb, 2020; Pavia et al., 2019; Van Zeeland & Schmitt, 2013a) and viewing (e.g., Montero Perez et al., 2014; Peters & Webb, 2018). Incidental vocabulary learning gains are perceived to be small in comparison with intentional vocabulary learning gains leading to suggestions that intentional learning is more effective (e.g., Laufer, 2003; Nation, 2013). Laufer's (2003) study provides some support for this as she found that three word-focused tasks contributed to greater gains in vocabulary knowledge than reading.

Studies of incidental vocabulary learning through reading have reported the following gains: 6.5%–8.6% (Pitts et al., 1989), 7.2% (Zahar et al., 2001), 18%–42% (Waring & Takaki, 2003), 28%–63% (Webb & Chang, 2015a), 43%–80% (Cho & Krashen, 1994), 44% (Webb & Chang, 2015b), 51% (Horst, 2005), and 65% (Pigada & Schmitt, 2006). There is similar variation in gains in studies of incidental word learning from listening with results ranging from 3.29%–8.67% (Pavia et al., 2019), 2%–29% (Brown et al., 2008), 7.8%–15.5% (Vidal, 2003), and 19%–29% (Van Zeeland & Schmitt, 2013a). There are relatively few studies of incidental word learning through viewing, and gains have been inconsistent: 8%–14% (Peters & Webb, 2018), 16–25% (Feng & Webb, 2020), and 23%–30% (Rodgers & Webb, 2020). The great deal of variation in the size of gains in studies of incidental vocabulary learning is likely owing to the many differences in participant, methodological, and treatment variables between studies.

4. Variables that may affect incidental vocabulary learning

4.1 L2 proficiency

Research indicates that learners who know more L2 words are likely to learn more words incidentally through reading (e.g., Webb & Chang, 2015a; Zahar et al., 2001) and watching television (Feng & Webb, 2020; Peters & Webb, 2018). This may occur because students who know more words have better reading (Laufer, 1989; Schmitt et al., 2011) and listening comprehension (Van Zeeland & Schmitt, 2013b), which may allow them to attend more to unknown language features that are encountered. Research indicates that the degree to which learners attend to unknown L2 words during reading is positively related to learning those words (Godfroid et al., 2018; Pellicer-Sánchez, 2016).

4.2 Institutional level

Although research tends to provide details indicating participants relative age, age itself, and place of study (primary or secondary school, university), age has not been explicitly examined in intervention studies as a variable. However, two recent meta-analyses indicate that older learners may make greater incidental vocabulary learning gains than younger learners. De Vos et al.'s (2018) meta-analysis of incidental vocabulary learning through encountering spoken input revealed that age had a positive effect on learning; participants who were university students made greater gains in vocabulary knowledge than children in elementary school and kindergarten. Uchihara et al.'s (2019) meta-analysis provides some justification for this result. They found that older learners were better able to make use of frequency effects to learn words incidentally than younger learners.

4.3 Text type

Gardner's (2004) corpus driven analysis of narrative and expository text written for children indicated several differences that may affect incidental vocabulary learning. Narrative text was found to include a higher proportion of high frequency words than expository text, while expository text consisted of a much larger number of different word types than narrative text. Gardner concluded that narrative text provided better conditions for incidental vocabulary learning than expository text. Because text type in studies of incidental vocabulary learning consist of both narrative (e.g., Brown et al., 2008; Rodgers & Webb, 2020) and expository text (e.g., Peters & Webb, 2018; Vidal, 2011), there is value in investigating the degree to which it moderates learning.

4.4 Text audience

The materials in studies of incidental vocabulary learning may be primarily targeted towards first language (L1) or L2 learners. Materials created for L1 learners include novels (e.g., Pellicer-Sánchez & Schmitt, 2010), songs (Pavia et al., 2019), and television programs (e.g., Peters & Webb, 2018; Rodgers & Webb, 2020). Materials aimed at L2 learners include graded readers (e.g., Horst, 2005; Waring & Takaki, 2003), teacher speech (Jin & Webb, 2020), and textbook passages (e.g., Teng & Reynolds, 2019). Materials such as novels and television shows created for native speakers of English are typically designed to inform and entertain. In contrast, materials aimed at L2 learners may be designed to promote the learning of language features that are encountered, as well as inform and entertain. Thus, it is useful to examine text audience as a variable to determine whether materials created for L2 learners enhance vocabulary gains.

4.5 Spacing

Incidental vocabulary learning is hypothesized to occur through repeated encounters with unknown words where knowledge is gained in small increments over time (Nagy et al., 1985; Webb & Nation, 2017). This suggests that the interval between encounters with unknown words may affect the degree to which they are learned. Most studies of incidental vocabulary learning have had very little spacing between encounters with all learning occurring through encounters in a single text in a single

learning session (e.g., Peters & Webb, 2018; Waring & Takaki, 2003). However, more recently there have been longitudinal studies of incidental learning over many sessions (e.g., Horst, 2005; Rodgers & Webb, 2020; Webb & Chang, 2015a, 2015b). Research on the spacing of encounters has revealed that spaced learning consistently leads to greater retention of vocabulary knowledge than massed learning (learning without an interval between repeated encounters) (e.g., Kim & Webb, 2022; Nakata, 2015).

4.6 Mode of input

Many studies have revealed that words can be learned incidentally through reading (e.g., Pigada & Schmitt, 2006; Waring & Takaki, 2003). An increasing number of studies have begun to show that words can also be learned incidentally through listening (e.g., Pavia et al., 2019; Van Zeeland & Schmitt, 2013b), reading while listening (e.g., Brown et al., 2008; Webb & Chang, 2012), and viewing (e.g., Peters & Webb, 2018; Rodgers & Webb, 2020). Several studies have compared learning across different modes of input. However, the results have been inconsistent. Brown et al. (2008) found that reading and reading while listening contributed to greater gains in vocabulary knowledge than listening. In contrast, Webb and Chang (2012) found that reading while listening led to greater gains in word learning than reading, while Feng and Webb (2020) found that reading, listening, and viewing all contributed to similar gains in vocabulary knowledge.

