



# Comparing incidental vocabulary learning from reading-only and reading-while-listening

Yingzhao Chen

Michigan State University, B220, 619 Red Cedar Road, East Lansing, MI, 48823, United States



## ARTICLE INFO

### Article history:

Received 28 March 2020

Received in revised form 14 December 2020

Accepted 14 December 2020

Available online 19 December 2020

### Keywords:

Incidental vocabulary learning

Reading-while-listening

Proficiency

Lexicalization

## ABSTRACT

This study compares incidental vocabulary learning in reading-while-listening and reading-only conditions. Using both offline and online outcome measures, I assessed explicit form and meaning knowledge and the lexicalization of new words. I also explored how L2 listening and reading proficiency moderated learning gains in the two treatment conditions. L2 learners first read or read-while-listened to four short stories in English for meaning, with embedded target vocabulary items. They then completed surprise vocabulary posttests in the order of a form priming lexical decision task, a form recognition test, a form-meaning connection test and a semantic priming lexical decision task. Results showed that while the reading-while-listening group outperformed the reading-only group in recognizing the form and meaning of the target vocabulary, neither group fully lexicalized the new words, which was crucial for fluent lexical retrieval. L2 listening and reading proficiency affected learning from reading-while-listening and reading-only differentially: the reading-while-listening group was negatively affected by L2 reading proficiency when controlling for listening proficiency while performance of the reading-only group was not predicted by L2 proficiency. Implications for the use of bimodal input in incidental vocabulary learning are discussed.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

Vocabulary knowledge is essential to language development and use. For English second language (L2) learners, adequate comprehension of written texts requires knowledge of around 8000 word families and that of spoken English requires 6000–7000 word families (Nation, 2006). With only a few hours of explicit instruction per week, learning new words incidentally helps fill the gap between the lexical demands and the limited amount of time devoted to vocabulary instruction in an L2 classroom.

Incidental vocabulary learning (IVL) takes place when learners are not told about a vocabulary posttest (Hulstijn, 2003). Webb (2019) further argued that from a pedagogical perspective, it is the purpose of the activity that matters rather than whether attention is involved. Recent IVL research has investigated whether reading-while-listening (RWL), where learners read a text and listen to its audio simultaneously, offers advantages over reading-only (RO) (Brown, Waring, & Donkaewbua, 2008; Malone, 2018; Vu & Peters, 2020; Webb & Chang, 2012). Results from these studies are mixed and not all of them operationalized learning as truly incidental based on Hulstijn's (2003) definition, which call for more research to compare RWL and RO. Moreover, while L2 proficiency has been shown to affect IVL (e.g., Chen, Ma, Liang, & Liu, 2017; Elgort, Perfetti,

E-mail address: [chenyi38@msu.edu](mailto:chenyi38@msu.edu).

Rickles, & Stafura, 2015; Lee & Pulido, 2017; Peters & Webb, 2018), few studies on IVL from RWL have directly examined the potential moderating effects of overall L2 proficiency, and particularly related to RWL, listening and reading proficiency. Finally, studies on IVL from RWL often measured learning outcomes with offline measures only, failing to capture whether learners have lexicalized newly learned words, which is important for fluent lexical access in real-time communication (Godfroid, 2019).

Against this background, I administered a strict incidental learning condition in the current study, where there was no prior test announcement and learners' attention was directed to meaning, to compare IVL from RO and RWL. I also explored the roles of L2 listening and reading proficiency in IVL, the understanding of which will provide more fine-grained information to language practitioners on the use of RWL materials based on different learner profiles. Finally, reaction time (RT) measures, in addition to offline tests, were used to see whether RWL had an advantage over RO in form and meaning lexicalization, offering a fuller picture of the effects of bimodal input in vocabulary learning.

## 2. Literature review

### 2.1. Reading-while-listening vs. reading-only in incidental vocabulary learning

A number of studies have explored RWL in various areas in L2 learning such as the development of listening skill (e.g., Chang, 2009; Chang & Millett, 2014), reading skill (e.g., Chang & Millett, 2015; Gorsuch & Taguchi, 2008), and learner perception (e.g., Mestres, Baró, & Garriga, 2019; Tragant & Vallbona, 2018). In general, they have shown benefits or advantages of RWL over RO. In the area of IVL, six studies to date have investigated the effects of RWL or its comparison with RO. Two of them, Webb and Chang (2015) and Webb, Newton, and Chang (2013) showed vocabulary gains from RWL but did not include RO as a comparison. Three studies (Brown et al., 2008; Malone, 2018; Webb & Chang, 2012) pointed to the superior effect of RWL over single-modal presentation but showed mixed results regarding the comparison between RWL and RO. In Malone (2018), a lab-based study, participants read or read-while-listened to four short stories with embedded target words appearing twice or four times. They were then tested on the form and meaning of the new words. The RWL group outperformed the RO group in the form recognition test when learners encountered the target words twice and in the form-meaning connection test regardless of frequency of occurrence. Webb and Chang (2012), a classroom study, obtained similar results regarding the comparison between RWL and RO. Brown et al. (2008), however, showed a different scenario: the RWL group outperformed the listening-only group but not the RO group in form and meaning learning. The divergent result could partly be attributed to the operationalization of incidental learning in the studies. In Brown et al. (2008), participants were explicitly told that they were participating in a vocabulary learning program with the purpose to determine the best way to learn new words, which potentially have rendered the learning condition more intentional than incidental. The advantage of RWL over RO might have been mitigated with learners trying to learn the words intentionally. Vu and Peters (2020), though with a small sample size, added evidence to the lack of advantage of RWL over RO. With inconsistent findings from previous studies, more research is needed to obtain a clearer picture regarding the comparative effectiveness of RWL and RO.

A learner variable that has seldom been investigated directly in the comparison between RWL and RO is L2 proficiency. Previous IVL studies have indicated a positive relationship between proficiency and learning gains: those who scored higher in standardized tests did better in translation posttests (Lee & Pulido, 2017), and reacted faster to the new words in RT tasks (Chen et al., 2017). In Elgort et al. (2015), learners who had higher self-rated proficiency, larger vocabulary size and processed words faster as well as in a more stable manner formed better initial lexical representations of newly learned words. One of the IVL-from-RWL studies, Brown et al. (2008) touched upon L2 proficiency in relation to learners' perception of RWL, RO, and listening-only. The authors found that among the intermediate-level participants, more proficient ones, as indicated by their vocabulary size and TOEFL scores, did not always prefer RWL whereas less proficient ones predominantly did. None of the studies with RWL input mentioned earlier in this section, however, looked directly into L2 proficiency and learning gains.

