ReCALL (2022), 34: 2, 235–252
doi:10.1017/S0958344021000239

RESEARCH ARTICLE

A meta-analysis examining technology-assisted L2 vocabulary learning

Aiqing Yu
Southwest Jiaotong University, China (aiqing.yu@huskers.unl.edu)

Guy Trainin
University of Nebraska–Lincoln, USA (gtrainin2@unl.edu)

Abstract

This meta-analysis examines the effectiveness of technology-assisted second language (L2) vocabulary learning as well as identifies factors that may play a role in their effectiveness. We found 34 studies with 2,511 participants yielding 49 separate effect sizes. Following the procedure developed by Hunter and Schmidt (2004), we corrected for sample size bias and measurement error. The overall effect size for using technology to learn L2 vocabulary was $d = 0.64$, which is a moderate effect size. The Q statistic indicated a significant variability in effect size, so we followed up with a theory-driven moderator analysis. The results of the moderator analysis revealed that learners benefited more from technology-assisted L2 vocabulary learning with incidental instruction than with intentional instruction; types of assessment were not significant moderators of the effect on technology-assisted L2 vocabulary learning; technology-assisted L2 vocabulary learning is more effective when the target language is close to the learner's first language; college students benefited more from technology-assisted L2 vocabulary learning than K–12 students; and, finally, mobile-assisted L2 vocabulary learning was more effective than computer-assisted L2 vocabulary learning.

Keywords: meta-analysis; technology-assisted learning; vocabulary learning; effectiveness

1. Introduction

Vocabulary is arguably the foundation of mastering a language because it comprises the building blocks of meaning. Extensive vocabulary can make speaking, listening, reading, and writing smoother and situationally precise (Webb & Nation, 2017). It is the key to communicating successfully. Vocabulary learning is not simply remembering a list of words but rather a complex process. For example, the learning burden of learning second language (L2) vocabulary can come from a variety of resource forms, which include the linguistic systems of learners' first language (L1), the similarities between learners' L1 and L2, the way in which the vocabulary is taught, and the learners' experience of the word (Webb & Nation, 2017). Hence, L2 learners often struggle to learn and to memorize vocabulary because lexical knowledge does not generalize easily.

The rapid development of new technologies with novel affordances provides new opportunities to meet the challenge of L2 vocabulary acquisition. Learners can now develop vocabulary through computer and mobile devices using language learning applications, online communication tools, computerized glosses, and games. The advantage of technology-supported vocabulary learning is predicated on the availability of practice and the use of media to support meaning-making in or out of context through the use of videos, pictures, audio, and L1 access. Nevertheless, researchers

Cite this article: Yu, A. & Trainin, G. (2022). A meta-analysis examining technology-assisted L2 vocabulary learning. *ReCALL* 34(2): 235–252. https://doi.org/10.1017/S0958344021000239

© The Author(s), 2021. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

also pointed out that the affordances have also increased the challenges for teachers, learners, and instructional designers (e.g. Chapelle, 2007; Golonka, Bowles, Frank, Richardson & Freynik, 2014; Ma, 2017). The challenge is in finding ways to select appropriate vocabulary learning apps, turn them into effective tasks for L2 learners, satisfy L2 learners' different needs, and develop self-regulated strategies.

A number of quantitative studies have been carried out to investigate the impact of technology-assisted vocabulary development, including vocabulary learning through digital games, instant messaging, mobile applications, and computer software (e.g. Dodigovic, 2013). The aim of a single experimental study is to decide if an intervention has a measurable effect on learners. A single study is not enough evidence for changing practice, but, once a field accumulates enough studies, a meta-analysis can provide adequate evidence for the efficacy of an approach.

Several meta-analyses have been conducted to shed light on the impact of technology-assisted L2 learning and Zhao's (2004) study is one of the most cited and the earliest meta-analysis in the field of technology-assisted language learning. His analysis included nine studies (nine effect sizes) with a sample size of 419 and found a large effect size, Cohen's $d = 1.12$. This early meta-analysis did not correct for bias nor explore potential moderators. Grgurović, Chapelle and Shelley's (2013) meta-analysis included 37 studies yielding 52 effect sizes corrected for sampling bias. They found a small effect size ($d = 0.25$) for the standard mean difference at post-test with the equivalence of the pre-test. They found $d = 0.35$ for the standard mean gain in studies in which the equivalence of the pre-test was not established after correction for sampling bias. Taj, Sulan, Sipra and Ahmad's (2016) meta-analysis is one of the latest in the field, which included 13 studies ($n = 813$). They discovered a large effect size of $d = 0.80$ after correcting for sampling bias. Two meta-analyses addressed vocabulary learning specifically. Chiu's (2013) meta-analysis examined the impact of computer-assisted L2 vocabulary learning from 16 studies with a sample size of 1,684 and discovered a moderate effect size of $d = 0.75$. Yun (2011) explored the efficacy of L2 vocabulary learning assisted by hypertext gloss from 10 studies ($n = 1{,}560$) and found a positive effect size, $d = 0.46$. The two meta-analyses each examined a specific technology. As a result, there is still only partial understanding of the overall effect of technology on L2 vocabulary learning. Smartphones started to grow rapidly in the late 2000s. These meta-analyses predate the dramatic increase in the popularity of mobile devices in education, including L2 vocabulary instruction. Since the meta-analyses in the field have been published, new technologies have emerged and substantial work has been done that justifies a follow-up.

2. Theoretical framework

The penetration of digital technology into education has introduced new opportunities for L2 teaching and learning. It has also posed challenges for teachers and learners. The main challenge is to figure out what digital applications and what general principles improve current practices. Research has shown that technology can have both positive and negative impacts on L2 learning (e.g. Zhao, 2004).

According to Clark and Paivio's (1991) dual coding theory, people encode information through two routes: visual and verbal. The verbal route encodes linguistic information in all its forms, whereas the visual route encodes images. When the inputs of the two routes overlap, encoding and retrieval improve. The referential connections between the two codes allow operations such as imagining words to reinforce input and accurate retrieval of information. Moeller *et al.*, 2009) pointed out that teaching with multimedia addresses individual learning needs by providing students opportunities to be exposed to language in multiple modalities, which will increase the speed of L2 learning and enhance vocabulary retention. Based on dual coding theory, the use of technology can enhance retrieval by incorporating images, sounds, and print to facilitate L2 vocabulary learning.

