

ORIGINAL ARTICLE

Open Access



The effectiveness of L2 vocabulary instruction: a meta-analysis

Mohammad Hossein Yousefi^{1*} and Reza Biria²

* Correspondence: mhh.yousefi@gmail.com

¹Department of English, Islamic Azad University, Bonab Branch, Bonab, Iran
Full list of author information is available at the end of the article

Abstract

The purpose of the present meta-analysis is to investigate the overall effectiveness of L2 vocabulary instruction and to find the moderator variables affecting its effectiveness. By defining a rigorous inclusion and exclusion criteria, a total number of 16 primary studies ($N = 1008$), 7 published and 9 Ph.D. dissertations, were included. Under Random-Effects Model, the overall effect size of ($d = 0.80$) was observed. After conducting Q test of heterogeneity, a number of moderator variables were examined; context of instruction, publication type, age and L2 learners' proficiency level. It was found that (a) studies conducted in foreign language contexts generated larger effect sizes than ones conducted in SL contexts. (b) intermediated learners show a larger effect size than advanced and elementary students. (c) child learners were better than adult learners in Learning L2 vocabulary. (d) Published studies generated larger effect size than doctoral dissertations. (e) employing "posters" for teaching L2 vocabulary items generated higher effect size than reading activities, CALL, and songs. (f) abstract words generated higher effect size than concrete ones. Possible explanations of the findings are discussed with regard to the similar meta-analyses in the field and directions for future research are proposed.

Keywords: Meta-analysis, Second language vocabulary learning, Effect size, Moderator variables, Random-effects model, Publication bias

Introduction

Nowadays, nobody has reservations about the effectiveness of instructed (tutored) language learning. In instructed second language acquisition, the learner typically focuses on some aspect of language system (Klein, 1986). Many primary studies conducted in the field of SLA provided support in favor of instructed language learning. In the same vein, a number of meta-analyses demonstrated overall effectiveness of teaching dimensions of second languages. For example, L2 grammar acquisition (Shintani, 2015) corrective feedback (Li, 2010) and second language strategy instruction (Plonsky, 2011).

Undoubtedly, it is generally agreed that language vocabulary is an essential part of learning a second language (Fehr et al., 2012; Ko, 2012; Nation, 2001; Schmitt, 2008) and the lexicon may be the most important language component for learners (Hamada & Koda, 2008; Yamamoto, 2013). Lexical proficiency is also crucial because the understanding of lexical acquisition in relation to its deeper, cognitive functions can lead to increased awareness of how learners process and produce an L2 (Crossley et al., 2009). In what follows, we review a number of issues related to L2 vocabulary teaching.

Several meta-analyses have been conducted on some aspects of L2 vocabulary teaching. Huang (2010) conducted a systematic statistical synthesis of the effects of output stimulus tasks on L2 incidental vocabulary learning. A total of 12 studies were included in this meta-analysis. Results showed that language learners gained more benefit from using output stimulus tasks to learn vocabulary than those who only read a text. For these 16 studies, the mean effect size was 1.39 (SE = .07).

Given the fuzziness of the variables affecting L2 vocabulary learning and in order to gain a more reliable picture of what factors actually affect L2 vocabulary teaching, conducting a quantitative meta-analysis is justified. Because meta-analysis is a standard, well-grounded statistical procedure for combining the evidence from independent studies that address the same research hypothesis (Normand, 1999). A meta-analysis has three advantages. First, it provides research findings in a sophisticated fashion, which differs from findings represented in statistical significance. Second, it is able to detect effects that are obscure in narrative summaries of findings. Third, it provides a systematic approach to analyzing information from a large number of research findings (Lipsey & Wilson, 2001).

Literature review

In this section, we first review two distinct approaches to L2 vocabulary teaching and critically discuss the empirical studies related to these theoretical underpinnings. Then, we discuss the effects of a number of input and output-based tasks and activities on L2 vocabulary learning. Finally, the related meta-analyses will be subject to critical review.

Many vocabulary learning theories divide vocabulary study into two distinct approaches: explicit vocabulary learning and implicit vocabulary learning (Hulstijn, 2001; Nassaji, 2003). Incidental vocabulary learning is “learning without an intent to learn, or as the learning of one thing, for example vocabulary, when the student’s primary objective is to do something else (Laufer & Hulstijn, 2001, p. 10).

Hulstijn (2001) suggested that it “is the quality and frequency of the information processing activities (i.e., elaborations on aspects of a word’s form and meaning, plus rehearsal) that determine retention of new information” (p. 275). However, the number of new words learned incidentally is relatively small compared to the number of words that can be learned intentionally (Hulstijn, 1992). Even with the use of a dictionary and the inferring strategy, incidental vocabulary learning tends to be incremental and slow (Hulstijn, 1992).

Nevertheless, incidental L2 vocabulary acquisition paradigm has not been free of criticisms: for instance, Paribakht and Wesche (1999) contend that it works for much advanced vocabulary acquisition. Moreover, they are of the belief that the process of incidental vocabulary acquisition is slow, often misguided, and seemingly haphazard, producing differential outcomes for different learners, word types, and contexts.

In intentional learning, on the other hand, learners try to commit new information to memory by using strategies, such as mnemonic devices (Paradis, 1994). In other words, intentional learning is a learning vocabulary out of context by using, for instance, word lists or word cards. One body of research employing the intentional learning model is

the keyword method (see e.g. Ellis & Beaton, 1993). This technique involves the creation of a mediating word that is meant to facilitate retention of a target word by allowing the learner to develop a connection between the form and the meaning of the target word (Rukholm, 2011). The mediating word is the keyword and ideally its phonology should resemble the form of the target word while also allowing the learner to associate the target word with a visual representation of the keyword.

Furthermore, retention rates under intentional learning are on average, much higher than under incidental conditions (Hulstijn, 2003). The findings of Elgort (2010) provided evidence that deliberate learning triggered the acquisition of representational and functional aspects of vocabulary knowledge. The benefits of vocabulary-list learning are to gain not only receptive vocabulary knowledge, but also productive vocabulary knowledge as well as to increase learners' breadth and depth of vocabulary knowledge (Yamamoto, 2013). Explicit teaching results in faster vocabulary gains and a higher level of vocabulary retention than learning vocabulary through reading (Schmitt, 2008). Nation recommends "the deliberate learning of vocabulary using word cards (as one way of speeding up learners' progress towards an effective vocabulary size" (Nation 2001: 533).

The role of input and output activities

It has been shown that reading is a powerful source of vocabulary acquisition for second and foreign English language learners. Research also indicates that vocabulary knowledge contributes significantly to learners' reading comprehension (Hu & Nassaji, 2014). Moreover, several research findings (Hulstijn, 1992; Nagy, 1997; Zahar et al., 2001) supported the idea that language learners acquire second language vocabulary from reading.