Taken together, more studies with a strictly operationalized incidental learning condition are needed to examine the benefits of RWL as compared to RO in the area of vocabulary learning and to explore how L2 proficiency might moderate the effects of RWL. Unlike in most IVL studies, where a single overall score for L2 proficiency or vocabulary size was used, in the current study, I will analyze L2 listening and reading proficiency separately to explore how listening and reading skills might differentially affect learning in the RWL and RO conditions.

### 2.2. Measures of gains

In assessing learning gains in IVL studies, a crucial question to ask is how word knowledge should be defined. Nation (1990, 2013) proposed a comprehensive framework, including form, meaning and use as the three dimensions of vocabulary knowledge. Many IVL studies used offline paper-and-pencil tasks such as multiple choice and fill-in-the-blank to measure these aspects. These tasks often place no time restriction on learners and primarily tap into explicit lexical knowledge; nevertheless, it is one's implicit or automatized explicit knowledge that underpins performance in real-time language tasks such as reading and listening (e.g., Elgort & Warren, 2014; Godfroid, 2019). Godfroid (2019) argued that offline measures only reveal the tip of the iceberg regarding word knowledge, i.e., explicit-declarative knowledge, and suggested the use of three

online measures, namely RT tasks, coefficient of variability and eye tracking, to tap into learners' automatized explicit and implicit knowledge that can be accessed without or with less controlled attention in language processing.

In psycholinguistics and vocabulary research, a few terms share some commonalities with “implicit” or “automatized explicit” knowledge, including lexicalization (Qiao & Forster, 2013), and integration into the mental lexicon (Bordag, Kirschenbaum, Tschirner, & Opitz, 2015). When one lexicalizes or integrates a word into their mental lexicon, the word interacts with known words and can be accessed fluently in real time. In the current study, I will use the term “lexicalization” because this is what the priming tasks employed here directly measure. Whether the word knowledge is implicit or automatized explicit in nature is not under the scope of the current study.

A few IVL studies have used eye tracking to measure how learners' processing of new words changes over time as the number of encounters of those words increases (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016). Elgort and Warren (2014) is one of the earliest IVL studies to use RT measures to assess learning outcomes. They used two lexical decision tasks (LDT) with either form priming or semantic priming to measure form and meaning lexicalization. An LDT is a task that asks participants to decide whether a stimulus, i.e., the target, is a word or not. Priming is a technique of stimuli presentation (Jiang, 2013), where the target is preceded by another stimulus, i.e., the prime (see Fig. 1 for an illustration of an LDT with priming). The prime and target can be related in form as in form priming or meaning as in semantic priming. A priming effect is the facilitation or inhibition of the prime on participants' RT or accuracy to the target. Elgort and Warren failed to find evidence of form or meaning lexicalization after at least six exposures to the target vocabulary in written mode. They did, however, find inhibition of the prime on the target in the semantic priming LDT, taking it as evidence that learners were in the process of lexicalization (Dagenbach, Carr, & Barnhardt, 1990). Another IVL study, Bordag et al. (2015), also obtained inhibitory priming in their semantic priming task, but with only three exposures to the targets.

Overall, despite the importance of implicit and automatized explicit vocabulary knowledge in real-time communication, few researchers have used online measures to assess vocabulary learning outcomes in IVL studies. The scarcity highlights the need to employ online tests to measure learning in incidental conditions, especially learning with multimodal input, e.g., RWL, to see whether the benefits of multimodality shown by offline measures hold when learning is assessed online. Research reviewed in the above sections also pointed to the importance of having more studies with strictly defined incidental learning condition to compare the effectiveness of RWL and RO in vocabulary learning and to explore whether RWL's benefits also apply to form and meaning lexicalization. In addition, little is known about how L2 listening and reading levels moderate learning from RWL and RO. The following research questions will guide the study:

RQ1: How do the effects of RWL and RO compare on the development of explicit form and meaning knowledge in IVL?

RQ2: To what extent does learners' self-reported proficiency in L2 reading and listening moderate the development of explicit form and meaning knowledge in IVL from RWL and RO?

RQ3: How do the effects of RWL and RO compare on form and meaning lexicalization in IVL?

RQ4: To what extent does learners' self-reported proficiency in L2 reading and listening moderate form and meaning lexicalization in IVL from RWL and RO?

### 3. Method

#### 3.1. Participants

Fifty participants were recruited from a US university (see Appendix A for participants' L1 backgrounds, age, age of arrival (AOA), and length of residence (LOR)) and were randomly assigned to either the RWL or the RO group. Mann-Whitney U tests revealed no significant difference between the two groups in age, AOA or LOR (age:  $W = 285$ ,  $p = .59$ ,  $r = -0.08$ ; AOA:  $W = 294$ ,  $p = .71$ ,  $r = -0.05$ ; LOR:  $W = 330$ ,  $p = .73$ ,  $r = -0.05$ ). On a five-point scale, their average self-rating was 3.94 ( $Mdn = 4$ ,  $SD = 0.82$ , range: 2–5) for listening proficiency, 3.98 ( $Mdn = 4$ ,  $SD = 0.71$ , range: 3–5) for reading, 3.54 ( $Mdn = 4$ ,  $SD = 0.93$ , range: 1–5) for speaking and 3.48 ( $Mdn = 3$ ,  $SD = 0.81$ , range: 2–5) for writing, indicating that they identified themselves as high-intermediate learners. Mann-Whitney U tests did not show significant difference between the two groups' self-rated proficiency, except in writing, which had a small effect size (listening:  $W = 369$ ,  $p = .25$ ,  $r = -0.16$ ; reading:  $W = 378$ ,  $p = .17$ ,  $r = -0.20$ ; speaking:  $W = 359$ ,  $p = .34$ ,  $r = -0.14$ ; writing:  $W = 412$ ,  $p = .04$ ,  $r = -0.29$ ).