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Table 1. Theoretical framework

Theoretical framework	Moderator
Incidental and intentional vocabulary instruction	Type of instruction
Receptive knowledge and productive knowledge	Type of assessment
Linguistic distance	Participants' native language
Cognitive load theory	Grade level
Individualized learning Accessibility	Type of technology

Although experiments show what works, it is also important to compile cumulative results to understand how multiple studies shed light on theories that explain the potential benefits and constraints of technology-assisted L2 vocabulary learning. A moderator is a third variable that affects the relationship between two variables. For example, many studies have shown that technology-assisted L2 vocabulary learning is more effective than traditional vocabulary learning, and type of instruction (incidental/intentional) might be a moderator that affects the result. In the following sections, we reviewed the relevant theories that led us to use specific moderators (see Table 1). Our goal in exploring moderators is to understand what affordances lead to better results in a way that allows practitioners and future digital designers to focus on effective practices. Table 1 provides a list of the main theories related to the moderators in the meta-analysis.

2.1 Incidental/Intentional vocabulary learning

With the question of how vocabulary should be taught, the ongoing conversation has centered on the two major types of vocabulary learning: incidental/implicit and intentional/explicit/deliberate vocabulary learning. Various terminologies for the two major types are employed in the research field. Researchers like Dodigovic (2013) and Webb and Nation (2017) are in favor of the term "incidental and deliberate" vocabulary learning, Hulstijn (2001) prefers to use "incidental and intentional" vocabulary learning, whereas Gu (2003) and Ma use the terms "explicit and implicit" vocabulary learning. Intentional instruction stresses use of deliberate retention techniques to commit new information to memory (Hulstijn, 2001). Some intentional strategies such as word part analysis, dictionary use, and mnemonic techniques use (Nation, 2001) focus on learning some words actively, which are valuable shortcuts for L2 vocabulary growth. Technology-assisted intentional vocabulary learning aims to help learners comprehend words with a focus on linguistic codes through digital technologies (e.g. L2 vocabulary learning through hyper gloss, e-dictionary, text message of target word definitions). Incidental instruction stresses learners' ability to infer the meaning of new words from the contextual clues by providing rich and plentiful comprehensive input, as well as opportunities for interactions (Webb & Nation, 2017). It provides learners with a rich sense of word use and meaning from context as well as promotes reading or listening and vocabulary learning at the same time. Technology-assisted incidental vocabulary learning aims to help learners acquire words incidentally through digital technologies (e.g. L2 vocabulary learning through game-based L2 learning, computer-mediated communication, text message of target words embedded in idioms, sentences, and stories). There is a need to know which vocabulary instruction with technology is more effective.

2.2 Receptive/Productive vocabulary knowledge

Nation (1990) categorized vocabulary knowledge into receptive vocabulary knowledge and productive vocabulary knowledge. Receptive knowledge is the ability to recognize words and recall

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

their meaning when heard or read. Productive knowledge is the ability to accurately use words in communicative and non-communicative contexts. Nation (2001) stated that a single type of assessment could not satisfactorily measure every aspect of learners' word knowledge. Many researchers in technology-assisted L2 vocabulary acquisition studies adopted different assessments to measure aspects of vocabulary knowledge. For example, multiple-choice tests assess vocabulary knowledge of recognition, and sentence translation tests and mixed-type tests assess vocabulary knowledge of production. It is important to know what aspects of vocabulary knowledge can be better acquired through technology. We hypothesize that multiple-choice tests assessing vocabulary knowledge of recognition will generate higher effect sizes than productive measures.

2.3 Linguistic distance

Languages differ from each other in a myriad of ways, such as phonology, morphology, syntax, and semantics. Linguistic distance is the degree of closeness between languages; it is one of the important factors that affects L2 acquisition (Chiswick & Miller, 2012). Researchers argue that if learners' L1 is structurally close to the target language, transfer of learning should be easier (Chiswick & Miller, 2012). The higher the percentage of cognate words and degree of lexical relatedness in the two languages, the lower their linguistic distance is and the easier learners acquire the words from one another. For example, English is lexically closer to Spanish than it is to Chinese; if all other factors remain equal, it would be expected that native Spanish learners would attain a higher level or the same level of lexical knowledge in English sooner than native Chinese learners. Participants' native language can be one of the factors that impact the effectiveness of technology intervention in their L2 learning. We hypothesize that learners whose native language is closer to the target language will be able to benefit more from technology.

2.4 Cognitive load

In their theory of cognitive load and multimodal learning, Sweller, van Merriënboer and Paas (1998) postulated that cognitive processing includes two parts: working memory and long-term memory. Working memory has limited capacity and short storage span, whereas long-term memory is virtually unlimited. When learners acquire novel information, working memory serves as temporary storage to register and process information for performing complex cognitive tasks (Baddeley, 1993; Sweller, 2017). Then, attentional mechanisms allow the registered information to be transferred into long-term memory. The transfer into long-term memory enables future retrieval and further reduces working memory load. When learning a lot of new information in a short span, learners may find it difficult to store the information in long-term memory because novel stimuli may overload working memory.

Kalyuga (2012) provided a comprehensive review of cognitive load effects when presenting visual and verbal instructions simultaneously and continuously. She concluded that instructions that contain redundant information might split students' attention and increase their cognitive load, leading to lower achievement. Working memory capacity can be exceeded when integrating too much information (e.g. new words, images) into vocabulary teaching, thus impeding students' learning. Paas and van Merriënboer (1994) conducted a comprehensive overview of factors determining the level of cognitive load and identified age as one of the causal factors. Researchers postulate that there is a "maturational increase" in working memory capacity (Cowan, 2011; Hitch & Halliday, 1983). Cowan (2011) defined chunks as the quantification of the capacity limit associated with short-term memory. He proposed that working memory capacity is four chunks in adults and fewer in children. Learners in different age groups have different working memory capacity and might benefit differently from technology intervention. By considering age as one of the potential moderators, we hypothesize that young adults will benefit more than children from technology because their cognitive load would be reduced.

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

2.5 Individualized learning accessibility

Kern (1995) claimed that L2 learning software can support individualized instruction by "offering the student the freedom to choose topics, to repeat input, to increase or to decrease task difficulty, and to get help whenever it is needed" (p. 457) as learners vary in reading skills and in their reliance on verbal and visual processing. Zhao (2005) pointed out that effective L2 teaching should be highly individualized and customizable so as to motivate all students, meet their diverse learning goals, and accommodate their individual psychological and cognitive needs. Technology can be tailored to differentiate the learning process, as it provides various paths to deliver content in order to satisfy different learning needs and allows students to work at an individual pace (Pederson, 1986). Ubiquitous digital devices with internet connectivity and a range of informational and communication tools are now available for many learners. It is a tool that facilitates access to language learning anytime and anywhere. As different devices afford different access, there is a need to know what devices and affordances have the largest effect on learners. In this research, we are concerned with the differences between mobile technology (phones and tablets) and more stationary computers.