4.7 Control of prior knowledge of target words

Most studies of incidental vocabulary learning have used target single word items that were real words (e.g., Horst, 2005) and used pretests to control for prior knowledge of target items. However, pseudo-words (e.g., *cader*, *denent*) are also used as target items (e.g., Van Zeeland & Schmitt, 2013a; Waring & Takaki, 2003) to ensure that participants have no knowledge of their forms and allow researchers to forgo the inclusion of pretests of vocabulary knowledge (Nation & Webb, 2011). Justification for the use of pseudowords is thus methodological. However, Uchihara et al. (2019) found that the use of pseudoword target items contributed to larger frequency effects in incidental vocabulary learning studies. Thus, examining the effects of pre-knowledge control in studies of incidental vocabulary learning is warranted.

4.8 Test format

Research indicates that the degree to which learners are able to demonstrate their knowledge of words may depend on the test format that is used to measure learning. Laufer and Goldstein (2004) found that the form recall test format in which test takers must produce the L2 forms of words when presented with their meanings was the most demanding test followed by meaning recall tests (test takers are presented with the L2 forms of items and must produce their meanings), form recognition (test takers are presented with the meanings of target items and must select the correct L2 word from several choices), and meaning recognition (test takers are presented with L2 words and must select their corresponding meanings from among several choices), in that order. Because incidental vocabulary learning gains tend to be relatively small, it is important to use tests that are sensitive to knowledge, and so meaning recognition tests are perhaps used most frequently in research while form recall tests are used less often. Other test formats that are occasionally used are Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) and a second type of form recognition test that involves indicating whether target words were encountered in spoken or written text (e.g., Peters & Webb, 2018). Examining the degree to which test formats moderate incidental learning gains may shed light on the degree to which gains vary across test types.

5. The present study

The present meta-analysis investigated the effects of meaning-focused learning through reading, listening, reading while listening, and viewing on incidental vocabulary learning. The following research questions were posed:

1. What is the overall effectiveness of L2 incidental vocabulary learning through exposure to L2 meaning-focused input?
2. What is the overall rate of incidental vocabulary learning in: (a) measures of form recognition, meaning recognition, and meaning recall, and (b) through reading, listening, reading while listening, and viewing?
3. To what extent do variables related to learners, materials, comprehension activities, and methodological differences moderate L2 incidental vocabulary learning gains?

6. Method

6.1 Literature search

As suggested by literature search guidelines (In'nami & Koizumi, 2010; Plonsky & Oswald, 2015), we identified articles relevant to the current meta-analysis through searching the following databases: the Education Resources Information Center (ERIC), Linguistics and Language Behavior Abstracts (LLBA), ProQuest Dissertations and Theses, PsycINFO, Google, and Google Scholar. We also searched 19 journals in second language acquisition and language teaching and learning, where vocabulary studies are widely published, using the search function on each journal's website: *Annual Review of Applied Linguistics, Applied Linguistics, Applied Psycholinguistics, Canadian Modern Language Review, ELT Journal, Foreign Language Annals, International Journal of Applied Linguistics, International Review of Applied Linguistics in Language Teaching, Language Learning, Language Teaching, Language Teaching Research, Modern Language Journal, Reading in a Foreign Language, Reading Research Quarterly, RELC Journal, Second Language Research, Studies in Second Language Acquisition, System,* and *TESOL Quarterly*. Abstracts published up to October 2022 were targeted using various combinations of the following key words: "incidental (OR contextualized) vocabulary (OR word OR lexical) learning (OR acquisition)," AND "reading OR listening OR reading while listening OR viewing." A total of 2,853 reports, including both published or unpublished articles (i.e., dissertations and conference presentations), appeared initially qualified for the meta-analysis. The second author read each abstract to remove duplicates and papers other than empirical studies (e.g., review papers, meta-analysis studies, commentary). Full manuscripts for the remaining reports were retrieved and then screened according to the following selection criteria.

6.2 Inclusion criteria

Eight selection criteria were created to assess the identified studies:

1. The study measured incidental vocabulary learning gains from comprehension-based meaning-focused learning conditions in which target words were not subject to explicit manners of instruction. In other words, studies using language-focused activities such as word card learning (see Webb et al. (2020) for a meta-analysis of intentional vocabulary learning) were excluded (see Webb & Nation (2017) for a list of language-focused activities).
2. Adopting a methodological operationalization of incidental learning (Hulstijn, 2001; Uchihara et al., 2019), the current meta-analysis included studies in which the participating learners were not told of an upcoming vocabulary posttest subsequent to the treatment.
3. The study focused on vocabulary learning through meaning-focused input activities such as reading, listening, reading while listening, and viewing. This meta-analysis focused on studies in which the perceived goal of task completion among students was to comprehend the content of reading, listening, or viewing materials rather than to learn new vocabulary.
4. The study did not use glossing or captioned videos in the treatment condition. Meta-analytic reviews examining glossing and captioning effects on vocabulary learning are available elsewhere (see Yanagisawa et al. (2020) and Ramezanali et al. (2020) for glossed reading; Montero Perez et al. (2013) and Reynolds et al. (2022) for L1 and L2 subtitling).

5. The study ensured that prior knowledge of target words was little to none by using one of the following design features: using pseudowords, pilot testing target words, or administering pretests.[1]
6. The study concerned L2 vocabulary acquisition, not L1 acquisition.
7. The study adopted a between-participants design and included a control condition in which participants were not expected to be exposed to target words. The comparison between experimental and control conditions is essential in order to control for practice effects (i.e., an increase in learning gains owing to multiple exposures to the same target items through multiple test-taking opportunities).
8. The study reported effect sizes or sufficient information to calculate effect sizes (e.g., means, standard deviations, mean differences, and sample sizes).

It should be noted that we excluded within-comparisons studies that did not use control groups. Including pretest-posttest designs without control groups might involve testing effects on results, making it difficult to attribute the findings to the learning conditions. Moreover, to minimize publication bias—the fact that studies finding large effect sizes tend to be published—this meta-analysis included unpublished studies such as M.A. and Ph.D. theses and conference presentations.

Studies satisfying all eight criteria were included in the current meta-analysis. When additional information was necessary to complete the analysis, we contacted authors and received information from five researchers that allowed inclusion of their data in the meta-analysis (i.e., Ana Pellicer-Sánchez, Marije Michel, Nina Daskalovska, Niousha Pavia, and Yanxue Feng). In total, 24 studies were identified and submitted for the subsequent coding procedures (see Supplementary Appendix 1 for the included studies).