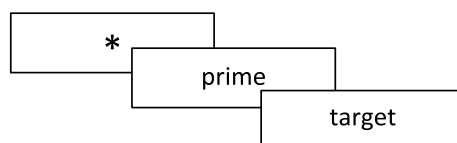


Fig. 1. Illustration of an LDT with priming.

### 3.2. Materials

I used the four stories from Malone (2018) as learning materials. The stories ranged from 694 to 773 words in length and according to the original study, around 96% of the words used belonged to the 4000 most frequent lemmas in the Corpus for Contemporary American English (COCA) (Davies, 2008). The materials were considered appropriate for comprehension given the similarity in participant profiles between the current study and Malone (2018). Each story contained eight target words and each target word appeared four times in the story. The texts were presented in timed PowerPoint slides in 32-point Calibri font. Each slide had five to nine lines of text. Participants saw two multiple-choice comprehension questions on the screen every two to three slides, with eight questions in total being asked for each story. The presentation order of the stories was randomized. For the RO group, a countdown clock appeared on the bottom right corner of each slide to show participants the amount of remaining time for reading. The maximum time allocated for an individual slide was determined by the duration of its audio recording to make sure the RWL and RO groups could have the same amount of time on each slide. For both groups, participants had 15s to answer each comprehension question before the screen automatically proceeded to the next slide. A male native speaker of English recorded the stories. Following Malone (2018), the recordings were at a rate of 120–140 words per minute.

I replaced the original 32 target words in Malone (2018) with 32 English pseudowords that were one letter different from real words so that the pseudowords could be included in the priming tasks to explore lexicalization. The pseudowords were four to six letters long and followed English spelling and phonotactic rules. Their meanings remained unchanged from the original target words in Malone (2018), which were all concrete nouns. None of target vocabulary items were important for answering the comprehension questions (See Appendix B for the list of target words.).

### 3.3. Instruments

#### 3.3.1. Form-recognition and form-meaning connection tests

To measure explicit form and form-meaning connection knowledge, I adopted the two tests used in Malone (2018). In the form-recognition test, participants saw a randomized list of the 32 target words and 32 distractors and were asked to circle the words that they had seen in the stories. To account for guessing, they received one point when they correctly identified a target word and lost one point when they circled a distractor. The total score for this test was 32 points. The form-meaning connection test asked participants to choose the meaning for each target word out of four choices. Each correct answer was awarded 1 point and the maximum score was 32 points (See Appendix C for full versions of the tests.).

#### 3.3.2. Form priming lexical decision task

I used a form priming LDT to measure form lexicalization of the target vocabulary. Following Elgort and Warren (2014), I looked for the Prime Lexicality Effect (PLE) (Forster & Veres, 1998), where priming effect is contingent on the lexical status, i.e., lexicality, of the prime: a form-related nonword prime facilitates the recognition of the target while a form-related real-word prime inhibits target recognition. For example, the nonword “stitus” would normally facilitate the recognition of a form-related word, e.g., “status”. But when a learner has lexicalized “stitus”, its lexical status changed. “Stitus” would behave similarly to a real English word and would produce inhibition. In the case of the current study, if the pseudowords in the learning materials [henceforth: pseudowords] inhibit the recognition of the targets in the LDT, one can argue that learners have lexicalized the forms of the pseudowords.

**3.3.2.1. Stimuli.** 96 word targets in the LDT were divided into three conditions and in each condition were preceded by one of the three types of primes: (1) form-related pseudowords [wender (meaning: a type of bird)- gender]; (2) form-related nonwords [shest-chest] and (3) unrelated word [teach-joke]. 96 nonword targets were created to balance the YES and NO responses in the LDT. The nonword targets were paired with (1) form-related word primes; (2) form-unrelated word primes and (3) nonword primes. There were altogether 192 trials (see Appendix D for a full set of task stimuli).

Following previous studies on the PLE (Davis & Lupker, 2006; Forster & Veres, 1998), all word targets in the LDT had low neighborhood density (Davis, 2005) (see Table 1) and all nonword targets were one letter different from real words to maximize the possibility of finding the effect. In order for the full list of 32 pseudowords, i.e., the target vocabulary items in the learning materials, to be included in the LDT so that participants could be tested on all of them, I created one list instead of three counterbalanced ones. Different word targets were used in the three conditions, and Kruskal-Wallis rank sum test

**Table 1**  
Means and SDs for word characteristics in the form priming LDT.

Condition	Related-pseudoword-word	Related-nonword-word	Unrelated-word-word
CELEX	35.61 (32.11)	37.80 (20.01)	40.85 (28.72)
Length	5.00 (0.62)	5.09 (0.64)	5.09 (0.59)
N	3.72 (2.04)	3.06 (2.15)	2.88 (1.50)

revealed that they did not differ significantly across conditions in CELEX frequency (Baayen, Piepenbrock, & Gulikers, 1995) ( $H(2) = 4.15, p = .13$ ), length ( $H(2) = 0.50, p = .78$ ) or neighborhood size ( $H(2) = 2.99, p = .22$ )<sup>1</sup> (see Table 1).

**3.3.2.2. Procedure.** Participants were instructed to decide whether the string of letters they saw on the screen was an English word or not as quickly and accurately as possible (see Appendix E for the instruction presented to participants). Each trial started with a row of five pairs of xw (xwxwxwxwxw) for 500 ms, followed by a prime presented in 10-point Times New Roman font for 56 ms. Finally, the target appeared in bold 11-point Arial Black font. I used different fonts and sizes for primes and targets to avoid graphemic overlap between them.

### 3.3.3. Semantic priming lexical decision task

In this task, if the pseudowords show a facilitative priming effect, it is taken as a sign that meaning lexicalization has taken place.