The review of relevant general theories of L2 learning led us to a set of theoretically driven questions that guide mediator analysis. Based on Clark and Paivio's (1991) dual coding theory, technology-assisted instruction can facilitate vocabulary learning as it enhances the language exposure by integrating verbal and visual codes. An appropriate application of technology in language learning can lower learners' affective filter so as to enhance learners' L2 learning. Although technology increases the omnipresence of information, the consequence is that our temporary working memory is overloaded, hence learners' learning anxiety increases and learning can be ultimately hindered. Vocabulary knowledge is multifaceted, and a good combination of intentional and incidental learning promotes vocabulary learning and retention. Technology might play a role in acquiring different aspects of vocabulary knowledge and facilitating different vocabulary learning. According to the linguistic distance hypothesis, the effectiveness of technology intervention in L2 learning can be impacted by participants' native language. Working memory capacity differs in different age groups, and learners might benefit differently from technology intervention. Technology-assisted instruction with better accessibility provides various ways to deliver content and improve practice not only anytime but also anywhere. We assume that the effectiveness of technology-assisted L2 vocabulary learning differs based on type of instruction, type of assessment adopted in the study, participants' grade level and their native language, and type of technology.

The study was guided by the following research questions:

1. What is the impact of digital technology on L2 vocabulary learning?
2. How are results affected by type of instruction, type of assessment, participants' native language and their grade level, and type of technology?

3. Methodology

3.1 Identification and selection of studies

The purpose of this study was to summarize evidence for the effectiveness of technology use in L2 vocabulary learning. We used a meta-analytic approach to investigate findings from experimental studies of L2 vocabulary learning that compare the use of various technologies with traditional methods or materials. The first step in the preparation for the meta-analysis was to conduct a systematic literature search for recent studies comparing technology-assisted L2 vocabulary learning and traditional L2 vocabulary learning. Technologies used in vocabulary learning included the following: computer-assisted instruction programs, mobile device–assisted instruction programs, audio, video, the web, e-books, and electronic dictionaries. The databases

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

searched were Education Resources Information Center (ERIC), and Dissertation Abstracts (DA), Education (SAGE), Academic Search Premier (EBSCO), and Google Scholar. Various combinations of terms used in the search included *vocabulary learning*, *technology*, *media*, *computer assisted language learning*, *mobile assisted language learning*, *computer instruction*, *traditional instruction*, *second language*, *foreign language*, *compare*, *electronic dictionary*. In addition to the database search, we conducted a manual search of three major technology and L2 journals: *Computers & Education*, *ReCALL*, and *Language Learning & Technology*. We searched these three publications as they elicited a proportion of flagged studies. Overall, 359 studies were identified in technology-assisted L2 vocabulary acquisition.

Rosenthal (1991) argues that the probability of publication is increased by the statistical significance of the results so that published studies may not be representative of all studies conducted in the field. Grgurović *et al.* (2013) argue that unpublished works provide details necessary for a comprehensive research synthesis as much as published journal articles do. To avoid publication bias, this study included articles in both published journals and unpublished dissertations and research reports that researchers may have overlooked. Another concern is about study quality in this meta-analysis. We used the Social Sciences Citation Index (SSCI) inclusion of journals as a proxy for quality.

3.2 Inclusion criteria and coding

In order to calculate effect sizes from the original study, descriptive or inferential statistics are needed. Studies that did not report statistics or those that reported insufficient results were excluded. In this meta-analysis, we used the following criteria to determine which studies were retained:

1. Written or published between 2006 and 2017.
2. Measured participants' performance on a vocabulary assessment in a general L2 context.
3. Used an experimental or quasi-experimental design; employed pre-test/post-test or post-test only in two or multiple group comparisons: technology-assisted vocabulary learning group versus traditional vocabulary learning group. Treatments for the technology-assisted vocabulary learning group include L2 vocabulary teaching and learning through computer and mobile devices using language learning applications, online communication tools, computerized glosses, and games. Treatments for the traditional vocabulary learning group include standard teaching and learning procedures without technology integration (e.g. use printed materials).

Each study was coded for location, sample size, average learners' age, age standard deviation (*SD*), percentage of female participants, native language, grade, type of instructional technology, year of learning, assessment name, study design, participant assignment, type of instruction, duration of treatment, assessment pre-test means, and descriptive statistics. A code book is presented in Table 2.

3.3 Statistical considerations

Cohen's *d* metric was used to calculate effect sizes in this meta-analysis because of its ease of interpretation and its common use in publication. The effect size *d* is the ratio of the difference between the means and *SD* (Hunter & Schmidt, 2004). This study compared the standardized mean difference between the post-test score of the experiment group and the control group in the two group comparison studies. The comparison is based on the post-test of the control and experimental groups.

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Table 2. Code book

Code	Subset	Definition	Examples
Type of instruction	Incidental ($k = 39$)	This instruction aims to teach vocabulary from the context without having a purpose to do so through digital technologies	L2 vocabulary learning through game-based L2 learning, computer-mediated communication, text message of target words embedded in idioms, sentences, and stories
	Intentional ($k = 10$)	This instruction aims to teach vocabulary actively with a focus on linguistic codes through digital technologies	L2 vocabulary learning through hyper gloss, e-dictionary, text message of target words definition
Type of assessment	Recognition ($k = 11$)	Assessments testing vocabulary knowledge of recognition	Studies adopted multiple-choice test for the post-test
	Production ($k = 24$)	Assessments testing vocabulary knowledge of production	Studies adopted sentence translation test and mixed-type test for the post-test
Linguistic distance	Non-Indo-European ($k = 32$)	Participants who speak Non-Indo-European languages learning an Indo-European language	Chinese speakers learning English
	Indo-European ($k = 17$)	Participants who speak Indo-European languages learning Indo-European languages	Spanish speakers learning English; English speakers learning Italian
Participant grade level	Undergraduate ($k = 24$)	College students	Age range: 17+
	K–12 ($k = 18$)	K–12 students	Age range: 6–18
Types of technology	Computer-assisted language learning ($k = 19$)	Computer-assisted L2 vocabulary learning	Computer programs originally designed for language learning, computer-mediated communication program, digital games, and the web
	Mobile-assisted language learning ($k = 17$)	Mobile-assisted L2 vocabulary learning	Mobile device applications originally made for language learning and text messaging

To correct for bias in sample size, we assigned weights to studies based on the number of participants. This study adopted bare-bones meta-analysis as a first step, using the random-effects model developed by Hunter and Schmidt (2004). Random-effects models assume that the true effect size can vary from study to study. Using a random-effects model, the mean of a distribution of true effects can be estimated.