6.3 Coding

Table 1 displays the coding scheme specifying publication information, moderator variables related to learner, material, activity, and methodology, and dependent measures for both first and follow-up posttests. We first coded all 24 studies according to our coding scheme, which produced a total of 53 effect sizes. However, some of these effect sizes came from performance based on the same participants in experimental and/or control groups, causing the issue of multiple effect sizes—that is, meta-analysis based on multiple effect sizes from the same participants violates independence of observations (Plonsky & Oswald, 2015). To address this problem, we averaged multiple effect sizes prior to the meta-analysis (In'nami et al., 2020). This averaging method enables the meta-analytic outcomes to be comprehensive without any unnecessary loss of valuable data. As a result, for the first posttest, 29 independent effect sizes with 2,771 participants (1,254 in control and 1,517 in treatment groups) were selected and meta-analyzed in the current study. For the follow-up (delayed) posttest (the average retention interval = 34.1 days), nine out of the 29 effect sizes ($N = 741$; 325 in control and 416 in treatment groups) were available and included for analysis to compare the data with the results based on the first posttest scores. The completed coding sheet and raw data files are available via OSF at https://osf.io/92r3t/.

In order to establish the reliability of the coding procedures, all 24 studies were coded independently by the two researchers whose expertise lies in L2 vocabulary studies. Following Boulton and Cobb's (2017) approach, the number of discrepancies between the two researchers' coding was counted, and the agreement was rated at 98%. All disagreements and ambiguities were resolved through discussion.

6.4 Coding of moderator variables

Eight moderator variables categorized into three groups (learner, material and activity, and methodology) were coded in reference to the following criteria.

6.4.1 L2 proficiency

Following Jeon and Yamashita's (2014) suggestion and recent meta-analyses on L2 vocabulary learning (Webb et al., 2020), this study adopted criteria defining L2 proficiency dichotomously (Basic or

Table 1. Coding scheme

Variables	Values			
Publication information				
Authors				
Year				
Source	Journal	M.A./Ph.D. thesis	Conference presentation	Book/book chapter
Learner variables				
Sample size				
Participants' L1				
Target language				
L2 proficiency	Basic	Beyond basic		
Institutional level	Primary	Secondary	University	
Material and activity features				
Text type	Narrative	Expository		
Text audience	L1 users	L2 learners		
Spacing	Spaced	Massed		
Mode of input	Reading	Listening	Reading while listening	Viewing
Methodological features				
Pre-knowledge control	No pretest	Pseudoword use	Pilot testing	Pretest
Test format	Form recognition	Meaning recognition	Meaning recall	Other formats
Dependent measures				
Pretest				
$M_{treatment}$				
$SD_{treatment}$				
$M_{control}$				
$SD_{control}$				
1st posttest				
$M_{treatment}$				
$SD_{treatment}$				
$M_{control}$				
$SD_{control}$				
Follow-up posttest				
$M_{treatment}$				
$SD_{treatment}$				
$M_{control}$				
$SD_{control}$				

Note: M = mean; SD = standard deviation.

Beyond basic) in order to avoid inconsistency of proficiency judgements owing to the considerable variation of researchers' reporting methods of L2 proficiency such as vocabulary test scores, standardized proficiency test scores (e.g., TOEFL), and teachers' intuitive judgements.

6.4.2 Institutional level
This variable was coded as primary school, secondary school, or university. Other types of institutions such as language school were not found in the current data set.

6.4.3 Text type
Following research suggesting differences in characteristics between narrative and academic texts (Gardner, 2004), we categorized the type of texts used for learning materials in two ways: narrative or expository texts.

6.4.4 Text audience
This variable was coded in terms of whether the text was created for L1 users or L2 learners. Texts made for L2 learners include graded readers, texts written by authors, and modified texts. Texts without any modification or with a minor modification by embedding target words or replacing real words with pseudowords were considered to be written for L1 speakers.

6.4.5 Spacing
The spacing effect occurs when the same target words are encountered multiple times with a certain gap between the encounters. Following Uchihara et al.'s (2019) definition of spaced and massed learning conditions, studies in which participants could encounter target words repeatedly over several treatment sessions were identified as spaced-treatment conditions. The current definition of spacing applies to the following scenarios: (a) participants are asked to complete certain tasks (e.g., reading a book) on their own time over an extended period of time (e.g., Daskalovska, 2016) and (b) participants are exposed to multiple texts in a series of classroom sessions (e.g., Webb & Chang, 2015a, 2015b). In contrast, studies in which exposure to L2 input was limited to a single-day/one-time treatment were coded as massed-treatment conditions.

6.4.6 Mode of input
This variable consists of four categories: reading, listening, reading while listening, and viewing. For viewing studies, conditions with L2 captions or L1 subtitles were not included in this category.

6.4.7 Pre-knowledge control
It is important to control for prior knowledge of target words in incidental vocabulary learning research to ensure that any gains can be attributed to the learning conditions. To examine this variable, we coded pre-knowledge control in three ways: using pseudowords, pilot testing with other learners of similar L1 and educational backgrounds, or administering pretests, and compared them with studies not administering pretests nor using pseudowords.

6.4.8 Test format
We initially coded test format in six ways: form recognition (e.g., lexical decision task, multiple choice task), meaning recognition (e.g., word matching task), form recall (e.g., L1 to L2 translation), meaning recall (e.g., L2 to L1 translation task), developmental scale (i.e., a vocabulary knowledge scale or VKS), and productive knowledge of spelling (i.e., dictation task). After coding, form recall ($k = 0$), developmental scales ($k = 2$) and productive knowledge of spelling ($k = 2$) were removed owing to an insufficient number of studies using these measures.

6.5 Effect size calculation

The first and follow-up posttest effect sizes were calculated according to Hedges' *g* (Hedges & Olkin, 1985), which offers slightly more conservative calculations compared with Cohen's *d*, especially for small samples (Borenstein et al., 2009). First, the standardized mean difference or *d* was calculated using the means for treatment and control groups and pooled standard deviations (*SD*s) according to the following formulae:

$$d = \frac{Mean_1 - Mean_2}{Pooled\ SD}$$

$$Pooled\ SD = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Following Morris's (2008) suggestion, we calculated pooled *SD*s using *SD*s based on pretest scores when available. Lastly, we used a bias correction factor or J, which was calculated according to the formulae: 1–[3/{4 × *df*–1}] (Hedges & Olkin, 1985) in order to obtain an unbiased effect size, Hedges' *g* (=*d* × *J*). In order to interpret effect size values for this study, we referred to Plonsky and Oswald's (2014) benchmarks defining the magnitude of effect sizes: small = 0.40, medium = 0.70, and large = 1.00.