Recently, Bolger and Zapata (2011) hypothesized that L2 learners' processing of context and completion of reading comprehension tasks would trigger deeper processing than merely lists of words. In this study, the use of context guided by the need to reflect this importance and common pedagogical practices (e.g., the communicative approach) but not by the debate on its value as a pedagogical tool for L2 learning.

Additionally, glossing has been argued to help vocabulary learning and assist reading comprehension (Ko, 2012). A number of studies have provided evidence that glosses are effective in helping learners learn new lexical items in a second language (Bowles, 2004; Cheng & Good, 2009), for example, the results of (Ko, 2012) indicated that glossing had a positive effect on L2 vocabulary learning. Additionally, Zhang (2007) showed that in terms of vocabulary gains, the provision of marginal glosses was the more beneficial than the availability of dictionary and non-dictionary use. The results also demonstrated that there would be a significant difference between gloss and no-gloss groups with respect to gaining word meaning.

Research indicates that lexical inferencing, or guessing the meaning of an unfamiliar word, is the main strategy learners use in initial comprehension of unfamiliar words while reading (Paribakht, 2005; Paribakht & Wesche, 1999). A word with a derived meaning is more likely to be retained in an L2 lexical system than a word with a glossed meaning (Nation, 2001).

Much research has focused on how to enhance the effectiveness of incidental vocabulary learning in reading by using stimulus techniques such as output tasks, textual glosses, and think-aloud activities (Min, 2008; Rott, 2004; Watanabe, 1997). On the contrary, research suggests that learning words from context while focusing on reading is an inefficient method because of the limitations inherent in deriving meanings from contextual cues (Nagy, 1997; Nation, 2001).

Meta-analyses on L2 vocabulary teaching

Several meta-analyses have been conducted on some aspects of L2 vocabulary teaching. For example, Chiu (2013) investigated the general effectiveness of L2 computer-assisted vocabulary instruction, with analysis of the features of treatment duration, educational level, and the use of games and the role of teachers in the CALL studies. In general, computer-assisted language learning in L2 vocabulary was shown to have positive effects with a medium effect size ($d = 0.745$, $p = 0.000$). The results of Abraham's meta-analysis (2008) showed that computer-mediated glosses had an overall medium effect on second language reading comprehension and a large effect on incidental vocabulary learning. Huang (2010) conducted a systematic statistical synthesis of the effects of output stimulus tasks on L2 incidental vocabulary learning. A total of 12 studies were included in this meta-analysis. Results showed that language learners gained more benefit from using output stimulus tasks to learn vocabulary than those who only read a text. For these 16 studies, the mean effect size was 1.39 (SE = .07).

Although the meta-analyses on L2 vocabulary teaching have highly contributed to the field of instructed L2 vocabulary learning, the effectiveness of receptive L2 vocabulary learning remains a relatively under-researched line of inquiry in the literature. Additionally, a number of contextual factors and moderator variables have rarely been investigated.

Recently, meta-analysis has been described more broadly as a research synthesis method with the aim of estimating an average association across studies and to explore the degree and sources of heterogeneity (Sutton & Higgs, 2008). Additionally, one of the most frequently cited reasons for conducting a meta-analysis are the increase in statistical power that it bestows a reviewer (Cohen & Becker, 2003; Card, 2012).

Admittedly, one of the problems that associated with conducting meta-analyses is the publication bias (Borenstein, et al. 2009; Card, 2012; Sutton & Higgs, 2008). Meta-analysis it is not without its critics particularly because of the difficulties of knowing which studies should be included and to which population final results actually apply (Sutton et al. 2000; Sutton & Higgs, 2008). If the included studies are a biased sample of all related studies, then the mean effect computed by the meta-analysis will reflect this bias (Borenstein, et al., 2010). Publication status cannot be used as a criterion for quality; and should not be used as a basis for inclusion or exclusion of studies (Borenstein, et al. 2009).

One way to reduce the possible influence of publication bias is to include doctoral dissertations in a research synthesis. As, Light and Pillemer (1984, p. 38) point out, dissertations have several advantages in that they are required to be approved by faculty, thereby enhancing quality, they often contain more detailed quantitative information than journals, and they also can provide more qualitative information about the

research. This study utilized a meta-analysis methodology to combine the quantitative results of primary studies identified in the existing research literature.

Purpose of the study

The primary purpose of the present study is to investigate the overall effectiveness of L2 vocabulary instruction. Second, it aims to assess the potential heterogeneity across effect size measures. Third, the study attempts to evaluate the moderator variables such as context of instruction, publication type, the age of the participants, and the L2 learners' proficiency level on the L2 vocabulary learning, type of technology, word type.

Research questions

The current meta-analysis is aimed to address the following research questions:

1. What is the overall effect of variables contributing to SLA vocabulary acquisition?
2. To what extent the effect sizes vary across studies?
3. What moderator variables affect the overall effectiveness of L2 vocabulary instruction?

Methodology

Literature search

For the purpose of data collection, documents were accessed electronically through *Web of Science*, *Academic Search Premier* and *Pro Quest Dissertations and theses* databases. Then, Oxford Journals, Cambridge Journals, Sage Journals, and Taylor & Francis Journals were subject to online search using the same search terms.

The second phase of study identification and retrieval stage of a meta-analytic review included: searching key applied linguistics and SLA journals, *Applied Linguistics*, *Language Awareness*, *Language Learning*, *Language Teaching Research*, *Modern Language Journal*, *RELC Journal*, *Second Language Research*, *Studies in Second Language Acquisition*, *System*, *TESOL Quarterly*.

Search terms

To retrieve the articles and dissertations, a set of search terms and combination of them were employed; *Foreign language vocabulary learning/ acquisition*, *L2 vocabulary acquisition*, *L2 vocabulary learning*, *second language vocabulary learning/ acquisition*, *L2 vocabulary knowledge*, *foreign language vocabulary knowledge*, *L2 lexical proficiency*, *second language vocabulary development*, *L2 vocabulary development*, *second language instruction*, *L2 vocabulary gain*, *L2 vocabulary retention*.

Inclusion criteria

The criteria stipulated for the inclusion of the studies for the current meta-analysis were as follows;

- 1) Dependent variable, in this meta-analysis, is second or foreign language vocabulary acquisition.