The formula used is:

$$\text{Ave}(d) = \sum w_i d_i / \sum w_i$$

$$\text{Var}(d) = \sum w_i [d_i - d]^2 / \sum w_i$$

$$\text{Var}(e) = \sum w_i \text{Var}(e_i) / \sum w_i$$

$$\text{Ave}(\delta) = \text{Ave}(d)$$

$$\text{Var}(\delta) == \text{Var(e)} = \left[(N-1)/(N-3)\right]\left[\left(4/N\right)\left(1 + \text{Ave}(d)2/8\right)\right]$$

$$\text{SD}(\delta) = \sqrt{\text{Var}(\delta)}$$

Ave(d) = the weighted average of d, where w_i = the sample size of the ith study, d_i = the effect size of the ith study; Var(d) = the correspondingly weighted variance; Var(e) = the average sampling error variance; Ave(δ) = the population effect size; Var(δ) = the variance of population effect sizes, where N = the average sample size; SD(δ) = the study population effect sizes (Hunter & Schmidt, 2004: 287).

One of the challenges in estimating effect size is the impact of measurement error. It is important to correct for the effects of measurement error to ensure accuracy of the result of the meta-analyses (Hunter & Schmidt, 2004). The reliability of the dependent variable is not known for all studies, so we imputed the average reliability. We corrected the *d* value for measurement error by using the following formula:

$$d_c = d_o / \sqrt{r_{YY}} \text{ (Hunter \& Schmidt, 2004 : 303)}$$

We used the *Q* statistic to assess whether there is true heterogeneity in the meta-analysis. If the *Q* test is significant, it suggests that a percentage of the variability in effect estimates is due to systematic heterogeneity rather than sampling error; in other words, we can proceed to examine the impact of potential moderators.

The formula used was as follows:

$$Q = K\ Var(d)/Var(e) \text{ (Hunter \& Schmidt, 2004 : 416)}$$

Moderator variables help explain the variance in effect sizes when the *Q* statistic indicates a high probability of systematic error. Lau, Ioannidis and Schmid (1997) claimed that a meta-analysis allows the researcher to examine whether the effect is influenced by study characteristic. In this study, we adopted subgroup analysis for the detection of the moderator variables. Hunter and Schmidt (2004) suggested two ways to detect a moderator variable if the data are broken into subsets. First, there should be a difference in the mean effect size between subsets. Second, there should be a reduction in variance within subsets.

4. Results

In total, we found 34 studies with 2,511 participants that met all study criteria, yielding 49 effect sizes. Journals indexed by SSCI are described as the world's leading journals. There were 20 effect sizes yielded from 12 SSCI journals and 29 effect sizes yielded from 23 non-SSCI journals. The difference between the mean of effect sizes from SSCI journals and non-SSCI journals that were included in this meta-analysis is $t(47) = 0.64$, $p > 0.01$, which indicates a non-significant

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

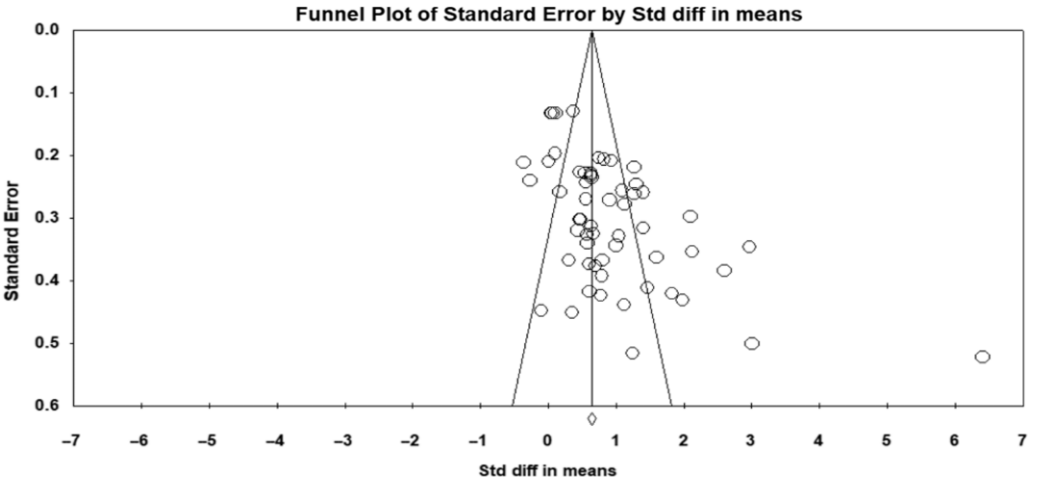


Figure 1. Funnel plot of standard error by standard mean differences (Std diff) in means

difference between the mean effect sizes of SSCI journals and non-SSCI journals. We assume that all included studies provided valid data to this study.

We used a funnel plot, a visual approach, to examine potential publication bias. A funnel plot is a scatter plot of effect sizes from each study against effect study precision. The funnel plot (Figure 1) in this study is asymmetrical, which raises the possibility of publication bias.

A forest plot is used to display the estimated effect from all included studies. In the forest plot, the y-axis represents the included studies and the x-axis represents the estimated corresponding effect of each of the studies. Each estimated effect is presented in the form of a square; the area of the square is proportional to the weight assigned to the study and the width of the line shows the confidence intervals of the effect estimate of individual studies (see Figure 2).

The number of effect sizes included in each subset is shown in Table 2. The results for the standardized mean difference between the post-test score of the experimental group and the control group are presented in Table 3. According to Cohen's (1988) guidelines for effect size magnitude, technology-assisted L2 vocabulary learning has a positive effect with a moderate effect size ($d = 0.64$, $SE = 0.08$, 95% CI [0.48, 0.80]) after correcting measurement error and sampling error. The result shows that L2 vocabulary learning supported by instructional technologies was more effective than instruction without technologies.

Tests of homogeneity of variance (Q test) was significant ($Q = 168$, $p < 0.001$), which indicates that the percentage of the variability in effect estimates is due to heterogeneity of variance. Hence, we can proceed to examine the impact of potential moderators.