6.6 Data analysis

The current study used the Comprehensive Meta-Analysis (version 3.3) (Borenstein et al., 2006) software to calculate the mean effect size and conduct moderator analysis. Prior to the effect size aggregation and moderator analysis, we conducted preliminary inspection of the data, revealing that one study (Vidal, 2011) was an outlier in first and second posttests (i.e., more than three *SD*s above the mean effect sizes); therefore, the study was excluded before proceeding to the subsequent statistical analysis. In addition, we conducted two measures to assess the extent to which publication bias influences the current data sets: fail-safe *N* and the trim-and-fill method based on the examination of funnel plots (Borenstein et al., 2009). These two measures indicated that there was little concern regarding the effect of publication bias on the current meta-analysis findings (see Supplementary Appendix 2 for detailed information regarding the results of the two measures).

A random-effects model was employed to compute the inverse-variance weighted mean effect for the first posttests as well as follow-up posttests, along with a mixed-effects model for subsequent moderator analysis. In effect size aggregation, the homogeneity test was conducted using a within-group *Q* statistic for the purpose of examining whether there would be a significant variation in true effect sizes across studies. For moderator analysis, between-group *Q* value was calculated for a total of eight categorical variables. In response to the second research question, we followed previous meta-analysis on L1 incidental vocabulary learning (Swanborn & De Glopper, 1999) to determine rates of learning by calculating the proportion of learned target items in relation to the total number of target words. To this end, we first revisited the full version of the completed coding sheet and sorted all effect sizes according to test task format (*k* = 38: form recognition, meaning recognition, and meaning recall) and mode of input (*k* = 34: reading, listening, reading while listening, and viewing). Next, the mean differences (i.e., learning gain scores in experimental groups – learning gain scores in control groups) were calculated per test format and used to provide an indication of learning gains MINUS practice effect. Finally, the resulting gain scores were divided by the total number (or maximum scores) of target words to produce rates of learning, which were then averaged for each of the three test formats and four modes of input.

7. Results

7.1 Overall effect of incidental vocabulary learning

In order to examine the overall effectiveness of incidental L2 vocabulary learning through different modes of input, the mean effect sizes along with 95% confidence intervals (CIs) were computed for

the first ($k = 28$) and follow-up ($k = 9$) posttests. The mean effect sizes for the first and follow-up posttests were large (Plonsky & Oswald, 2014), $g = 1.14$, 95% CI [0.86, 1.41], $p < .001$ and $g = 0.93$, 95% CI [0.44, 1.42], $p < .001$, respectively. The homogeneity test was statistically significant for both the first and follow-up posttests, $Q = 269.08$, $p < .001$ and $Q = 55.85$, $p < .001$, indicating the possibility that moderator variables exist accounting for the difference in the magnitude of incidental learning across studies.

7.2 Rate of learning

Table 2 shows that for knowledge of form recognition, approximately 18% of the target words were learned on the first posttest, and average retention of form recognition declined very sharply to 6%. The pick-up rates for meaning recognition were 15% on the first posttest and 17% on the second posttest. The rate of learning for meaning recall was the lowest (9%) of the three test formats. Yet, meaning recall appeared to be durable as the rate of learning on the follow-up posttests was slightly higher (12%) than the initial learning rate.

Table 3 presents the learning rates for the four modes of input. Learning rates were highest for reading (17%), followed by listening (15%), reading while listening (13%), and viewing (7%). For the second posttest, it is interesting to note that reading while listening (17%) appeared to help learners retain or enhance word knowledge more than reading (15%) or listening (13%). Viewing appears to be least effective among all modes of input (5%).

7.3 Moderator analysis (first posttest)

The results of moderator analyses explore the extent to which the selected moderator variables could account for learning variability across studies. Summaries of results for eight categorical variables for the first and follow-up posttests are presented in Tables 4 and 5, respectively.

7.3.1 Learner variables (proficiency, institutional level)

For L2 proficiency, the effect size was larger for Beyond basic learners ($g = 1.40$) than for Basic learners ($g = 0.70$) and the difference was approaching significance ($p = .051$). This indicates that more proficient learners are more likely to learn L2 words incidentally than less proficient learners. Regarding institutional level, although not statistically significant, a sizable variation in the effectiveness of incidental vocabulary learning was found across the three groups, indicating that older individuals tended to learn more words than younger ones: university ($g = 1.36$), secondary ($g = 0.87$), and primary school ($g = 0.72$) students.

Table 2. Rate of learning for form recognition, meaning recognition, and meaning recall

				95% CI	
	k	M	SE	Lower	Upper
First posttest					
Form recognition	7	.181	.052	.056	.309
Meaning recognition	22	.151	.025	.098	.204
Meaning recall	9	.086	.038	−.001	.174
Follow-up posttest					
Form recognition	2	.064	.027	−.280	.409
Meaning recognition	7	.173	.056	.035	.311
Meaning recall	4	.116	.061	−.080	.313

Table 3. Rate of learning for mode of input: Reading, listening, reading while listening, and viewing

				95% CI	
	k	*M*	*SE*	Lower	Upper
First posttest					
Reading	17	.174	.034	.103	.245
Listening	6	.151	.039	.051	.250
Reading while listening	2	.128	.024	−.181	.438
Viewing	9	.071	.021	.024	.119
Follow-up posttest					
Reading	7	.151	.056	.013	.289
Listening	4	.133	.065	−.075	.340
Reading while listening	2	.170	.065	−.658	.998
Viewing	2	.051	.002	.027	.075

Note: k = number of studies.