**3.3.3.1. Stimuli.** 64 word targets were divided into two conditions and in each condition were preceded by one of the following types of primes: (1) semantically related pseudowords [wender (a type of bird) -bird] and (2) semantically unrelated words [body-mail]. 64 nonword targets were created to pair with 32 word primes and 32 nonword primes for NO responses. There were a total of 128 trials in this task (see Appendix F for the list of stimuli). Semantic relatedness was operationalized as hyponymy, hypernymy, synonymy or semantic overlap. The mean similarity values between the primes and the targets in the control condition was 0.07 ( $SD = 0.08$ ), as obtained from the Latent Semantic Analysis Online Tool (<http://lsa.colorado.edu>), suggesting that they were not semantically related. For similar reasons as in the form priming task, different word targets were used in the two conditions and Mann-Whitney U tests revealed that the word targets did not differ across conditions (CELEX frequency:  $W = 545, p = .07, r = -0.06$ ; length:  $W = 497, p = .83, r = -0.03$ )<sup>2</sup> (see Table 2).

**3.3.3.2. Procedure.** Participants were instructed to decide as fast and accurately as possible whether a string of letters is an English word (see Appendix G for the instruction presented to participants). A fixation point (\*) was presented for 500 ms, followed by the prime for 450 ms and then the target for 500 ms. The primes and targets were both in 10-point Times New Roman font.

### 3.3.4. Reliability of measures

Table 3 presents the reliabilities of instruments calculated with the CTT package (version 2.3.3) (Willse, 2018) for Cronbach alphas and with the Multicon package (version 1.6) (Sherman, 2015) for split-half reliabilities in R.

### 3.3.5. Language background questionnaire and exit interview

Participants provided information about their language learning experiences and self-rated proficiency scores in an online language background questionnaire (see Appendix H). They answered the following questions in an exit interview at the end of the experiment: (1) whether they noticed the target words in the four short stories; (2) whether they tried to learn or memorized the target words; (3) whether they expected vocabulary posttests.

## 3.4. Procedure

Participants completed the language background questionnaire online before coming to the lab to finish the remaining tasks in the following order: reading-only or reading-while-listening comprehension, form priming LDT, form-recognition task, form-meaning connection task, and semantic priming LDT, which altogether took around an hour. The ordering of the tasks was to maximally prevent participants from getting hints from a previous task. No more than two participants were present in the lab at the same time.

The LDTs were conducted with the Windows DMDX software (Forster & Forster, 2003). For all LDTs in the study, participants pressed the right shift key for YES responses and left shift key for NO responses. Feedback on the speed and accuracy of the response was given after each trial. All test items were randomized and 10 practice items were presented before each LDT.

**Table 2**  
Means and SDs for word characteristics in the semantic priming LDT.

Condition	Related-pseudoword-word	Unrelated-word-word
CELEX	86.33 (85.90)	56.13 (27.70)
Length	5.00 (1.05)	5.6 (0.88)



**Table 3**  
Reliability of measures.

Test	Reliability	Coefficient
Form recognition	Cronbach alpha	.83
Form-meaning connection	Cronbach alpha	.80
Form-priming LDT	Split half	.99
Semantic-priming LDT	Split half	.98

### 3.5. Analysis

All statistical analyses were performed in R (version 3.6.1, [R Core Team, 2019](#)). Assumptions of statistical tests were checked and in cases where they were violated, non-parametric tests were used. Raw data and R scripts are available on Open Science Framework (<https://osf.io/uc4h9/>).

All participants got over 81% accuracy in the reading comprehension, indicating adequate comprehension of the texts. One participant did not complete the form-meaning connection task and their data was excluded from the analyses of form-meaning connection test scores. Analysis of the exit interview revealed that while all of the participants noticed the target words, only two expected vocabulary posttests, confirming that the learning condition was incidental. However, in terms of the attempt to learn the target words, out of the 47 answers collected, 14 were a “No”, 19 were an affirmative “Yes”, and the remaining were ambiguous responses like “sort of” and “not always”, showing that in spite of the learning condition being incidental, learning could be intentional ([Godfroid et al., 2018](#); [Hulstijn, 2003](#)). The effects of intention on learning gains will be addressed in the Discussion session (See [Appendix I](#) for participants’ responses.).

To address RQ1, which compared the effects of RO and RWL on gains of explicit form and meaning knowledge, I conducted multiple t-tests to compare the means between the RO and RWL groups. For RQ2, which concerned the effects of L2 listening and reading proficiency on explicit learning gains of the RO and RWL groups, I first fit two multiple regression models, each for the form-recognition and form-meaning connection tests. Both had Group (RO vs. RWL), participants’ self-rated proficiency scores on listening and reading, and the interaction between Group and the two proficiency scores as predictors. Self-rated listening and reading proficiency scores were mean-centered to avoid multicollinearity, making sure that the models’ VIFs are smaller than 5 ([Heiberger & Holland, 2004](#)) and that correlations between predictors are lower than 0.90 ([Plonsky & Ghanbar, 2018](#)). When I found a significant interaction between Group and the proficiency scores, I built separate models for the RO and RWL groups. I removed influential cases with Cook’s distance higher than one ([Cook & Weisberg, 1982](#)) and will present the final models in the Results session.

To answer RQ3 and RQ4, which examined the effects of RO and RWL on form and meaning lexicalization and the roles of L2 listening and reading proficiency, I fit mixed-effects models to RT of correct YES responses in the form and semantic priming LDTs. Participants with an accuracy rate below 80% in the LDTs were excluded from the analysis. I trimmed RT that were below 300 ms or above 2000 ms. I also excluded RT that were 3 SDs away from each participant’s mean. Such criteria for RT outliers were set based on [Jiang \(2013\)](#). The trimming was conducted with the trimr package (version 1.0.1) ([Grange, 2015](#)). All RT data was then inverse transformed ( $RT_{inv} = -1000/RT$ ) to bring the distribution closer to normal. [Table 4](#) presents the percentage of RT data trimmed and the total observations included in the final analysis.

The lmer package (version 1.1–21) ([Bates, 2005](#)) was used to build linear mixed-effects models and p values were calculated using the lmerTest package (version 3.1–0) ([Kuznetsova, Brockhoff, & Christensen, 2017](#)). In building the models, I pre-specified the fixed effects (Condition, L2 reading and listening proficiency, and their interactions) and engaged in modelling procedure for random effects by first including random intercepts for participants and items, followed by adding by-participant and by-item random slopes. The modelling process ceased when adding more random effects did not improve the model or when the model failed to converge. I used the Akaike information criterion (AIC) and ANOVA to compare different models. In the Results session, I will report the final models. All models were fit using a maximum likelihood technique.