4.1 Type of instruction

Based on the ongoing discussion of how vocabulary should be taught (incidental instruction vs. intentional instruction), we categorized the studies based on the types of instruction. One group included studies that adopted intentional instruction (e.g. hyper gloss, e-dictionary). It contained 26 studies yielding 39 effect sizes. Another group included studies that adopted incidental instruction (e.g. game-based L2 vocabulary learning, computer-mediated communication). It contained eight studies yielding 10 effect sizes (see Table 4). A medium effect size was found for intentional instruction subset ($d = 0.57$, 95% CI [0.39, 0.75]); a large effect size was found for incidental subset ($d = 1.04$, 95% CI [0.90, 1.18]). The difference between the mean for these two subsets is $t(47) = 2.67$, $p < 0.01$, which indicates a significant difference between mean effect

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

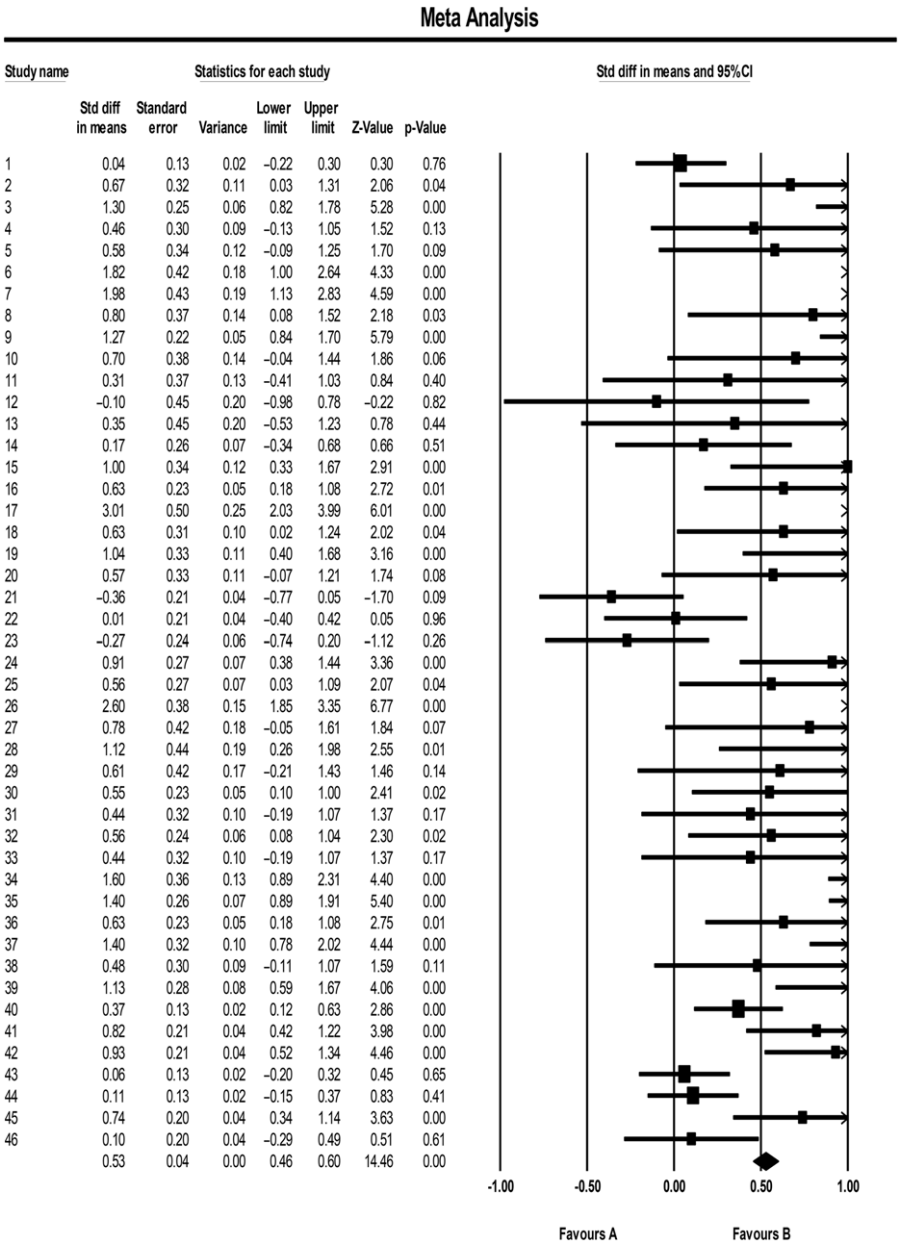


Figure 2. Forest plot meta-analysis of technology-assisted L2 vocabulary learning

sizes of the two subsets. It indicates that learners benefited more from technology-assisted L2 vocabulary learning with incidental instruction than with intentional instruction.

4.2 Types of assessment

Given that different types of assessments address different skills, we categorized the studies based on the types of vocabulary assessment. One group included studies that adopted multiple-choice tests assessing vocabulary knowledge of recognition. This contained 11 studies yielding 11 effect

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Table 3. Meta-analysis performed on the 34 two group comparison studies (49 effect sizes)

Bare-bones meta-analysis analysis
T = 3,227
$n = 71$
Ave(d) = 0.64
Var(d) = 0.37
Var(e) = 0.07
Ave(δ) = Ave(d) = 0.64
Var(δ) = Var(d)–Var(e) = 0.31
SD(δ) = 0.55
Q = 168

Table 4. Within-subset meta-analysis for type of instruction

Incidental	Intentional
T = 448	T = 2779
$n = 45$	$n = 77$
Ave(d) = 1.04	Ave(d) = 0.57
Var(d) = 0.14	Var(d) = 0.40
Var(e) = 0.09	Var(e) = 0.05
Ave(δ) = 1.04	Ave(δ) = 0.57
Var(δ) = 0.05	Var(δ) = 0.34
SD(δ) = 0.22	SD(δ) = 0.54

sizes. Another group included studies that adopted sentence translation tests and mixed-type tests assessing vocabulary knowledge of production. This contained 16 studies yielding 24 effect sizes. Seven studies were excluded due to insufficient description of the adopted outcome measures. As shown in Table 5, a medium effect size was found for the recognition subset ($d = 0.69$, 95% CI [0.45, 0.93]), and small effect size was found for the production subset ($d = 0.47$, 95% CI [0.25, 0.69]). Although the recognition subset generated a medium effect size and the production subset generated a small effect size, there was no statistical difference found between these two subsets, $t(33) = 1.19$, $p = 0.18$. These results indicate that there is no difference between the receptive vocabulary knowledge and the productive vocabulary knowledge that L2 learners acquired through technology.