7.3.2 Material and activity (text type, audience, spacing, mode)

Both text-related variables (text type, text audience) turned out to be significant moderators. Learners exposed to narrative texts ($g = 1.43$) picked up more L2 words compared with those exposed to expository texts ($g = 0.61$). Learners exposed to texts designed for L2 learners ($g = 1.56$) acquired more words than those exposed to texts designed for L1 users ($g = 0.71$). In addition, the effect size for learning in spaced conditions ($g = 1.51$) was larger than in massed conditions ($g = 0.97$), and the difference approached statistical significance ($p = .080$). With respect to mode of input, significant variation was found across the four modes with a large effect for reading ($g = 1.45$), medium to large effects for listening ($g = 0.97$) and reading while listening ($g = 0.78$), and a slightly lower effect for viewing ($g = 0.60$).

7.3.3 Methodology (pre-knowledge control, test format)

Significant variation was found across the three approaches to pre-knowledge control of target words revealing a noticeable pattern with studies using pseudowords ($g = 1.90$) tending to find larger effects compared with studies either pilot testing target words ($g = 0.80$), administering pretests ($g = 1.01$), or no pretests ($g = 0.27$). There was no noticeable difference between test formats: form recognition (g = 1.04), meaning recognition (g = 1.10), and meaning recall (g = 1.08).

7.4 Moderator analysis (follow-up posttest)

Overall, the patterns of results for the follow-up posttests were similar to the first posttest results. However, fewer statistically significant differences and effect size values were found owing to smaller sample sizes. The results of three variables (spacing, mode of input, pre-knowledge control) are notable for their contrast to the first posttest results. First, the positive effect of spaced learning ($g = 1.71$) relative to massed learning ($g = 0.58$) became even more accentuated in the second posttest (first posttest: $Q = 3.06$, $p = .080$ vs. second posttest: $Q = 6.22$, $p = .013$). Regarding mode of input, the between-study variation in effect size became non-significant with the results showing a sharp decrease in effect size for reading from the first posttest ($g = 1.45$) to the second posttest ($g = 0.96$). In addition, a large effect size for reading while listening ($g = 1.07$), a medium to large effect size for listening ($g = 0.76$), and a small effect for viewing (g = 0.43) were found. However, the results for listening, reading while listening, and viewing should be interpreted cautiously owing to the small sample of studies with delayed posttests for these modes of input ($k = 2$, $k = 3$, $k = 2$ respectively). Lastly, it is notable that

Table 4. *Moderator analysis (first posttest)*

	k	g	95% CI		p	Q tests	p
			Lower	Upper		Q	
Learner variables							
L2 proficiency						3.81	.051
Basic	7	0.70	0.10	1.29	.022		
Beyond basic	18	1.40	1.02	1.77	.000		
Institutional level						3.82	.148
Primary school	5	0.72	0.07	1.37	.029		
Secondary school	6	0.87	0.29	1.46	.003		
University	17	1.36	1.00	1.71	.000		
Material and activity							
Text type						8.68**	.003
Narrative	18	1.43	1.10	1.76	.000		
Expository	10	0.61	0.17	1.04	.041		
Text audience						10.49**	.001
L1 users	14	0.71	0.35	1.07	.000		
L2 learners	14	1.56	1.19	1.93	.000		
Spacing						3.06	.080
Spaced condition	9	1.51	1.01	2.02	.000		
Massed condition	19	0.97	0.63	1.31	.000		
Mode of input						9.50*	.023
Reading	16	1.45	1.11	1.80	.000		
Listening	5	0.97	0.35	1.59	.002		
Reading while listening	2	0.78	−0.18	1.74	.110		
Viewing	9	0.60	0.15	1.05	.009		
Methodology							
Pre-knowledge control						13.34**	.004
No pretest	2	0.27	−0.66	1.19	.570		
Pseudoword use	7	1.90	1.38	2.42	.000		
Pilot testing	5	0.80	0.22	1.38	.007		
Pretest	14	1.01	0.65	1.38	.000		
Test format						0.04	.983
Form recognition	7	1.04	0.48	1.59	.000		
Meaning recognition	22	1.10	0.78	1.41	.000		
Meaning recall	9	1.08	0.59	1.58	.000		

Note: **indicates $p < .01$; *indicates $p < .05$.

Table 5. Moderator analysis (follow-up posttest)

	k	g	95% CI		p	Q tests	
			Lower	Upper		Q	p
Learner variables							
L2 proficiency						0.03	.869
Basic	3	0.87	−0.04	1.78	.060		
Beyond basic	6	0.97	0.33	1.60	.003		
Institutional level						0.65	.722
Primary	2	0.52	−0.67	1.71	.390		
Secondary	2	0.96	−0.21	2.12	.108		
University	5	1.11	0.35	1.84	.004		
Material and activity							
Text type						3.67	.056
Narrative	5	1.33	0.71	1.94	.000		
Expository	4	0.43	−0.26	1.11	.222		
Text audience						4.00*	.045
L1 users	4	0.44	−0.21	1.09	.189		
L2 learners	5	1.34	0.74	1.94	.000		
Spacing						6.22*	.013
Spaced condition	3	1.71	0.97	2.46	.000		
Massed condition	5	0.58	0.08	1.07	.022		
Mode of input						0.80	.849
Reading	6	0.94	0.29	1.58	.005		
Listening	3	0.76	−0.14	1.66	.098		
Reading while listening	2	1.07	−0.05	2.20	.061		
Viewing	2	0.43	−0.68	1.54	.448		
Methodology							
Pre-knowledge control						0.80	.372
Pilot testing	2	0.45	−0.67	1.57	.431		
Pretest	6	1.04	0.38	1.71	.002		
Test format						1.69	.430
Form recognition	2	0.24	−0.94	1.42	.695		
Meaning recognition	7	1.10	0.47	1.74	.001		
Meaning recall	4	1.06	0.23	1.89	.012		

Note: *indicates $p < .05$. Pseudoword use (under pre-knowledge control) was removed owing to the small number of effect sizes ($k = 1$).

the effect size for studies in which learners completed pretests appeared to be less subject to change between the first ($g = 1.01$) and second ($g = 1.04$) posttests, pointing to the possibility that practice effects occurred in studies with pretest/posttest designs.