- 2) Studies included for the current meta-analysis should be experimental or quasi-experimental. Studies included in the statistical analysis, must utilize an experimental design, quasi-experimental design, or pre-post design.
- 3) Eligible studies have interventions or treatments. So, the correlational studies were excluded.
- 4) Eligible studies must report sufficient statistical and descriptive data for inclusion in the analysis.
- 5) The current meta-analysis included both published and unpublished studies. Among unpublished studies, doctoral dissertations will be included in the current meta-analysis to the exclusions of the proceedings of the conferences.
- 6) To take account for the latest development in the field of L2 vocabulary instruction, the studies should be published between 2004 and May 2014. Thus, studies published before 2004 were excluded from the present meta-analysis.
- 7) this study concentrated on the acquisition of “receptive vocabularies”. So “productive words” was excluded from current meta-analysis.

Exclusion criteria

The criteria for exclusion of papers or dissertations are as follows:

- 1) The study did not examine L2 vocabulary learning, development or retention. For example, the study may have examined learners' perception of L2 vocabulary learning strategies.
- 2) The study was a literature review, synthesis, or meta-analysis.
- 3) Studies on L2 vocabulary learning of people with language impairment were excluded.

Coding the studies

The primary investigator screened all articles for inclusion. To promote consistency in the screening process, a minimum of 50% of the studies were double-screened by a trained graduate research assistant. All articles selected for inclusion were coded and rated by the primary investigator and a graduate research assistant. The outcome of the coding was compared and any discrepancies resolved through discussion. The graduate assistant and the lead author coded 8 randomly selected studies and intercoder reliability was calculated through Cohen's Kappa (k) coefficient. The agreement rate was 98.5% and the differences were resolved through discussion. Coding measurement procedures and research settings would enable the reviewer to assess whether *effect size* estimates had been affected by the choice of instrument or the location of the study (Ellis, 2010).

After identifying the body of research literature that meets the stipulated inclusion and exclusion criteria, a coding scheme was developed to classify common characteristics of the studies. Final comprehensive coding scheme was included two major categories for methodological features: 1) learner characteristics and 2) research design. Studies were coded for the number of participants, age of the participants, publication type, types of the target words, length of instruction, the technology used, context of L2 study, and the proficiency level of the participants. For the present meta-analysis, the coding scheme was constructed by reviewing previously published meta-analyses and based on the research questions that guided the present study.

Random –effects vs. fixed effects model

Borenstein et al. (2010) pointed out that the selection of the model is critically important. In addition to affecting the computations, the model helps us to define the goals of the analysis and the interpretation of the statistics. In the same way, Lau et al. (1992) recommend using random-effects (RE) analyses rather than fixed-effects (FE) analyses because RE analyses yield wider confidence intervals around the weighted average effect size, thereby reducing the likelihood of committing a Type I error. Perhaps most importantly, RE analyses may permit generalizations that extend beyond the studies included in a review, whereas FE analyses are more restrictive and only permit inferences about estimated parameters (Cohen & Becker, 2003). Likewise, Borenstein, et al. (2009) pointed out that under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate.

Calculation and interpretation of the effect sizes

All the analyses (including *effect size* measures) were run by using professional meta-analysis software called Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). Hunter and Schmidt (2004) believe that this software is all-purpose meta-analysis program. There are different ways of interpreting the effect size measures. The most commonly used one is Cohen (1998) benchmark in that he suggested the following guidelines for designating effects as small, medium, and large: $d = .20$ or $r = .10$ is considered a small effect size, $d = .50$ or $r = .30$ is a medium effect size, and $d = .80$ or $r = .50$ is a large effect size. “The larger this value, the greater the extent to which the phenomenon under study is manifested” (Cohen, 1988, p. 10). recently, however, Oswald and Plonsky, (2010) suggested a more field-sensitive criterion for SLA research. For mean differences between groups, d values in the neighborhood of .40 should be considered small, .70 medium, and 1.00 large. These estimates of (roughly) small, medium, and large effects were chosen based on their approximate correspondence to the 25th, 50th, and 75th percentiles, respectively, for between-group contrasts in primary and meta-analytic research (Plonsky & Oswald, 2014). The present study interprets the findings based on the latter one.

Results

Approximately 2322 articles and PhD dissertations that have been published or not published between 2004 and 2014 were retrieved through first filtering. Eighty-two of these documents were selected through second filtering. Finally, 16 published articles and Ph.D. dissertations met the inclusion criteria and were included in this meta-analysis. All studies investigated the effects of different factors and variables on the acquisition of L2 receptive vocabulary. Nine of these documents were PhD dissertations and 7 were published papers. The principle of “one study, one effect size” was followed as much as possible to minimize the presence of sample size inflation and nonindependence of events. Only group contrasts, control vs. experimental groups, were gained and analyzed. Table 1 shows all the studies as well as the included studies.

Table 1 Descriptive effect size statistics

K	N	Effect Size (d)	SE sm	Z-Value	95% CI	Q	I^2
16	1008	0.80	0.14	5.60	0.52 to 1.08	59.94	74.94

Descriptive data

In order to address the overall effectiveness of L2 vocabulary instruction, the random-effects effect size, Cohen’s d, of the effects of the treatments on L2 vocabulary instruction was examined. Figure 1. demonstrates forest plot of standardized mean effect for overall L2 vocabulary instruction.

Heterogeneity of effect sizes

The second research question asked, “To what extent the effect sizes varied across studies?” The Q test of homogeneity of effect size was conducted based on the random-effects model of meta-analysis. It indicated that the null hypothesis should be rejected, $Q(16) = 59.94, p < .01$, finding that effect sizes varied significantly across studies. The tau-squared (T^2) refers to the estimation of the variance of effect sizes, $T^2 = 0.23$. It indicated sizable variation in parameter effect sizes. The I^2 statistic (Higgins et al. 2003) was 74.97 which indicate that a high proportion of the between-effect size variance reflects real differences in effect sizes. Thus, the answer to the second research question is that there is sizable variation of effect sizes across studies. Table 2 demonstrates the Cohen ‘s d, upper limit and lower limit.

Publication bias

If publication bias were present, the bottom of the funnel plot would show a higher concentration of studies on one side of the mean than the other. This type of distribution would reflect the tendency for smaller studies with larger than average effect sizes, making them more likely to achieve statistical significance, to be published (Borenstein et al., 2009).