## 4. Results

### 4.1. RQ1: Explicit word knowledge

[Table 5](#) presents the descriptive statistics of the RO and RWL groups for the form-recognition and form-meaning connection tests. T-test analyses (see [Table 6](#)) showed that the two groups differed significantly in both tests. The effect

**Table 4**  
Percentage of RT data trimmed and total observations.

	RO		RWL	
	% trimmed	Total observations	% trimmed	Total observations
Form priming LDT	3.05	4095	2.40	4317
Semantic priming LDT	3.33	2872	2.12	2812

**Table 5**

Descriptive statistics of the RO and RWL groups.

	Form-recognition				Form-meaning connection			
	n	Mean	SD	95% CI	n	Mean	SD	95% CI
RO	25	16.68	7.20	[13.71, 19.65]	25	19.48	4.81	[17.49, 21.47]
RWL	25	21.16	5.59	[18.85, 23.47]	24	23.13	5.02	[21.00, 25.25]

sizes were medium, based on [Plonsky and Oswald \(2014\)](#)'s guidelines, indicating that the differences were important in magnitude.

#### 4.2. RQ2: The role of L2 proficiency

[Table 7](#) presents the Spearman correlation coefficients between test scores and proficiency. Multiple regression models (see [Tables 8 and 9](#)) showed that Group but not self-rated listening and reading proficiency significantly predicted form and meaning test scores. For the form-recognition test, there was a significant interaction between Group and self-rated listening proficiency, indicating that listening proficiency had a differential effect on explicit form learning of the RWL and RO groups.

[Tables 10 and 11](#) present the models for the RWL and RO groups respectively. One participant was removed from the model for the RWL group due to a high Cook's distance ( $D = 2.68$ ). For the RWL group, L2 listening proficiency was a more important predictor than L2 reading proficiency. Interestingly, the effects of the L2 listening and reading proficiency scores went in opposite directions as shown by their regression coefficients. Neither of the L2 proficiency scores significantly predicted the form-recognition test scores of the RO group.

#### 4.3. RQ3 and 4: Form and meaning lexicalization and the role of L2 proficiency

[Tables 12 and 13](#) present the mean RT of the RWL and RO groups in the LDTs. The RWL group was faster than the RO group in responding to words as well as nonwords, though linear mixed-effects modelling showed no significant group difference in the overall RT (form priming LDT:  $t = .67$ ,  $SE = 0.05$ ,  $p = .51$ ; semantic priming LDT:  $t = 1.77$ ,  $SE = 0.05$ ,  $p = .08$ ). Linear mixed-effects models showed no PLE or semantic priming for either the RWL or the RO group regardless of their L2 listening and reading proficiency (see [Tables 14–17](#)).

## 5. Discussion

The present study implemented a strict incidental learning condition to compare the effects of RWL and RO on vocabulary learning. In an attempt to contribute to a better understanding of bimodal input in IVL, the study extended previous studies by assessing both explicit word knowledge and lexicalization of form and meaning. It also explored the effects of skill-based proficiency levels of L2 reading and listening on learning from RWL and RO, revealing how the effects of bimodal input varied across learner profiles.

#### 5.1. RWL is better, but not in the same way for everyone

With regard to RQ1 on the development of explicit word knowledge, results show that participants were able to establish form-meaning connection for more than half of the target words and that RWL offered an advantage over RO in the learning of both form and meaning knowledge, largely corroborating findings in [Webb and Chang \(2012\)](#) and [Malone \(2018\)](#). It should be noted, however, that in [Malone \(2018\)](#), when learners were exposed to the new words four times, the advantage of RWL manifested only in the learning of explicit meaning knowledge while in the current study, the RWL group outperformed the RO group in both form and meaning learning. A possible explanation for the discrepancy is the use of pseudowords in the current study, as compared to infrequent real words in [Malone \(2018\)](#). In a meta-analysis, [Uchiyara, Webb, and Yanagisawa \(2019\)](#) revealed that the use of pseudowords could inflate learning gains because pseudowords are more likely to stand out than real words when surrounded by high-frequency words in the learning materials.

**Table 6**

T test results.

	<i>t</i>	<i>df</i>	<i>p</i>	FDR adjusted <i>p</i>	<i>d</i>	95% CI of group difference
Form-recognition	2.46	48	.02*	.02*	.70	[1.06, 8.04]
Form-meaning connection	2.59	47	.01**	.02*	.74	[-.84, 6.39]

\* $p < .05$ .\*\* $p < .01$ .\*\*\* $p < .001$ .

**Table 7**

Correlation matrix of test scores and self-rated proficiency.

Variable	Form-recognition	Form-meaning connection	Self-rated listening proficiency	Self-rated reading proficiency
Form-recognition	1.00	.43	.18	.23
Form-meaning connection		1.00	.16	.09
Self-rated listening proficiency			1.00	.70
Self-rated reading proficiency				1.00

**Table 8**

Multiple regression for the form-recognition test.

	<i>B</i> [95% CI]	$\beta$	<i>sr</i> <sup>2</sup>	<i>p</i>
<i>R</i> <sup>2</sup>	.25			.02*
Adjusted <i>R</i> <sup>2</sup>	.16			
Intercept	16.82 [14.26, 19.37]			
Group ( <i>RO</i> = 0, <i>RWL</i> = 1)	3.87 [.25, 7.48]	-.29	.01	.04*
Self-rated listening	-2.48 [-6.32, 1.37]	-.30	4e-04	.20
Self-rated reading	3.48 [-1.36, 8.32]	.37	.002	.15
Group * Self-rated listening	7.34 [1.05, 13.63]	.58	.006	.02*
Group * Self-rated reading	-4.98 [-12.09, 2.13]	-.38	4e-04	.17

\**p* < .05.\*\**p* < .01.\*\*\**p* < .001.**Table 9**

Multiple regression for the form-meaning connection test.