8. Discussion

In answer to the first research question, the analysis revealed that L2 meaning-focused input contributed to large learning effects for vocabulary knowledge on both first and second posttests. This finding is important because it clarifies the value of encountering L2 meaning-focused input on vocabulary learning. Moreover, when considering this result, it is useful to also reflect on aspects of vocabulary learning that were not included in the meta-analysis, but may also occur through encountering L2 meaning-focused input. For example, in the research literature, incidental vocabulary learning is limited to gains in knowledge of whichever target words are tested. These words are typically the lowest frequency words that are least likely to be known by most participants. However, there are likely to be many non-target words that are unknown or partially known that may also potentially be learned in these studies. In addition, the value of encountering L2 input for vocabulary learning might primarily be that it reveals how words can be used. While there are a small number of studies (e.g., Chen & Truscott, 2010; Pigada & Schmitt, 2006; Webb, 2007, 2008) that have investigated the contribution of meaning-focused input on learning aspects of knowledge other than form-meaning connection (e.g., collocation, grammatical function, association), these studies do reveal that gains in vocabulary knowledge will typically extend beyond form-meaning connection. Thus, it is important to note that the studies included in this meta-analysis were unlikely to reveal the full extent of gains in L2 vocabulary knowledge through encountering meaning-focused input.

In answer to the second research question, the mean proportions of target words learned as indicated by the three test formats (form recognition, meaning recognition, form recall) ranged from 9–18% on the first posttest, and 6–17% on the second posttest. Higher mean proportions of target words learned for meaning recognition (15% -> 17%) and meaning recall (9% -> 12%) on delayed posttests suggest practice effects may often occur from immediate to delayed posttests. Researchers should try to control for practice effects by counterbalancing target items between immediate and delayed posttests and measuring knowledge of half of the target items in the immediate posttest and the remaining items in the delayed posttest. Teachers should also be aware of the positive effects and potential gains that can occur through assessing knowledge of target vocabulary encountered in L2 input.

The mean proportions of target words learned in the different modes of input on the first posttest were 7% for viewing, 13% for reading while listening, 15% for listening, and 17% for reading. On the second posttest, the mean proportions of target words learned were 15% for reading, 13% for listening, 17% for reading while listening, and 5% for viewing. The results suggest that the proportions of words learned through reading, listening, and reading while listening may be similar, but that all of these modes of input might lead to greater gains than viewing on immediate and delayed posttests. This may be owing to: (a) learners' familiarity with learning from L2 written and spoken input in the classroom in comparison to audiovisual input, (b) the ability for researchers to more easily manipulate the frequency of occurrence of target items in spoken and written input than in audiovisual input, and (c) the use of spoken and written learning materials designed for L2 learning in studies in comparison with audiovisual materials created for L1 viewers' entertainment, education, and interest.

The analyses indicate that incidental vocabulary learning gains are likely to be smaller than occur through intentional vocabulary learning activities. The findings of the present study suggest that less than 20% of target words are likely to be learned through encountering meaning-focused input on delayed posttests. In comparison, a meta-analyses of intentional vocabulary learning tasks revealed the mean proportions of target words retained was 39.4% on meaning recall delayed posttests (Webb et al., 2020). It is important to note, however, that the incidental vocabulary learning gains reported in the present meta-analysis accounted for possible learning effects by calculating the degree to which the gains of experimental groups exceeded those of control groups. The lack of control groups in L2 intentional vocabulary learning studies limited Webb et al. (2020) from carrying out a similar calculation. Thus, we urge readers to be cautious when interpreting the difference in proportional gains between the two studies.

The present study and Webb et al.'s (2020) meta-analyses of intentional learning activities reveal that both meaning-focused and intentional learning activities have large positive effects on vocabulary learning. However, both approaches have advantages and disadvantages, and neither is likely to contribute to comprehensive knowledge of words on their own over a short period of time. There is a great deal to learn about a word including its spelling, pronunciation, derivations, associations, meanings, collocations, grammatical functions, and constraints on use (Nation, 2013). Research indicates that learners make incremental gains in word knowledge and that developing comprehensive knowledge of words is a slow process in which gains and losses occur (Webb & Nation, 2017). Thus, the gains demonstrated through tests in any cross-sectional study likely represent a fraction of the vocabulary knowledge that could be gained about each word. Rather than suggesting that any one approach to learning is best, there is greater value in determining the degrees to which different types of vocabulary knowledge (e.g., form-meaning connection, collocation, word parts) are gained over different lengths of time though the different learning approaches.

In answer to the third research question, the results revealed that learner, material and activity, and methodological variables all moderated the incidental learning gains. Of the learner variables, the results showed that more proficient learners make larger gains than less proficient learners; however, the difference between the proficiency levels diminished on delayed posttests. Institutional level did not significantly moderate gains (although it should be noted that the size of gains showed a similar trend i.e., university ($g = 1.36$), secondary school ($g = 0.87$), and primary school ($g = 0.72$)). The reason why more proficient learners make larger gains is likely owing to better comprehension of L2 input allowing them to devote greater processing resources towards understanding unfamiliar language. Lower proficiency learners may have to devote greater processing resources towards understanding the input as a whole, thus making it more difficult to attend to unfamiliar words. The difference in findings between the two learner variables indicates the greater precision of proficiency tests as an indicator of level than institutional background. This should be expected because there is likely to be greater variation among learners within an institutional level than among those within a proficiency band.

Of the material and activity moderator variables, the analyses of immediate posttests revealed that incidental gains were larger for narrative texts than expository texts, texts oriented towards L2 learners rather than L1 learners, and through reading and reading while listening, followed by listening and then viewing. The greater effects of learning from narrative text than expository text may be because the former provides better conditions for word learning than the latter; narrative text consists of a higher proportion of higher frequency words and is less lexically dense than expository text (Gardner, 2004). Greater vocabulary gains occurring from exposure to materials created for L2 learners rather than those oriented towards L1 learners should be expected, because L2 materials are designed to be at the appropriate lexical and syntactic levels for L2 learners while L1 materials are not. Materials that are less linguistically demanding should allow learners to attend more to unfamiliar words and increase the potential for vocabulary learning. The results of the delayed posttests revealed significant positive effects of learning occurring over multiple sessions rather than in a single session, and through learning in each of the different modes of input. Greater incidental vocabulary learning gains through learning in multiple sessions on delayed posttests is supported by research on distributed practice, which suggests that spacing study sessions has a positive effect on learning particularly when there is an interval between learning and testing (Kim & Webb, 2022). This finding demonstrates the value of extensive learning programs that involve regular meaning-focused learning over longer periods of time.