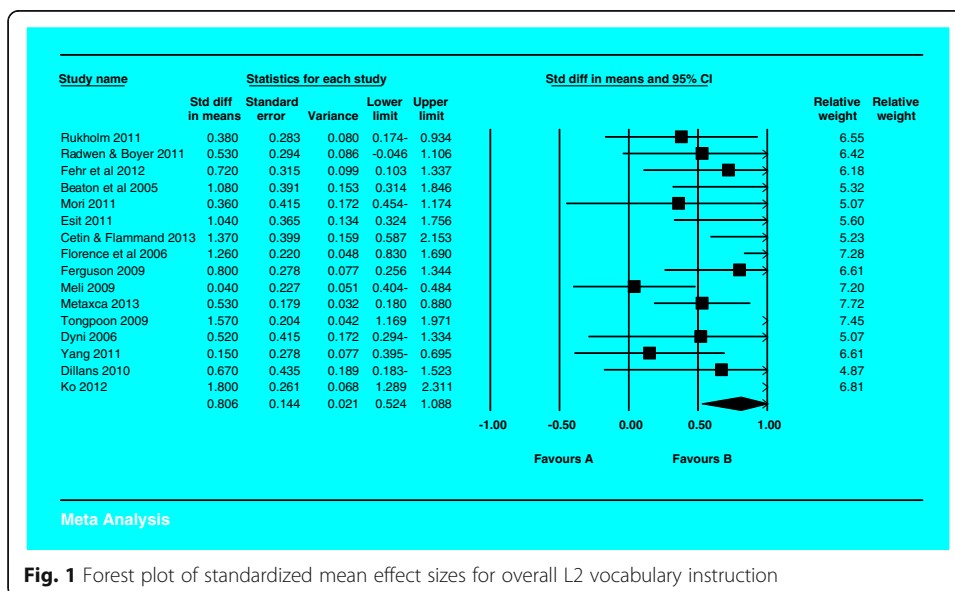


Fig. 1 Forest plot of standardized mean effect sizes for overall L2 vocabulary instruction

Table 2 the Cohen’s *d*, upper limit and lower limit

Study Name	Cohen’s <i>d</i>	Standard error	Variance	Lower limit	Upper limit
Rukholm 2011	0.38	0.283	0.80	0.174	0.934
Radwen & Boyer 2011	0.53	0.294	0.086	- 0046	1.106
Fehr et al. 2012	0.72	0.315	0.099	0.103	1.337
Beaten et al. 2005	1.08	0.391	0.153	0.314	1.846
Mori 2011	0.36	0.415	0.172	0.454	1.174
Esit 2011	1.04	0.365	0.134	0.324	1.756
Cetin & Flamand 2012	1.37	0.399	0.159	0.587	2.153
Florence et al. 2006	1.26	0.220	0.048	0.830	1.690
Ferguson 2009	0.80	0.278	0.077	0.256	1.344
Meli 2009	0.04	0.227	0.051	0.404	0.484
Metaxca 2013	0.53	0.179	0.032	0.180	0.880
Tongpoon 2009	1.57	0.204	0.042	1.69	1.971
Dyni 2006	0.52	0.415	0.172	0.294	1.334
Yang 2011	0.15	0.278	0.077	0.395	0.695
Dilans 2010	0.67	0.435	0.189	0.183	1.523
Ko 2012	1.80	0.261	0.68	1.289	2.311
Random-Effect	0.80	0.144	0.021	0.524	1.088

Funnel Plot (Light & Pillemer, 1984) is one of the approaches to display the relationship between effect size and study size and illustrate potential evidence of publication bias. When publication bias is not present, the studies should be distributed symmetrically around the average effect size because of random sampling error. Large studies cluster around the mean effect size on the top and smaller studies spread across wider range near the bottom.

Figure 2 demonstrates that the majority of effect sizes were equally distributed around the mean, indicating the absence of publication bias. Studies with larger sample sizes appear towards the upper portion of the funnel and are relatively evenly

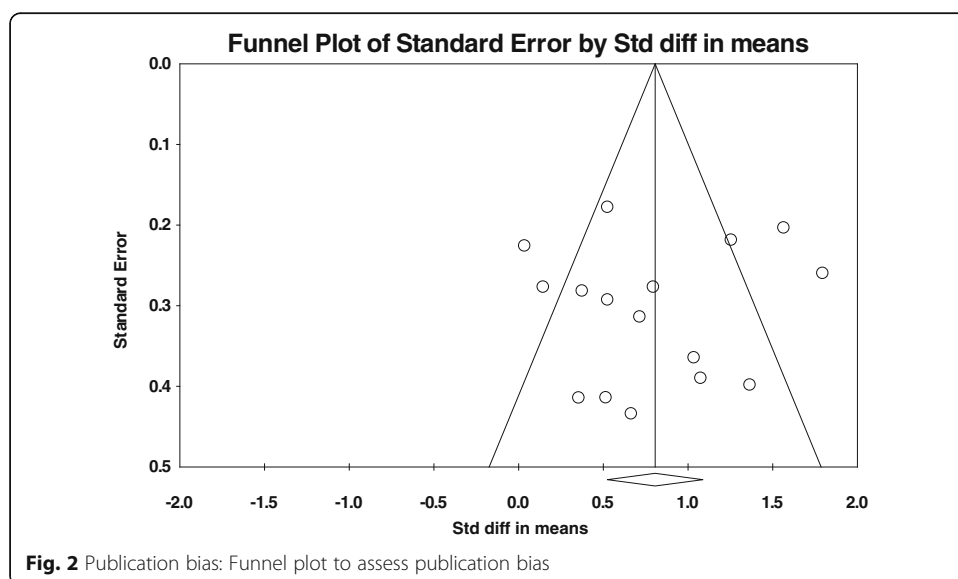


Fig. 2 Publication bias: Funnel plot to assess publication bias

distributed about the mean, with the graph indicating that medium and larger scale studies with medium effect sizes were well represented. Additionally, to address the ‘file-drawer problem’ that is characteristic of meta-analysis, Rosenthal’s (1979), Fail-Safe N test was conducted (using CMA software). The test showed $N = 1,600,000$, $z = 11.25464$, $p < 0.00000$. This statistic indicated that 1,600,000 studies would need to be added to the analysis to yield a statistically non-significant result that is a large Fail-safe.

Moderator variable analysis

Table 3 delineates the characteristics of the moderator variables of the primary studies. Table 4 shows the Moderator analysis: Means and Q-statistics for group contrasts of the study.