4.3 Linguistic distance

Given that linguistic distance can influence L2 acquisition, we categorized the studies based on linguistic distance between the participants' native language and the target language. In one group, participants' native language and the target language differ significantly. This group includes studies with participants who natively speak a Non-Indo-European language (Chinese, Japanese, Thai, and Turkish) learning an Indo-European language (English). This group

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Table 5. Within-subset meta-analysis for type of assessment

Recognition	Production
T = 935	T = 2292
$n = 85$	$n = 72$
Ave(d) = 0.69	Ave(d) = 0.47
Var(d) = 0.22	Var(d) = 0.35
Var(e) = 0.05	Var(e) = 0.06
Ave(δ) = 0.69	Ave(δ) = 0.47
Var(δ) = 0.17	Var(δ) = 0.29
SD(δ) = 0.41	SD(δ) = 0.54

Table 6. Within-subset meta-analysis for participants' native language

Non-Indo-European	Indo-European
T = 2228	T = 959
$n = 74$	$n = 64$
Ave(d) = 0.48	Ave(d) = 0.85
Var(d) = 0.37	Var(d) = 0.29
Var(e) = 0.06	Var(e) = 0.07
Ave(δ) = 0.48	Ave(δ) = 0.85
Var(δ) = 0.31	Var(δ) = 0.22
SD(δ) = 0.56	SD(δ) = 0.47

contained 22 studies yielding 32 effect sizes. In another group, participants' native language and the target language do not differ significantly. This group includes studies with participants who natively speak Indo-European languages (Spanish, Persian, English) learning Indo-European languages (English, Spanish, and Italian). This group contained 12 studies yielding 17 effect sizes.

As shown in Table 6, after correcting the measurement and sampling error, a small effect size was found for learners who natively speak a Non-Indo-European language learning an Indo-European language ($d = 0.48$, 95% CI [0.17, 0.67]), and a large effect size was found for learners who natively speak an Indo-European language learning another Indo-European language ($d = 0.85$, 95% CI [0.69, 1.03]). The difference between the means for these two subsets is $t(47) = 2.20$, $p < 0.05$, which indicates that learners who are learning a similar language to their native language benefited more from technology-assisted L2 vocabulary learning than those who are learning a language that differs significantly from their native language.

4.4 Participant grade level

We categorized participant grade level into two subsets given that maturity level is one of the factors that may impact students' learning. The undergraduate subset contained 20 studies and 24 effect sizes, and the K–12 subset contained 13 studies and 18 effect sizes. One study was excluded because it did not provide information on the age range of its population. As shown in Table 7, after correcting the measurement and sampling error, a large effect size was found for undergraduate students ($d = 0.84$, 95% CI [0.57, 1.10]), and a small effect size was found for K–12

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Table 7. Within-subset meta-analysis for participants' grade level

Undergraduate	K–12
T = 1571	T = 1352
$n = 65$	$n = 75$
Ave(d) = 0.84	Ave(d) = 0.30
Var(d) = 0.52	Var(d) = 0.10
Var(e) = 0.07	Var(e) = 0.05
Ave(δ) = 0.84	Ave(δ) = 0.30
Var(δ) = 0.45	Var(δ) = −0.05
SD(δ) = 0.67	SD(δ) = 0.21

Table 8. Within-subset meta-analysis for type of technology

CALL	MALL
T = 1946	T = 1081
$n = 85$	$n = 64$
Ave(d) = 0.46	Ave(d) = 0.85
Var(d) = 0.34	Var(d) = 0.31
Var(e) = 0.05	Var(e) = 0.07
Ave(δ) = 0.46	Ave(δ) = 0.85
Var(δ) = 0.29	Var(δ) = 0.24
SD(δ) = 0.54	SD(δ) = 0.49

Note. CALL = computer-assisted language learning; MALL = mobile-assisted language learning.

students ($d = 0.30$, 95% CI [0.20, 0.39]). The difference between the mean for these two subsets is $t(30) = 3.29$, $p < 0.01$, which indicates a significant difference between mean effect sizes of undergraduate students and K–12 students. The study shows that college students benefited more from technology-assisted L2 vocabulary learning than K–12 students.

4.5 Types of technology

Technology for L2 vocabulary teaching can be used in many different ways. In order to examine the influence of the types of technology used in the L2 vocabulary instruction, we categorized the studies into two groups: computer-assisted L2 vocabulary learning (CALL) and mobile-assisted L2 vocabulary learning (MALL). Computer-assisted L2 vocabulary learning includes computer programs originally designed for language learning, computer-mediated communication programs, digital games, and the web. This group contains 14 studies yielding 19 effect sizes. Mobile-assisted L2 vocabulary learning includes mobile device applications originally made for language learning and text messaging. This group contains 17 studies yielding 17 effect sizes. Four studies were excluded due to the insufficient description of types of technology adopted. As shown in Table 8, a medium effect size was found for computer-assisted L2 vocabulary learning ($d = 0.46$, 95% CI [0.22, 0.70]) after correcting the measurement error and sampling bias, and a

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

large effect size was found for mobile-assisted L2 vocabulary learning ($d = 0.85$, 95% CI [0.62, 1.08]) after correcting the measurement error and sampling bias. The difference between the means for these two subsets is $t(36) = 2.26$, $p < 0.05$, which indicates that studies using mobile-assisted L2 vocabulary learning performed better than studies using computer-assisted L2 vocabulary learning.

5. Discussion

This meta-analysis represents a comprehensive approach to the efficiency of technology-assisted L2 vocabulary learning over the past decade. Through the comprehensive research, we found 34 contemporary studies yielding 49 effect sizes that met the inclusion criteria. Our results indicated that L2 vocabulary learning assisted by technology across various conditions was more effective than instruction without technology. In addition to the overall effect of technology-assisted L2 vocabulary learning, this study also analyzed the relationship between technology-assisted L2 vocabulary learning and five variables identified as important moderators of outcomes.

The study showed that learners benefited more from technology-assisted L2 vocabulary learning with incidental instruction than with intentional instruction. A possible explanation for this might be that incidental instruction emphasizes learners' ability to infer the meaning of new words from the contextual clues, which requires a deeper level of cognitive processing than intentional instruction. A number of studies (e.g. Ma, 2017) have pointed out that technology provides learners with authentic spoken input, simulative communication opportunities, and multimodal and individualized learning environments, and creates opportunities for incidental L2 vocabulary learning. It may be that these affordances helped learners reach a higher level of cognition.