The comparisons between the different modes of input varied on first and second posttests. On immediate posttests, reading contributed to the largest effect size, whereas on delayed posttests, reading while listening contributed to the largest effect size (although it should be noted that there were only two reading while listening studies that included delayed posttests). It might be expected that written materials contribute to greater gains than aural and audiovisual materials, because in the EFL context, participants are most likely to be familiar with learning with written texts, followed by

aural and audiovisual materials. However, the results tended to show medium to large effects for all modes (except a small effect for viewing on immediate and delayed posttests). Further incidental vocabulary learning studies of listening, reading while listening, and viewing would be useful as there were far fewer studies investigating these modes of input than for reading.

The moderator analyses also revealed greater effects on learning when pre-knowledge control involved pseudowords followed by pretests, pilot testing, and no pretests. There are several possible reasons for this result. First, pseudoword forms may be more salient than real word target items because the former have never been encountered before while the latter may be unfamiliar (i.e., partially known but not to the degree that participants can demonstrate knowledge on pretests) rather than completely unknown. Second, it may be that practice effects in which participants gain knowledge of target items through pretests reduce the extent to which gains are revealed in pretest-posttest designs. However, it is important to note that pretest control was found to be the most common form of pre-knowledge control, and it contributed to large effect sizes at both retention intervals. Thus, there appears to be strong support for the use of pretests for pre-knowledge control. Third, although pilot testing other learners with a similar profile to participants also revealed medium and small effect sizes on immediate and delayed posttests, these smaller effect sizes in relation to the other two pre-knowledge control options suggest that this is the least effective of the three approaches perhaps owing to variation in knowledge of pilot test and experimental participants.

The moderator analysis did not reveal a significant difference between the three test formats. Large effect sizes were found at both retention intervals for meaning recognition and meaning recall while large and small effects were found for form recognition on the immediate and delayed tests, respectively. The similar effect sizes between meaning recognition and recall might be considered surprising because recall formats are less sensitive to knowledge than recognition formats (e.g., Laufer & Goldstein, 2004; Webb, 2007), and may limit the degree to which incidental learning gains are found (Nagy et al., 1985). However, it may be that because L2 incidental learning studies tend to include optimal conditions for word learning (i.e., higher than normal frequencies of occurrence of target items), meaning recall tests have been similarly effective for revealing word learning. A second reason for the similarity in findings between these test formats may be that guessing on meaning recognition pretests limits the degree to which gains are found through exposure to input.

9. Limitations and future directions

It is important to note several limitations of the present study. First, research on incidental vocabulary learning tends to include careful control of learning conditions. This has methodological value because it increases the likelihood that findings can be attributed to the learning conditions. However, it also means that the results may not reflect how words are typically learned incidentally inside and outside of the classroom. For example, within studies, participants typically do not have access to the support that they would typically have within and outside of the classroom. For example, during meaning-focused learning, students may consult dictionaries, teachers, classmates, and parents to aid comprehension of L2 input. Thus, the results found in studies of incidental vocabulary learning studies may be less than those that occur in less controlled conditions. Second, incidental vocabulary learning studies reveal gains that occur for target items. These target items are typically low frequency words that are encountered more often than the other low frequency words within the L2 input. However, low frequency words that are not target items may also be learned to some degree. In addition, higher frequency words that are unknown or partially known may also be learned. Third, the meta-analysis examined gains in knowledge of form recognition and form-meaning connections of words. The degree to which other aspects of knowledge (i.e., spoken form, collocation, word parts, association) may also be gained through exposure to meaning-focused input was not explored. There are relatively few studies examining aspects of knowledge besides form-meaning connection and this is clearly an area where further research is needed.

The meta-analysis also revealed several other areas where research on incidental vocabulary learning is warranted. First, the results showed that there are few studies that include delayed posttests

($k = 9$) in comparison with immediate posttests ($k = 29$). We urge readers to be cautious when interpreting the moderator analyses on the delayed tests as further research is necessary to be confident about the generalizability of these findings. The lack of studies including delayed posttests may be owing in part to practice effects occurring from immediate posttest to delayed posttests (e.g., Webb et al., 2013). There are two options that might be considered to avoid this issue. First, target items could be counterbalanced with half assessed in immediate posttests and the remaining items assessed in delayed posttests. Another option would be to eliminate immediate posttests and assess learning only on delayed posttests. It would be essential to include a no-treatment control group in such a design to control for outside learning. A second area where further research is needed is incidental learning through listening, reading while listening, and viewing. The number of studies that examined immediate posttest gains made through reading ($k = 17$) was the same as the combined number of studies of viewing ($k = 9$), listening ($k = 6$), and reading while listening ($k = 2$). In addition, there is also a need for more studies investigating incidental learning with younger participants. There were relatively few studies with participants in primary ($k = 5$) and secondary school ($k = 6$) in comparison with university ($k = 17$).

10. Conclusion

The present study revealed that exposure to L2 meaning-focused input contributed to large learning effects for knowledge of form-meaning connections of unknown words. The mean proportions of target words learned ranged from 9–18% on immediate posttests, and 6–17% on delayed posttests. Incidental L2 vocabulary learning gains were similar across reading (17%, 15%), listening (15%, 13%), and reading while listening (13%, 17%) conditions on immediate and delayed posttests. In contrast, the proportion of words learned in viewing conditions on immediate posttests was much smaller (7%, 5%). Moderator analyses revealed that each of learner (proficiency), material and activity (text type, text audience, mode of input), and methodological (pre-knowledge control) variables affected the size of gains on immediate posttests. However, only material and activity (spacing, mode of input) variables moderated gains on delayed posttests. Difference in findings for moderator variables between retention intervals was likely owing in part to fewer studies including delayed posttests.

Note
[1] Koolstra and Beentjes (1999) conducted none of the control measures mentioned (pseudoword use, pretest, and pilot testing). However, this study conducted a preliminary examination into the potential confounding variable of prior word knowledge by administering vocabulary tests (e.g., part of the Peabody Picture Vocabulary Test) in order to mitigate the impact of the initial difference in vocabulary size across experimental groups on the study results.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2006). *Comprehensive meta-analysis (version 3.3.070) [computer software]*. Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009).*Introduction to meta-analysis*. Wiley. doi:10.1002/9780470743386

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. doi:10.1111/lang.12224

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening. *Reading in A Foreign Language*, 20(2), 136–163.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(3), 693–713. doi:10.1093/applin/amq031

Cho, K., & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662–667.