The context of L2 vocabulary instruction

Research setting can be divided into foreign language (FL) and second language (SL). A foreign language setting is one where the learner studies a language that is not the primary language of the linguistic community. A second language setting, on the other hand, is one in which the learner’s target language is the primary language of the linguistic community. A small to medium effect ($d = 0.53$) for Second language contexts

Table 3 Characteristics of the moderator variables of the primary studies

Study Name	No of participants	Context	Age Range	Proficiency level	Technology Type	Types of Target words	Publication type
Rukholm 2011	66	SL	adult	elementary	Authentic Songs	Abstract	dissertation
Radwen & Boyer 2011	48	FL	adult	Intermediate	Reading Task	Abstract	paper
Fehr et al.,2012	43	SL	Child	Elementary	CALL	NA	paper
Beaten et al. 2005	30	SL	adult	Intermediate	CALL	Concrete	paper
Mori 2011	30	FL	adult	Intermediate	Authentic songs	Abstract	dissertation
Esit 2011	33	FL	adult	Intermediate	CALL	NA	paper
Cetin & Flamand 2012	31	FL	Child	Elementary	Poster	Concrete	paper
Florence et al., 2006	100	FL	adult	Intermediate	CALL	NA	paper
Ferguson 2009	56	SL	adult	Intermediate	CALL	Abstract	dissertation
Meli 2009	93	SL	adult	Elementary	CALL	NA	dissertation
Metaxca 2013	130	EFL	adult	Advanced	Authentic songs	Abstract	dissertation
Tongpoon 2009	160	FL	adult	Intermediate	Reading Task	Abstract	dissertation
Dyni 2006	24	SL	child	Elementary	Reading Task	concrete	dissertation
Yang 2011	160	FL	adult	Intermediate	CALL	concrete	dissertation
Dilans 2010	23	SL	adult	Intermediate	Reading Task	concrete	dissertation
Ko 2012	90	FL	adult	Intermediate	Reading Task	Abstract	paper

Table 4 Moderator analysis: Means and Q-statistics for group contrasts

Moderator	K	<i>d</i>	95% CI	(Low, High)	Q
Context					3.026
FL	9	0.96	0.57	1.36	
SL	7	0.53	0.53	0.81	
Age Group					0.04
> 12	13	0.79	0.46	1.11	
< 12	3	0.85	0.38	1.32	
Publication Type					4.75
Published	7	1.12	0.78	1.46	
Dissertations	9	1.12	0.20	0.93	
Proficiency Level					2.23
Elementary	4	0.61	0.04	1.18	
Intermediate	11	0.61	0.56	1.24	
Advanced	1	0.53	0.18	0.88	
Word Type					0.24
Abstract	7	0.92	0.73	1.10	
Concrete	5	0.65	0.33	0.97	
NA	4	0.72	0.46	0.98	
Technology Type					7.05
CALL	7	0.68	0.47	0.88	
CALL	1	1.37	0.58	2.15	
Reading	5	1.25	1.00	1.50	
Song	3	0.47	0.19	0.75	

and large effect for foreign language settings ($d = 0.96$) were obtained. 9 and 7 studies were conducted in foreign language and second language contexts, respectively. The difference between foreign language and second language contexts was not statistically significant ($Q = 3.02$, $df = 1$, $P = 0.08$).

The age of the participants

Following Jeon and Yamashita (2014), All participants who were at or below grade six (or age 12) were coded as Child and the participants who were at or older than grade seven (13 or older) were coded as Adult .we sought to account for variation in effect size measures by investigating the influence of the age of the participants in the primary studies. As shown in Table 2, ($d = 0.79$) was observed for adult and ($d = 0.85$) was found for child participants. However, the differences are not statistically significant ($Q = 0.47$, $df = 1$, $p = 0.82$).

L2 learners' proficiency level

The third moderator variable of the current meta-analysis was the impact of the participants' proficiency level on the overall effect size. To estimate it, three levels of L2 proficiency levels were coded in the included studies (elementary, intermediate, and advanced). Ten primary studies were conducted targeting intermediate l2 learners and 5 studies included participants in elementary level of L2 proficiency. Only one study

was done with advanced L2 learners. With respect to L2 proficiency level, small *effect size* ($d = 0.53$) was obtained for both advanced and elementary levels ($d = 0.54$). However, large effect size ($d = 0.95$) was gained for intermediate L2 learners. However, the difference between three groups was not statistically significant ($Q = 3.46$, $df = 2$, $P = 0.17$).

Publication type

To account for the variation in *effect sizes*, another moderator factor, publication type, was examined. 7 published and 9 Ph.D. dissertations were included in the present meta-analysis. Published articles generated effect size of ($d = 1.12$), whereas, Ph.D. dissertations produced the effect size of ($d = 0.57$). The difference is statistically significant ($Q = 4.75$, $df = 1$, $p = 0.02$).

Word type

In order to examine the variation in *effect size*, another moderator variable, word type, was analyzed. This variable included; abstract words, and concrete words. Since some studies did not report type of the target words in the studies, another category labeled *not mentioned*. The effect size observed for abstract words was ($d = 0.92$) whereas, concrete words generated the effect size of ($d = 0.65$). Statistically speaking, the difference is not significant ($Q = 0.24$, $df = 2$, $p = 0.88$).

Technology (technique) type

Four types of technology (technique) were classified in the included studies; Computer-assisted Language learning (CALL), poster, reading, and song. Applying "poster" generated the largest effect size ($d = 1.37$, $k = 1$). Employing reading activities to teach target words produced ($d = 1.25$, $k = 5$). CALL technology produced the *effect size* of ($d = 0.68$, $k = 7$). The smallest effect size was gained for studies that employed song to teach the target words ($d = 0.47$, $k = 0.47$). The differences, however, are not statistically significant ($Q = 7.05$, $df = 3$, $p = 0.07$).

General discussion

This meta-analysis sought to determine the effectiveness of L2 vocabulary instruction and to identify the moderator variables for its effectiveness. The overall effect size for L2 vocabulary instruction was ($d = 0.80$). Based on Oswald and Plonsky (2010) criterion, this effect size is medium to large. The findings indicate that L2 vocabulary instruction is an effective instructional approach for improving L2 proficiency and should be incorporated as an integral part of L2 syllabus. The results of the present meta-analysis should be discussed considering other similar meta-analyses. As Plonsky and Oswald (2014) suggested that meta-analysts can look to the results of other meta-analyses when explaining their finding. Chiu (2013) investigated the general effectiveness of L2 computer-assisted vocabulary instruction, with analysis of the features of treatment duration, educational level, and the use of games and the role of teachers in the CALL studies. In general, computer-assisted language learning in L2 vocabulary was shown to have positive effects with a medium effect size ($d = 0.745$, $p = 0.000$). The results of Abraham's meta-analysis (2008) showed that computer-mediated glosses had

an overall medium effect on second language reading comprehension and a large effect on incidental vocabulary learning. Huang (2010) conducted a systematic statistical synthesis of the effects of output stimulus tasks on L2 incidental vocabulary learning. A total of 12 studies were included in this meta-analysis. Results showed that language learners gained more benefit from using output stimulus tasks to learn vocabulary than those who only read a text. For these 16 studies, the mean effect size was 1.39 ($SE = .07$).