	<i>B</i> [95% CI]	$\beta$	<i>sr</i> <sup>2</sup>	<i>p</i>
<i>R</i> <sup>2</sup>	.21			.07
Adjusted <i>R</i> <sup>2</sup>	.11			
Intercept	19.20 [17.18, 21.22]			
Group ( <i>RO</i> = 0, <i>RWL</i> = 1)	3.68 [.79, 6.56]	.36	.01	.01**
Self-rated listening	1.21 [-1.84, 4.25]	.19	1e-04	.43
Self-rated reading	-3.12 [-7.01, .65]	-.44	4.9e-05	.10
Group * Self-rated listening	-.20 [-5.17, 4.78]	-.02	1e-04	.94
Group * Self-rated reading	3.88 [-1.75, 9.50]	.39	.003	.17

\**p* < .05.\*\**p* < .01.\*\*\**p* < .001.**Table 10**Multiple regression for *RWL* in the form-recognition test.

	Total <i>R</i> <sup>2</sup>	Adjusted <i>R</i> <sup>2</sup>	Intercept	Listening	Reading
<i>B</i>	.46	.41	19.98	9.94	-6.68
95% CI			[18.04, 21.92]	[4.56, 15.31]	[-12.24, -1.13]
$\beta$				1.29	-.84
<i>sr</i> <sup>2</sup>				.12	.01
<i>p</i>	.001***			.0009***	.02*

\**p* < .05.\*\**p* < .01.\*\*\**p* < .001.**Table 11**Multiple regression for *RO* in the form-recognition test.

	Total <i>R</i> <sup>2</sup>	Adjusted <i>R</i> <sup>2</sup>	Intercept	Listening	Reading
<i>B</i>	.07	-.01	16.82	-2.48	3.48
95% CI			[13.75, 19.89]	[-7.10, 2.15]	[-2.34, 9.30]
$\beta$				-.30	.33
<i>sr</i> <sup>2</sup>				9e-04	.002
<i>p</i>	.44			.28	.23

\**p* < .05.\*\**p* < .01.\*\*\**p* < .001.



**Table 12**  
Mean RT (SD) for the form priming LDT.

Condition	RWL	RO
Related-pseudoword-word	698 (187)	724 (215)
Related-nonword-word	707 (183)	734 (219)
Unrelated-word-word	699 (160)	726 (215)

As for RQ2 about the roles of L2 listening and reading skills in the learning of explicit knowledge, results revealed that L2 listening and reading proficiency predicted explicit form learning of the RWL but not the RO group. This suggested that bimodal input, requiring learners to read and listen simultaneously, might induce a heavier cognitive load, which accentuated the role of language proficiency. L2 listening proficiency was a more important predictor than reading proficiency of form learning for the RWL group, suggesting that higher L2 listening proficiency might have helped ease the cognitive load imposed by bimodal input. Another hypothesis is that only when one can easily process and understand the auditory input can one enjoy the benefits of simultaneous input from two modalities, i.e., hearing and seeing a word; otherwise, one would struggle to figure out what they hear while reading, only taking advantage of the visual input. Interestingly, for the RWL group, while gains in explicit form knowledge were positively predicted by L2 listening proficiency, they were negatively related to L2 reading level when controlling for L2 listening proficiency. That is, higher L2 reading proficiency reduced the effectiveness of RWL. A possible scenario is that skilled L2 readers read faster and hence were ahead of the audio. In fact, using eye tracking, [Conklin, Alotaibi, Pellicer-Sánchez, and Vilkaitė-Lozdienė \(2020\)](#) showed that L2 learners with higher proficiency were more likely to read ahead of the audio in the RWL condition. The misalignment between reading and listening of skilled readers might render the audio disruptive because they will be looking at one word and hearing a different one. Proficiency did not predict performance in the form-meaning connection test, indicating that fast decoding of auditory and visual input does not necessarily lead to successful guess or memory encoding of word meanings. The results resonated with [Malone \(2018\)](#), who only found an effect of working memory on the learning of form but not meaning. To establish form-meaning connection in incidental learning, more than working memory and L2 proficiency might be needed, such as effort to figure out word meaning ([Godfroid et al., 2018](#)).

## 5.2. Lexicalization

Similar to [Elgort and Warren \(2014\)](#), I failed to find the PLE in the form priming LDT. Nor was I able to find any priming in the semantic priming LDT, unlike [Bordag et al. \(2015\)](#) and [Elgort and Warren \(2014\)](#), who uncovered inhibitory semantic priming. The lack of PLE or facilitatory semantic priming in these studies shows that full lexicalization is a slow process and suggests that it takes more than six encounters of a word in a single modal context ([Elgort & Warren, 2014](#)) or four in a bimodal condition (the current study) in incidental learning. The results largely align with eye-tracking studies, which showed that new words were processed similarly to familiar words after eight ([Pellicer-Sánchez, 2016](#)) or eleven ([Godfroid et al., 2018](#)) exposures.

An important caveat is that the lack of PLE in this study might be related to other factors than the absence of form lexicalization. It has been shown that some time for knowledge consolidation was necessary to obtain PLE ([Qiao & Forster, 2013](#)). In the current study, participants completed the form priming task right after the learning phase, which might account for the null result. [Qiao and Forster \(2017\)](#) also found that newly learned words showed facilitatory priming and the more training L2 learners received on the new words, the stronger the facilitation was. This phenomenon contradicted with what the PLE predicts, i.e., inhibition, and with the behavior of English native speakers. The authors thus hypothesized that the L1 and L2 words were stored in different memory systems and the PLE would not be borne out with L2 learners.

## 5.3. Intention to learn

Although the learning condition in the current study was incidental, nearly half of the participants reported trying to learn the target words in the exit interview. I conducted a follow-up analysis to explore how their intention to learn has affected learning gains. I classified the participants into two groups based on their response to the question of whether they tried to learn or memorize the target words: those who gave a “Yes” or an ambiguous answer, such as “a bit” and “sometimes”, as long as their answer indicated the intention to learn, were put in the intentional group and the others in the unintentional group. ANOVA revealed that there was a Group (RO vs. RWL) by Intention interaction ( $F_{(1,42)} = 4.14, p = .05, \text{partial } \eta^2 = 0.09$ ) in the learning of word meanings: while intention did not affect the RWL group ( $t = 0.75, p = .47, d = 0.36, 95\% \text{ CI } [-2.99, 6.09]$ ), it adversely affected the RO group ( $t = -2.49, p = .03, \text{FDR adjusted } p = .06, d = 0.98, 95\% \text{ CI } [-8.11, -0.56]$ ). Intention did not have an impact on form learning ( $F_{(1,42)} = 0.93, p = .34, \text{partial } \eta^2 = 0.02$ ) of either the RWL or the RO group. One reason for the negative impact of intention on the RO group could be that the attempt to figure out word meanings interfered with their reading pace while with the guidance of the auditory input, the pace of the RWL was disrupted to a lesser extent.