Although the study showed that there is no significant difference between the mean effect sizes of the recognition subset and the production subset, technology-assisted L2 vocabulary learning generated a medium effect size for receptive vocabulary knowledge but a small effect size for productive vocabulary knowledge. Nation stated that receptive knowledge is the knowledge required to listen or read and productive knowledge is the knowledge required to speak or write. This result, although striking, may be because teaching materials are designed to develop receptive skills rather than productive skills. The challenge in developing technology-based teaching materials is to better design materials to enhance productive skills. There might be more elements to consider when developing productive vocabulary knowledge, such as interactions with peers and teachers.

In terms of linguistic distance, our result showed that technology-assisted L2 vocabulary learning is more effective when the target language is close to the learners' L1. Previous studies have pointed out that transfer of learning is easier if the learners' L1 is structurally closer to the target language (e.g. Chiswick & Miller 2012). Learners who learn an L2 from a different system might need extra help and support from different perspectives to achieve the same proficiency level as learners who learn an L2 from the same system.

The effectiveness of technology-assisted L2 vocabulary learning for college students yielded a significantly larger effect size than for K–12 students. It is analogous to the findings of Chiu (2013) that high school and college students can benefit more from a CALL program than elementary school students. There are two possible explanations for this result. One reason might be that motivation and self-regulation levels differ across different age groups. Undergraduate students have clearer life goals and they can see how learning an L2 will contribute to those goals. They may also be more motivated and self-regulated due to their age and experience. In addition, Cummins' (1976) thresholds hypothesis claims that the learner must have a minimum competence and proficiency in either their L1 or L2 in order to avoid cognitive overload and allow "the potentially beneficial aspects to influence their cognitive functioning" (p. 1). College students may have

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

higher linguistic proficiency in L1 and L2 so that they can benefit more from technology-assisted L2 learning.

This study found that mobile-assisted L2 vocabulary learning is more effective than computer-assisted L2 vocabulary learning. This finding is contrary to that of Stockwell (2010), who compared learner's vocabulary learning achievement on mobile phones and computers and found no significant difference in terms of student scores. Many researchers have pointed out MALL's unique characteristics compared with CALL, which include immediacy, flexibility, and portability (e.g. Ma, 2017). These unique characteristics may explain the relatively larger effective size of the MALL subset.

6. Implications for practice

Although this meta-analysis showed that the overall use of technology in L2 vocabulary learning was more effective than traditional instruction, new technologies introduce uncertainty for students and teachers about how to use it to support language learning. In addition, instructional designers and developers need to pay attention to the factors that may affect students' learning in order to design more effective tools.

6.1 Recommendations for instructors

Instructors need to thoughtfully consider students' age group, affective filter, and language threshold while integrating technology in language instruction. It would be beneficial if instructors could provide enough comprehensive input based on students' language threshold and adapt technologies with multiple modalities to enable learners to choose whichever method they prefer.

Professional skills such as curriculum design and technical and routine skills are also needed in technology-assisted L2 vocabulary learning. The main purpose of technology-assisted L2 learning is to use technology effectively to create truly augmented experiences that would help students succeed academically. Instructors should begin by choosing the learning goals for each of the lessons by considering what is important and what the students already know and need to know in order to walk away with new knowledge. They then need to make pedagogical decisions while planning for the lesson, by considering students' prior experience that the teachers could draw from, how this would affect learning, and what activities are appropriate for achieving the learning goals. Finally, instructors need to be aware of the variety of resources and technologies available for improving students' language skills and then choose appropriate technologies that will support the activity type and assist the students in achieving the learning goals.

Instructors also need to closely evaluate technology selections, as technologies with better portability and flexibility may facilitate more effective L2 vocabulary learning. For example, an application that works from both computers and mobile devices is more effective than one that can only be accessed through computers. This choice would allow students to access learning materials not only anytime but also anywhere.

In addition, taking linguistic distance into consideration, teachers need to select technologies with more support for the learners who are learning an L2 from a different system. Some examples of such supportive elements could include definition, pronunciation, image, derived forms, synonyms, example sentences, and opportunities to practice learned knowledge through negotiations with others.

6.2 Recommendations for instructional designers

Technology designers should attend to creating different contexts for classroom instruction. An application should provide meaningful contexts in which target vocabulary is embedded in

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

sentences/stories and presented in multiple inputs: audio, pictorial, and textual. Learners could also benefit from applications with carefully designed tasks to practice learned vocabularies.

In order to develop vocabulary knowledge comprehensively including both receptive knowledge and productive knowledge, it would be beneficial if L2 vocabulary teaching and learning applications included as many supportive elements as possible to facilitate students' L2 vocabulary learning processes. Some examples of supportive elements include comprehensive language input, feedback for vocabulary use, access to extensive language data, and opportunities for interactions and communication.

L2 vocabulary apps need to be age appropriate to address different cognitive load capacities. It would be beneficial if an L2 vocabulary learning program could have a children's version and an adult's version with different topics or themes according to learners' interests. Take learning vocabularies for shopping as an example: the context for the children's version could be at a toy store, whereas shopping in a supermarket would be for the adults' version. The program could also be differentiated for the complexity of operation. The children's version should be easy to operate, whereas the adults' version could be more complicated and include more functions.

It is also important for technology designers to develop self-regulated strategies in L2 learning applications in order to make them more effective and efficient in the classroom setting. It would be beneficial if an L2 vocabulary learning program could allow learners to identify the type of tasks and goals, the amount of effort/time to achieve them, and the type of resources to use for accomplishing learning goals.

Mobile devices in education, including L2 vocabulary instruction, have increased dramatically in the past few years. Learners often switch between computers and mobile devices based on their needs and environment. Technology designers should develop applications by taking the compatibility of computers and mobile devices into consideration in order to facilitate learners to learn the target language anytime and anywhere.

6.3 Recommendation for researchers

Meta-analyses depend greatly on the quality of the studies that are included. In order to include as many studies as possible and increase the validity and reliability of results, we used very liberal criteria for inclusion. To increase our understanding of individual results and overall effect, we highly recommend that researchers use more rigorous research methods (e.g. include pre-tests). Furthermore, they should report future studies by considering the inclusion of greater detail about the methods and participants in the study, thereby allowing a deeper understanding of the moderators. The majority of the studies examined outcomes of technology-assisted intentional L2 vocabulary learning; therefore, we highly recommend that researchers explore more on the effectiveness of technology-assisted incidental L2 vocabulary learning. We also suggest that researchers incorporate instruction and outcomes that combine receptive and productive outputs.