Daskalovska, N. (2016). Acquisition of three word knowledge aspects through reading. *Journal of Educational Research*, *109*(1), 68–80. doi:10.1080/00220671.2014.918530

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in A Foreign Language*, *7*(2), 541–551.

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*(4), 906–941. doi:10.1111/lang.12296

Ellis, R. (1999). *Learning a second language through interaction*. John Benjamins.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. doi:10.1017/S0272263119000494

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics*, *25*(1), 1–37. doi:10.1093/applin/25.1.1

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A. & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. doi:10.1017/S1366728917000219

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. doi:10.1016/C2009-0-03396-0

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, *61*(3), 355–382. doi:10.1353/cml.2005.0018

Hulstijn, J. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instructions* (pp. 258–286). Cambridge University Press. doi:10.1017/CBO9781139524780

In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, *44*(1), 169–184. doi:10.5054/tq.2010.215253

In'nami, Y., Koizumi, R., & Tomita, Y. (2020). Meta-analysis in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 240–252). Routledge. doi:10.4324/9780367824471-4

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. doi:10.1111/lang.12034

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning from listening to teacher talk. *Modern Language Journal*, *104*(3), 550–566. doi:10.1111/modl.12661

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, *72*(1), 269–319. doi:10.1111/lang.12479

Koolstra, C. M., & Beentjes, J. W. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, *47*(1), 51–60. doi:10.1007/BF02299476

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567–587. doi:10.3138/cmlr.59.4.567

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. doi:10.1111/j.0023-8333.2004.00260.x

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, *39*(1), 1–20. doi:10.24539/let.39.0_1

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118–141. http://dx.doi.org/10125/44357

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. doi:10.1177/1094428106291059

Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233–253. doi:10.2307/747758

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*(4), 677–711. doi:10.1017/S0272263114000825

Nation, I. S. P. (2013). *Learning vocabulary in another language*. (2nd ed.). Cambridge University Press. doi:10.1017/CBO9781139858656

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

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning from listening to L2 songs. *Studies in Second Language Acquisition*, *41*(4), 745–768. doi:10.1017/S0272263119000020

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, *38*(1), 97–130. doi:10.1017/S0272263115000224

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart?. *Reading in A Foreign Language*, *22*(1), 31–55.

Perez, M. M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. doi:10.1016/j.system.2013.07.013

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through watching a single episode of L2 television. *Studies in Second Language Acquisition*, *40*(3), 551–577. doi:10.1017/S0272263117000407

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

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in A Foreign Language*, *5*(2), 271–275.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. doi: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). Routledge. doi:10.4324/9781315870908

Ramezanali, N., Uchihara, T., & Faez, F. (2020). Efficacy of multimodal glossing on L2 vocabulary learning: A meta-analysis. *TESOL Quarterly*, *55*(1), 105–133. doi:10.1002/tesq.579

Reynolds, B. L., Cui, Y., Kao, C. W., & Thomas, N. (2022). Vocabulary acquisition through viewing captioned and subtitled video: A scoping review and meta-analysis. *Systems*, *10*(5), 133. doi:10.3390/systems10050133

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through watching television. *ITL - International Journal of Applied Linguistics*, *171*(2), 191–220. doi:10.1075/itl.18034.rod

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, *21*(4), 589–619. doi:10.1017/S0272263199004039

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Swanborn, M. S., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261–285. doi:10.3102/00346543069003261

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245–265. doi:10.1111/ijal.12092

Teng, F., & Reynolds, B. L. (2019). Effects of individual and group metacognitive prompts on EFL reading comprehension and incidental vocabulary learning. *PLoS One*, *14*(5), 1–24. doi:10.1371/journal.pone.0215902

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599. doi:10.1111/lang.12343

Van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. doi:10.1016/j.system.2013.07.012

Van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. doi:10.1093/applin/ams074

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*(1), 56–89. doi:10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258. doi:10.1111/j.1467-9922.2010.00593.x

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in A Foreign Language*, *15*(2), 130–163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. doi:10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in A Foreign Language*, *20*(2), 232–245.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge. doi:10.4324/9780429291586

Webb, S., & Chang, A. C.-S. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651–675. doi:10.1017/S0272263114000606

Webb, S., & Chang, A. C.-S. (2015b). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686. doi:10.1177/1362168814559800

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

Webb, S., Newton, J., & Chang, A. C.-S. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120. doi:10.1111/j.1467-9922.2012.00729.x

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary learning activities? A meta-analysis. *Modern Language Journal*, *104*(4), 715–738. doi:10.1111/modl.12671

Webb, S. A., & Chang, A. C.-S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267–290. doi:10.3138/cmlr.1204.1

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13–40. doi:10.3138/cmlr.53.1.13

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition*, *42*(2), 411–438. doi:10.1017/S0272263119000688

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, *57*(4), 541–572. doi:10.3138/cmlr.57.4.541

Stuart Webb is a Professor of Applied Linguistics at the University of Western Ontario, Canada. He currently teaches on the Masters in TESOL program and supervises students at the M.A., Ph.D., and post-doctorate levels. His articles have been published in journals such as *Applied Linguistics* and *Language Learning*. His latest books are *The Routledge handbook of vocabulary studies* (Routledge), and *How vocabulary is learned* (Oxford University Press, with Paul Nation).

Takumi Uchihara is Assistant Professor at the Center for the English Language Education at Waseda University, Japan. His research interests include the teaching and learning of second language vocabulary. He is particularly interested in acquisition and assessment of spoken vocabulary knowledge. He has also been involved in several meta-analysis projects on various SLA-related topics including vocabulary and speech learning. He has served as a member of the *TESOL Quarterly* editorial board since March 2022.

Akifumi Yanagisawa is Assistant Professor at the Faculty of Humanities and Social Sciences at the University of Tsukuba, Japan. His research focuses on second language vocabulary acquisition, and he is particularly interested in cognitive factors that influence vocabulary learning. His studies have examined different factors and learning conditions such as task-induced involvement load, retrieval, and glossed reading. His work has appeared in journals such as *Language Learning*, *Studies in Second Language Acquisition*, and *Modern Language Journal*.