In terms of lexicalization, results pointed to an advantage of RWL over RO in meaning lexicalization: unintentional learners in the RWL group showed inhibitory semantic priming ( $t = 2.30, p = .03$ ), which is an indication that learners were in the process of lexicalization ([Bordag et al., 2015](#); [Dagenbach et al., 1990](#); [Elgort & Warren, 2014](#)). Intention did not affect form

**Table 13**  
Mean RT (SD) for the semantic priming LDT.

Condition	RWL	RO
Related-pseudoword-word	673 (167)	718 (221)
Unrelated-word-word	655 (149)	723 (217)

**Table 14**  
Linear mixed-effects model for RWL group's RT in form priming LDT.

	Fixed effects				Random effects		
	Estimate	SE	t	p	Group	Intercept	SD
Intercept	−1.44	0.17	−8.37	<.001	Participant	.02	.13
Related-pseudoword-word	−.06	.09	−.65	.52	Item	.01	.11
Related-word-word	.01	.09	.11	.92			
Listening proficiency	.005	.06	.09	.93			
Reading proficiency	−.02	.06	−.31	.76			
Listening proficiency* Related-pseudoword-word	−.01	.03	−.40	.69			
Reading proficiency * Related-pseudoword-word	.02	.03	.72	.47			
Listening proficiency* Related-word-word	.03	.03	.94	.35			
Reading proficiency*Related-word-word	−.03	.03	−.92	.36			

\*p < .05.

\*\*p < .01.

\*\*\*p < .001.

**Table 15**  
Linear mixed-effects model for RO group's RT in form priming LDT.

	Fixed effects				Random effects		
	Estimate	SE	t	p	Group	Intercept	SD
Intercept	−1.27	0.24	−5.31	<.001	Participant	.04	.19
Related-pseudoword-word	−.001	.09	−.02	.99	Item	.01	.10
Related-word-word	.01	.09	.14	.89			
Listening proficiency	−.13	.06	−2.18	.04*			
Reading proficiency	.08	.08	1.03	.31			
Listening proficiency* Related-pseudoword-word	−.01	.02	−.71	.48			
Reading proficiency * Related-pseudoword-word	.01	.03	.45	.65			
Listening proficiency* Related-word-word	−.002	.02	−.07	.94			
Reading proficiency*Related-word-word	.003	.03	.09	.93			

\*p < .05.

\*\*p < .01.

\*\*\*p < .001.

**Table 16**  
Linear mixed-effects model for RWL group's RT in semantic priming LDT.

	Fixed effects				Random effects		
	Estimate	SE	t	p	Group	Intercept	SD
Intercept	−1.81	0.20	−8.91	<.001	Participant	.02	.15
Related-pseudoword-word	.04	.09	.47	.64	Item	.01	.10
Listening proficiency	.02	.07	.28	.78			
Reading proficiency	.03	.07	.48	.64			
Listening proficiency* Related-pseudoword-word	−.009	.03	−.33	.74			
Reading proficiency * Related-pseudoword-word	.008	.03	.28	.78			

\*p < .05.

\*\*p < .01.

\*\*\*p < .001.

lexicalization (see [Appendix J](#) for full results). Inhibitory priming obtained here seems to correspond to the early familiarization stage, which was indexed by a rapid decrease of reading time within a few exposures to a new word in eye-tacking studies ([Godfroid et al., 2018](#); [Pellicer-Sánchez, 2016](#)). The results are also comparable to those in [Toomer and Elgort \(2019\)](#), where implicit knowledge only developed in the unenhanced reading treatment, i.e., when attention was not explicitly drawn to the target words. Jointly, the results suggest that while bimodal input might accelerate the process of lexicalization, full lexicalization takes time to develop and is less amenable to the depth of processing afforded by the intention to learn.

**Table 17**

Linear mixed-effects model for RO group's RT in semantic priming LDT.

	Fixed effects				Random effects		
	Estimate	SE	t	p	Group	Intercept	SD
Intercept	−1.39	0.22	−6.13	<.001	Participant	.03	.19
Related-pseudoword-word	−.10	.09	−1.18	.24			
Listening proficiency	−.14	.06	−2.36	.03*	Item	.01	.11
Reading proficiency	.12	.07	1.59	.12			
Listening proficiency* Related-pseudoword-word	.01	.02	.68	.50			
Reading proficiency * Related-pseudoword-word	.008	.03	.29	.78			

\*p &lt; .05.

\*\*p &lt; .01.

\*\*\*p &lt; .001.

## 6. Limitations and conclusions

The results of this study confirm the advantage of RWL over RO in learning explicit form and meaning knowledge incidentally while point out that the effects of RWL were moderated differentially by learners' L2 listening and reading skills. The study also adds to our understanding of how bimodal input affects lexicalization in IVL. RT measures revealed that RWL might have an advantage over RO in meaning lexicalization only for unintentional learners but not in form lexicalization nor for intentional learners. The findings support the use of bimodal input to facilitate IVL and underscore the importance to consider learners' proficiency profiles. While better listeners learned more from RWL input, RWL might be more beneficial for poor readers than skilled ones. It is also advised that the speed of the audio in RWL be adjusted based on one's reading speed, which would reduce the chance of misalignment between reading and listening to avoid any potential harmful effects of RWL.

The results, however, should be interpreted with several caveats. First, using pseudowords, while eliminating the possibility of prior knowledge, compromised ecological validity. The use of pseudowords might also have inflated learning because pseudowords were more likely to stand out. Alternatively, pseudowords represent concepts that are already connected to known words (Pellicer-Sánchez, 2016) and it is unclear how this might have affected the development of explicit knowledge and lexicalization. Further, self-rated proficiency, particularly when measured with a five-point scale only, is not the most accurate representation of one's abilities. The L1 backgrounds of the participants were primarily Chinese, which could potentially have moderated the effects of bimodal input given the differences in writing system between English and Chinese. Finally, lexicalization, in its initial stage, might have gone undetected due to the limited number of items included in the LDTs and insufficient sensitivity of RT measures.