7. Limitations

Although the meta-analysis offers an opportunity to combine independent research findings across studies and find an overall effect, there are a number of limitations in conducting a meta-analysis. This study inherits the limitations of the research method used by the primary researchers. This meta-analysis does not overcome the problems that are inherent in the primary studies, such as measurement error. Second, the funnel plot of the included studies is asymmetrical, which raises the possibility of publication bias (see Figure 1). Third, this study was limited to quasi-experimental studies involving groups with access to technology supports and control groups without access to such supports. Other research designs including within-group designs and qualitative studies make important contributions not recognized here. Furthermore, L2 vocabulary learning can be impacted by teaching methods, different views of word knowledge,

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

types of tests, and so on. More moderators can be investigated for future research, such as the vocabulary measures, instructional approaches, among others.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344021000239

Ethical statement. We confirm that this research has not been submitted to any other journal and that all data included were used in accordance with ethical guidelines.

References

Baddeley, A. (1993) Working memory and conscious awareness. In Collins, A. F., Gathercole, S. E., Conway, M. A. & Morris, P. E. (eds.), *Theories of memory*. Hove: Lawrence Erlbaum Associates, 11–20. https://doi.org/10.4324/9781315782119-2

Chapelle, C. A. (2007) Technology and second language acquisition. *Annual Review of Applied Linguistics*, 27: 98–114. https://doi.org/10.1017/S0267190508070050

Chiswick, B. R. & Miller, P. W. (2012) Negative and positive assimilation, skill transferability, and linguistic distance. *Journal of Human Capital*, 6(1): 35–55. https://doi.org/10.1086/664794

Chiu, Y.-H. (2013) Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology*, 44(2): E52–E56. https://doi.org/10.1111/j.1467-8535.2012.01342.x

Clark, J. M. & Paivio, A. (1991) Dual coding theory and education. *Educational Psychology Review*, 3(3): 149–210. https://doi.org/10.1007/BF01320076

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Cowan, N. (2011) The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia*, 49(6): 1401–1406. https://doi.org/10.1016/j.neuropsychologia.2011.01.035

Cummins, J. (1976) The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, 19: 1–43.

Dodigovic, M. (2013) Vocabulary learning with electronic flashcards: Teacher design vs. student design. *Voices in Asia Journal*, 1(1): 15–33.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L. & Freynik, S. (2014) Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1): 70–105. https://doi.org/10.1080/09588221.2012.700315

Grgurović, M., Chapelle, C. A. & Shelley, M. C. (2013) A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2): 165–198. https://doi.org/10.1017/S0958344013000013

Gu, P. Y. (2003) Vocabulary learning in a second language: Person, task, context and strategies. *TESL-EJ*, 7(2): 1–25.

Hitch, G. J. & Halliday, M. S. (1983) Working memory in children. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 302(1110): 325–340. https://doi.org/10.1098/rstb.1983.0058

Hulstijn J. H. (2001) Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (eds.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 258–286. https://doi.org/10.1017/CBO9781139524780.011

Hunter, J. E. & Schmidt, F. L. (2004) *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks: SAGE.

Kalyuga, S. (2012) Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, 7(2): 145–159. https://doi.org/10.1016/j.edurev.2011.12.002

Kern, R. G. (1995) Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, 79(4): 457–476. https://doi.org/10.1111/j.1540-4781.1995.tb05445.x

Lau, J., Ioannidis, J. P. A. & Schmid, C. H. (1997) Quantitative synthesis in systematic reviews. *Annals of Internal Medicine*, 127(9): 820–826. https://doi.org/10.7326/0003-4819-127-9-199711010-00008

Ma, Q. (2009) *Second language vocabulary acquisition*. Bern: Peter Lang.

Ma, Q. (2017). Technologies for teaching and learning L2 vocabulary. In Chapelle, C. A. & Sauro, S. (eds.) *The handbook of technology and second language teaching and learning*. Hoboken, NJ: Wiley, 45–61.

Moeller, A. K., Ketsman, O. & Masmaliyeva, L. (2009) The essentials of vocabulary teaching: From theory to practice. *Faculty Publications: Department of Teaching, Learning and Teacher Education*, 171: 1–16. http://digitalcommons.unl.edu/teachlearnfacpub/171

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

Nation, I. S. P. (2001) Learning vocabulary in another language. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Paas, F. G. W. C. & van Merriënboer, J. J. G. (1994) Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4): 351–371. https://doi.org/10.1007/BF02213420

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press

Pederson, K. M. (1986) An experiment in computer-assisted second-language reading. *The Modern Language Journal*, 70(1): 36–40. https://doi.org/10.1111/j.1540-4781.1986.tb05242.x

Rosenthal, R. (1991) *Meta-analytic procedures for social research*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412984997

Stockwell, G. (2010) Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology*, 14(2): 95–110.

Sweller, J. (2017) Cognitive load theory and teaching English as a second language to adult learners. *Contact Magazine*, 43(1): 10–14.

Sweller, J., van Merriënboer, J. J. G. & Paas, F. G. W. C. (1998) Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3): 251–296. https://doi.org/10.1023/A:1022193728205

Taj, I. H., Sulan, N. B., Sipra, M. A. & Ahmad, W. (2016) Impact of mobile assisted language learning (MALL) on EFL: A meta-analysis. *Advances in Language and Literary Studies*, 7(2): 76–83. https://doi.org/10.7575/aiac.alls.v.7n.2p.76

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

Yun, J. (2011) The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, 24(1): 39–58. https://doi.org/10.1080/09588221.2010.523285

Zhao, Y. (2004) Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, 21(1): 7–27. https://doi.org/10.1558/cj.v21i1.7-27

Zhao, Y. (2005) The future of research in technology and second language education: Challenges and possibilities. In Zhao. Y. (ed.), *Research in technology and second language education: Developments and directions*. Greenwich:Information Age Publishing, 445–457.

About the authors

Aiqing Yu is an assistant professor in the International Chinese Department at Southwest Jiaotong University with a focus on second language teaching and learning and technology-assisted second language learning.

Guy Trainin is the department chair of the Department of Teaching, Learning and Teacher Education at University of Nebraska–Lincoln. His research focuses on the intersection of literacy development, teacher education, and literacy integration with technology and the arts.

Author ORCiD. Aiqing Yu, https://orcid.org/0000-0003-1869-8016
Author ORCiD. Guy Trainin, https://orcid.org/0000-0002-2116-7155

https://doi.org/10.1017/S0958344021000239 Published online by Cambridge University Press