Context

The mean *effect size* associated with the studies conducted in FL contexts was larger than those conducted in SL contexts, indicating that L2 vocabulary instruction was more effective in FL contexts than in SL ones ($d = 0.96$ vs. $d = 0.53$). This finding is similar to other studies. For example, Cobb (2010) meta-analysis of task-based interaction found a strong advantage for studies carried out in foreign-language settings ($d = 0.89$ vs. 0.14 in L2 settings). Likewise, Li (2010) found larger effect for studies conducted in foreign language contexts than for studies conducted in second language contexts. Li (2010) attributes this difference to the instructional dynamics of FL contexts. We believe that one explanation is that teachers in FL contexts mainly tend to teach lexical items and grammatical structures whereas teachers in SL contexts might concentrate on the overall communication. We also hypothesize that language learners in foreign language contexts presumably have different objectives in language teaching. One of the reasons behind the difference of *effect size* across different contexts can be “language teaching system orientation” (Yousefi & Biria, 2011, P.14). In addition, Liu (2007) surveyed 800 teachers of English throughout the world and found that EFL teachers tended to focus more on linguistic forms than ESL teachers. Likewise, Won (2008) suggested that ESL and EFL classroom teachers need to consider the differences of first and second language vocabulary acquisition as well as student learner characteristics.

Publication bias

With respect to the effect of publication type on the variation among primary studies, it was indicated that the published studies generated more effect size than PhD dissertations and the difference was statistically significant. This finding highlights one of the big threats and concerns about conducting meta-analyses. It also confirms the fact that studies with larger effect sizes give their ways to the publication more easily than those with smaller effect size and non-significant ones. We propose that in order to reduce publication bias, it is up to meta-analysts that include both published and unpublished studies including doctoral dissertations, conference proceedings, and working papers. We also believe that L2 researchers should report the effect size in their primary studies and larger effect size should not be interpreted as contributing to the field more than small effect size measures. In order to advance our understanding of SLA processes, the researchers should report the perceived phenomenon and justify the findings in the light of the current theories and hypotheses.

Similarly, Plonsky and Oswald (2014) believe that there is growing evidence of publication bias among L2 meta-analyses that have investigated this issue. Lee and Huang (2008) grouped and compared the effects of textual enhancement among (a) published results (not based on a dissertation; $d = .55$, $k = 8$), (b) published results based on a dissertation

($d = .24$, $k = 4$), and (c) unpublished dissertation results ($d = -.01$, $k = 4$). In Li (2010) study, Published studies did not show a larger effect than PhD dissertations; in fact, the mean effect size for dissertations was larger than that yielded by published articles.

Proficiency level

The effect size that was obtained for intermediate learners was larger than elementary and advanced learners. This finding should be interpreted with caution. Since only one study has included the advanced learners. The larger effect size of intermediate participants can be attributed to the fact that they have already achieved a threshold level of L2 vocabulary. Intermediate learners also attained L reading strategies that enable them to benefit much from reading activities.

In Yun (2011) Learner proficiency was found a statistically significant moderator to affect the treatment effects with $Q = 15.304$, $p < 0.05$; that is, studies with beginning learners had the largest mean effect size, 0.698 while those with intermediate learners had the least mean effect size, 0.233. That is, beginning learners who had access to multiple hypertext glosses most benefited from multiple glosses in reading. Abraham (2008) believes that Intermediate learners may possess deeper lexical knowledge allowing them to connect vocabulary encountered in the glosses more easily to a pre-existing semantic system and network of L2 vocabulary than beginners who are still developing their vocabulary base. The results of Huang's (2010) meta-analysis showed that the vocabulary learning of language learners with low proficiency levels and vocabulary sizes may benefit more from L1 textual glosses than those who have higher proficiency levels and larger vocabulary sizes.

Li (2010) did not include proficiency measure as one the moderator analyses due to the high degree of heterogeneity in primary researchers' use of proficiency measures. The researcher believes that the primary researchers' decisions on the proficiency levels of participants were arbitrary and highly context-specific. Chiu (2013) showed that high school or college students ($d = 1.032$, $p = 0.001$) can benefit more from computer-assisted language learning program than elementary school students ($d = 0.321$, $p = 0.004$). Learners would have different learning styles and strategies. This may be due to the maturity level of high school or college students enabling more effective use of technology for English vocabulary learning. In the same vein, in Yun's (2011) study, Learner proficiency was found to be a statistically significant moderator to affect the treatment effects with $Q = 15.304$, $p < 0.05$: studies with beginning learners had the largest mean effect size, 0.698 while those with intermediate learners had the least mean effect size 0.233.

Age of the participants

Following Jeon and Yamashita (2014), all participants who were at or below grade six (or age 12) were coded as Child and the participants who were at or older than grade seven (13 or older) were coded as Adult. The present meta-analysis revealed that the effect size observed for child learners was larger than adult participants in the primary studies ($d = 0.85$ vs. 0.79). However, the difference statistically speaking, however, is not significant. With this in mind, this finding should be interpreted with caution.

The results of Nakanishi (2015) suggest that the effect of extensive reading might increase with older participants. The researcher attributes the reason to the beneficial for older learners who have learned the foreign language explicitly, as it might lead them to draw on and proceduralize their explicit knowledge. Nakanishi (2014) goes on to argue that another factor concerns the maturity of the participants in terms of their cognitive processing. As individuals age, they are able to understand and process more complex information, a development that could lead them to read more.

The influence of the age at which words are acquired on various measures of lexical processing was acknowledged (Balota et al., 2006). There have been a number of reports suggesting that age of acquisition produces a unique influence on word recognition performance above and beyond correlated variables such as word frequency Balota et al., (2006) believe that the intriguing argument here is that early acquired words could play a special role in laying down the initial orthographic, phonological, and/or semantic representations that the rest of the lexicon is built upon. Moreover, early acquired words will also have a much larger cumulative frequency of exposure across the lifetime.

Simply put, from the perspective of information processing theory, differences in problem-solving abilities have been identified as one of the main explanations for the difference between second language learning by younger and older learners (Munoz, 2006). With biological maturation, aspects such as rate of information processing increase regularly from childhood to adulthood.

Word type

Contrary to the research findings so far, the findings indicate that abstract words generated higher effect size than concrete words. We believe that one justification may be the fact that abstract words do not make extra cognitive processing demands on adult Language learners. However, it might be more demanding for young children to acquire abstract words than concrete ones. In first language (L1) acquisition, concrete words (e.g., table, paper) are typically learned prior to abstract words (e.g., liberty, myth) (Schwanenflugel, Akin, & Luh, 1992). Schwanenflugel et al. (1992) noted that the advantages demonstrated by concrete words may stem from the fact that concrete words have greater 'context-availability' than abstract words. It is typically easier to think of a context in which concrete words appear than it is to think of a context in which a given abstract word appears.