Overall, the study demonstrated the advantage of RWL over RO in the development of explicit knowledge. It also took initial steps in exploring lexicalization and the roles of skill-based L2 proficiency, providing a fuller picture of the use of bimodal input in IVL. Future studies are needed to address the limitations of the current study by implementing more reliable measures of L2 proficiency and using other online tasks to further our understanding of RWL in IVL.

## Notes

1,2: Participants' baseline RT to the targets was also obtained to make sure targets across conditions were comparable. Analyses revealed that the targets' baseline RT did not differ significantly across conditions. Statistical analyses with the inclusion of baseline RT did not change the conclusion of the current study (see Appendix K for details).

## Funding

This study was supported by the graduate student research funding from the Second Language Studies Program at Michigan State University.

## Acknowledgements

I would like to thank the anonymous reviewers, Dr. Aline Godfroid, Curtis Green-Eneix, Bronson Hui, Dr. Nan Jiang, Ryo Maie, Eva Puimege, and Dr. Patti Spinner for their constructive feedback. Any errors remain my own.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.system.2020.102442>.

## Author Statement

Yingzhao Chen is the sole author of the manuscript titled “Comparing incidental vocabulary learning from reading-only and reading-while-listening”.

The author declares that she has no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

## References

- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database [CD-ROM]*. Linguistic Data Consortium. University of Pennsylvania.
- Bates, D. (2005). Fitting linear models in R: Using the lme4 package. *R News*, 5, 27–30.
- Bordag, D., Kirschenbaum, A., Tschirner, E., & Opatz, A. (2015). Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon. *Bilingualism: Language and Cognition*, 18(3), 372–390. <https://doi.org/10.1017/S1366728914000078>
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136–163.
- Chang, A. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, 37(4), 652–663. <https://doi.org/10.1016/j.system.2009.09.009>
- Chang, A., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, 68(1), 31–40. <https://doi.org/10.1093/elt/cct052>
- Chang, A., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, 52, 91–102. <https://doi.org/10.1016/j.system.2015.05.003>
- Chen, B., Ma, T., Liang, L., & Liu, H. (2017). Rapid L2 word learning through high constraint sentence context: An event-related potential study. *Frontiers in Psychology*, 8, 2285. <https://doi.org/10.3389/fpsyg.2017.02285>
- Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaite-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*. <https://doi.org/10.1177/0267658320921496>
- Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. Chapman & Hall.
- Dagenbach, D., Carr, T. H., & Barnhardt, T. M. (1990). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 328–340. <https://doi.org/10.1037/0278-7393.16.2.328>
- Davies, M. (2008). The corpus of contemporary American English: 450 million words, 1990–present. Retrieved from <http://corpus.byu.edu/coca/>.
- Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37, 65–70. <https://doi.org/10.3758/BF03206399>
- Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 668–687. <https://doi.org/10.1037/0096-1523.32.3.668>
- Elgart, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, 30(5), 506–528. <https://doi.org/10.1080/23273798.2014.942673>
- Elgart, 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(2), 365–414. <https://doi.org/10.1111/lang.12052>
- Forster, K. I., & Forster, J. C. (2003). Dmxd: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 498–514. <https://doi.org/10.1037/0278-7393.24.2.498>
- Godfroid, A. (2019). Sensitive measures of vocabulary knowledge and processing: Expanding Nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). Routledge. <https://doi.org/10.4324/9780429291586-28>
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., et al. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584. <https://doi.org/10.1017/S1366728917000219>
- Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System*, 36(2), 253–278. <https://doi.org/10.1016/j.system.2007.09.009>
- Grange, J. A. (2015). rimir: An implementation of common response time trimming methods. T. R package version 1.0.1 <https://cran.r-project.org/web/packages/trimr/index.html>
- Heiberger, R. M., & Holland, B. (2004). *Statistical analysis and data display*. Springer. <https://doi.org/10.1007/978-1-4939-2122-5>
- Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell. <https://doi.org/10.1002/9780470756492.ch12>
- Jiang, N. (2013). *Conducting reaction time research in second language studies*. Routledge.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, 21(1), 118–135. <https://doi.org/10.1177/1362168816637381>
- Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651–675. <https://doi.org/10.1017/S0272263117000341>
- Mestres, E. T., Baró, A. L., & Garriga, A. P. (2019). Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English. *Reading and Writing*, 32(3), 819–838. <https://doi.org/10.1007/s1145-018-9886-x>
- Nation, P. (1990). *Teaching and learning vocabulary*. Newbury House.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97–130. <https://doi.org/10.1017/S0272263115000224>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R<sup>2</sup> values. *The Modern Language Journal*, 102(4), 713–731. <https://doi.org/10.1111/modl.12509>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Qiao, X., & Forster, K. I. (2013). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1064–1074. <https://doi.org/10.1037/a0030528>

- Qiao, X., & Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition*, 158, 147–152. <https://doi.org/10.1016/j.cognition.2016.10.026>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sherman, R. A. (2015). Multicon: Multivariate constructs. R package version 1.6 <https://CRAN.R-project.org/package=multicon>.
- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>
- Tragant, E., & Vallbona, A. (2018). Reading while listening to learn: Young EFL learners' perceptions. *ELT Journal*, 72(4), 395–404. <https://doi.org/10.1093/elt/ccy009>
- 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. <https://doi.org/10.1111/lang.12343>
- Vu, D. V., & Peters, E. (2020). Learning L2 vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*. <https://doi.org/10.1177/0033688220911485>
- Webb, S. (2019). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge. <https://doi.org/10.4324/9780429291586-15>.
- Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review*, 68, 276–290. <https://doi.org/10.3138/cmlr.1204.1>
- 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(6), 667–686. <https://doi.org/10.1177/1362168814559800>
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91–120. <https://doi.org/10.1111/j.1467-9922.2012.00729.x>
- Willse, J. T. (2018). CTT: Classical test theory functions. R package version 2.3.3 <https://CRAN.R-project.org/package=CTT>.