Technology (technique) type

The last moderator variable of the present study was the effect of technology (technique) used for the purpose of teaching L2 vocabulary items. It was revealed that employing "posters" generated the highest effect size following by "reading activities and tasks". CALL technology produced the third highest effect size. While using authentic songs for the purpose of L2 vocabulary teaching generated the smallest effect size.

This finding should be interpreted with caution since only one study has employed "posters" for the purpose of teaching L2 vocabulary items (Cetin &

Flammand, 2012). Cetin and Flammand (2012) believe that using poster in the classroom provide support for the usefulness of the concept of self-directed inferential learning, raise students' awareness, arouse their interest, and will allow them to take an interest in their own surroundings. The fact that "reading activities and tasks" generated higher effect size than CALL technology should be verified by more longitudinal studies. We believe that there is much room for manipulation of reading tasks on the part of language teachers and paving the way for input enhancement and making the target words more salient. As Fehr, et al. (2012) argued it is unrealistic to suggest that computer-delivered vocabulary instruction can be the sole vehicle for remediation of significant vocabulary deficits or L2 vocabulary learning. One possible explanation for this finding is that students welcome a higher degree of autonomy in their learning and they tend to be in control of their own learning when learning from vocabulary web sites with games (Yip & Kwan, 2006). Yip and Kwan (2006) suggested that sophisticated experiential games, such as simulated tasks, are needed, as they are more interactive and collaborative and can address cognitive issues and foster active learning. We propose that language teachers should incorporate CALL as well as reading activities and tasks into their syllabi to meet learners' ongoing needs and expectations.

Suggestions for further studies

The findings of this study have practical implications for educators, Language teachers, and other scholars that advance our understanding of the mechanisms responsible for the most effective techniques of L2 vocabulary teaching. Research must try to establish what variations in participants, as well as in treatments, will provide the most benefit for most L2 learners. This meta-analysis highlighted important gaps in the following areas of research: first, the effects of the context of L2 vocabulary instruction on the acquisition and retention of the target words. Second, the modifying effects of background knowledge, L1 and L2 distance, type of different tests and tasks, different ways of operationalizing vocabulary learning and retention, duration of instruction. Future work aimed at understanding the interplay among language- learner related factors and language learning connected variables can illuminate our understanding of the mechanisms underlying L2 vocabulary learning and account for cost-effective L2 vocabulary learning model. We propose that different word types (concrete, abstract, emotion, and pseudo word) may be acquired differently. As Altarriba and Basnight-Brown (2011) suggest that the three word types – concrete, abstract, and emotion – were not acquired in the same way, even though the same basic mode of acquisition was used to teach these words in a new language.

Future research should examine other potential moderators, including setting (e.g., instructed vs. naturalistic setting), instructional variables (e.g., instructional tasks and activities), teacher orientation (e.g., beliefs and attitudes), and L2 learner variables (e.g., type of motivation, cognitive style, and learning strategies) that may influence the effectiveness of L2 vocabulary instruction.

We recommend that meta-analysts include PhD dissertations in their syntheses. By so doing, researchers will reduce publication bias and gain access to rich descriptions of the research procedures. In addition, by including doctoral dissertations,

meta-analysts will gain access to rich data that would be able them to analyze more moderating variables that otherwise will go untouched.

Limitations

This review was intentionally limited to experimental-control studies. The strict inclusion criterion led to the relatively small number of included studies. Although the inclusion of studies with within-subject designs utilizing pre-post comparisons may contribute significantly to our understanding, the effect size statistics for these types of studies may add to the inflation of effect sizes when pooled with studies utilizing a separate control group. There are several issues that pose limitations and warrant consideration when evaluating the results of this study. Due to the relatively small number of studies, care should be exercised as to the generalization of its findings. Many of the included studies have employed relatively short duration of instruction. In order to grasp a total picture and construct an integrative model of L2 vocabulary learning more and more longitudinal studies should be conducted and analyzed through meta-analyses and Structural Equation Modeling (SEM).

Conclusion

The overall effectiveness of L2 vocabulary instruction gained through the present meta-analysis was ($d = 0.80$) which means that L2 vocabulary treatment programs have the effect size of medium to large. The research synthesis indicates that L2 vocabulary instruction was effective and given the significance of vocabulary, L2 vocabulary teaching should be incorporated as indispensable part of L2 syllabus. What remains unresolved, here, is the question of what factors and variables enhance L2 vocabulary development more effectively than the other variables. To gain such an insight, we call for constructing L2 vocabulary models and hypotheses that provide syllabus designers and language teachers with cost-effective techniques of teaching L2 vocabulary items. We are sure that this can be achieved through application of sophisticated statistical analyses and capitalizing on the development in the field of SLA.

Acknowledgements

We would like to appreciate two anonymous reviewers of the journal for their invaluable comments.

Funding

No funding was available to the present study.

Availability of data and materials

All authors must include an "Availability of Data and Materials" section in their manuscript detailing where the data supporting their findings can be found.

Authors' contributions

MHY conducted the meta-analysis and reported the results. RB coded the included studies and constructed the coding sheet. MHY discussed the findings in the light of other meta-analyses. RB edited the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of English, Islamic Azad University, Bonab Branch, Bonab, Iran. ²Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.

Received: 9 November 2017 Accepted: 16 November 2018

Published online: 22 December 2018

References

One asterisk indicates that the study was included in the meta-analysis.

- Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 21(3), 199–226.
- Altarriba, J., & Basnight-brown, D. M. (2011). The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism*, 16(4), 446–452.
- Balota, D., A., Yap, M. J., & Cortes, M. J. (2006). In M. J. Traxler & M., A. Gernsbacher (Eds.), *Visual word recognition: The journey from features to meaning (a travel update)*. *Handbook of psycholinguistics*. Elsevier.
- Beaton, A. A., Gruneberg, M. M., Hyde, C., Shufflebottom, A., & Skyes, R. N. (2005). Facilitation of receptive and productive foreign vocabulary learning using keyword method; the role of image quality. *Memory*, 13(5), 458–471.
- Bolger, D., & Zapata, G. (2011). Semantic categories and context in L2vocabulary learning. *Language Learning*, 61(2), 614–646.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis (Version 2.2.027)* [Computer software]. Englewood, NJ: Biosta.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Computing Effect Sizes for Meta-analysis*. Chichester, UK: John Wiley & Sons, Ltd.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed effect & random effects models for meta-analysis. *Research Synthesis methods*, 1, 97–111.
- Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania*, 87, 541–552.
- Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Publications, Inc..
- *Cetin, Y. and Flammand, L. (2012). Posters, self-directed learning, and L2 vocabulary acquisition. *ELT*, 67(1), 52–71.
- Cheng, Y., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, 21, 119–142.
- Chiu, Y. H. (2013). Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology*, 44(2), E52–E56.
- Cobb, M. (2010). *Meta-analysis of the effectiveness of task-based interaction in form-focused instruction of adult learners in foreign and second Language Teaching* Unpublished doctoral dissertation, University of San Francisco.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. San Diego, CA: Academic Press.
- Cohen, L., & Becker, B. J. (2003). How Meta-Analysis Increases Statistical Power. *Psychological Methods*, 8(3), 243–253.
- Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using Hypernymic relationships. *Language Learning*, 59(2), 307–334.
- Dilans, G. (2010). Oral corrective feedback and L2 vocabulary development: Prompts and recasts in the adult ESL classroom. Unpublished doctoral. dissertation. The University of Texas at San Antonio.
- *Dyni, L. (2006). *Promoting vocabulary acquisition among grade one and two ESL students with word explanation and repeated reading using audiotaped books*. Unpublished doctoral dissertation. University of Toronto.
- Elgort, I. (2010). Deliberate learning and vocabulary Acquisition in a Second Language. *Language Learning*, 61(2), 367–413.
- Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary. *Language Learning*, 43, 559–617.
- Ellis, P. (2010). *The essential guide to effect sizes. Statistical power, Meta- analysis and the interpretation of research result*. Cambridge: Cambridge University Press.
- Esit, O. (2011). Your verbal zone: An intelligent computer-assisted language learning program in support of Turkish learners' vocabulary learning. *Computer Assisted Language Learning*, 24(3), 211–232.
- Fehr, C. N., Davison, M. L., Graves, F. M., Sales, C. G., Seipel, B., & Sharma, S. S. (2012). The effects of individualized, online vocabulary instruction on picture vocabulary scores: An efficacy study. *Computer Assisted Language Learning*, 25(1), 82–102.
- *Ferguson, J.L. (2009). *Explicit second language vocabulary learning: an investigation of a gloss- embedded text plus form, meaning, and use exercises*. Unpublished doctoral dissertation. The Pennsylvania State University.
- Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, 58(1), 1–31.
- Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta- analysis. *British Medical Journal*, 327, 557–560.
- Hu, H. M., & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferencers. *System*, 45, 27–34.
- Huang, S. F. (2010). *Effects of tasks and glosses on L2 incidental vocabulary learning: Meta- analysis, unpublished doctoral dissertation*. USA: Texas A&M University.
- Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Bejoing (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.
- Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: Areappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.
- Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis*. London: SAGE Publications.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64, 160–212.
- Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.
- *Ko, H.M. (2012). Glossing and second language vocabulary learning. *TESOL QUARTERLY*, 46, 1. 56–79.
- Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). Cumulativemeta-analysis of therapeutic trials for myocardial infarction. *The New England Journal of Medicine*, 327, 248–254.
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.

- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand oaks, CA: Sage.
- Liu, J. (2007). The place of methods in teaching English around the world. In J. Liu (Ed.), *English language teaching in China: New approaches, perspectives, and standards* (pp. 13–41). New York: Continuum International Publishing Group.
- *Meli, R. (2009). *hypermedia and vocabulary acquisition for second Language*. Unpublished doctoral dissertation. Walden University.
- *Metaxa, X. T. (2013). *The effect of authentic songs on vocabulary acquisition in the English foreign language classroom*, unpublished doctoral dissertation. Saint Louis university.
- Min, H. (2008). EFL vocabulary acquisition and retention reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, 58(1), 73–115.
- *Mori, M. (2011). *The effects of singing on the vocabulary acquisition of university Japanese foreign language students*, unpublished doctoral dissertation. The University of Kansas.
- Munoz, C. (2006). Age-related differences and second language practice. In R. M. Dekeyser (Ed.), *Practice in a second language*. UK: Cambridge University Press.
- Nagy, W. (1997). On the role of context in first-and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 64–83). New York, NY: Cambridge University Press.
- Nakanishi, T. (2014). *A meta-analysis of extensive reading research*, unpublished doctorals dissertation. Temple University.
- Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, 49(1), 6–34.
- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645–670.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York, NY: Cambridge University Press.
- Normand, S. L. T. (1999). Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3), 321–359.
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research : Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85–110.
- Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). London: Academic Press.
- Paribakht, T., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21, 195–224.
- Paribakht, T. S. (2005). The influence of L1 lexicalization on L2 lexical inferencing: A study of Farsi-speaking EFL learners. *Language Learning*, 55(4), 701–748.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61, 993–1038.
- Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? interpreting effect sizes in L2 research. *Language Learning*, 1(62), 1–35.
- *Radwen, A.A. & Boyer, J.R. (2011). Semantic processing and vocabulary development of adult ESL learners. *Asian Journal of English Language Teaching*, 21, 1–21.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Rott, S. (2004). A comparison of output interventions and un-enhanced reading conditions on vocabulary acquisition and text comprehension. *The Canadian Modern Language Review*, 61(2), 169–202.
- *Rukhlom, V.N. (2011). *Facilitating Lexical Acquisition in Beginner Learners of Italian Through Popular Song*, unpublished doctoral dissertation. University of Toronto.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
- Schwabenflugel, P. J., Akin, C., & Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, 20, 96–104.
- Shintani, N. (2015). The effectiveness of processing instruction on L2 grammar acquisition: Meta-analysis. *Applied Linguistics*, 36(3), 306–325.
- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ*, 320, 1574–1577.
- Sutton, A. J., & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5), 625–650.
- *Tongpoon, A. (2009). *The enhancement of EFL learners' receptive and productive vocabulary knowledge through concordance-based methods*. Unpublished doctoral dissertation. Northern Arizona University.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19, 287–307.
- Won, M. (2008). *The effects of vocabulary instruction on English language learners: A meta-analysis*. Unpublished doctoral dissertation, Texas Tech University, Lubbock, TX.
- Yamamoto, Y. (2013). Multidimensional vocabulary acquisition through deliberate vocabulary list learning. *System*, 42, 232–243.
- *Yang, B. (2011). *Investigating noticing of errors and vocabulary acquisition in a multimedia environment*, Unpublished doctoral dissertation. Steinhardt School of Culture, Education, and Human Development New York University.
- *Yip F. W. M. & Kwan, A. C. M. (2006) Online vocabulary games as a tool for teaching and learning English vocabulary, *Educational Media International*, 43:3, 233–249.
- Yousefi, M. H., & Biria, R. (2011). Interactional feedback, task-based interaction and learner uptake. *Contemporary Online Language Education Journal*, 1(1), 1–19.
- Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, 24(1), 39–58.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, 57, 541–572.
- Zhang, L. (2007). *Study of incidental vocabulary learning of Chinese intermediate English learners*. MA dissertation. Shanghai: Shanghai Jiao Tong University.