Journal of Research on Educational Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

# Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis

Tao Hao, Zhe Wang & Yuliya Ardasheva

**To cite this article:** Tao Hao, Zhe Wang & Yuliya Ardasheva (2021): Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2021.1917028

**To link to this article:** https://doi.org/10.1080/19345747.2021.1917028

Published online: 15 Jun 2021.

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
https://doi.org/10.1080/19345747.2021.1917028



THEORY, CONTEXTS, AND MECHANISMS

# Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis

Tao Hao[a], Zhe Wang[b] and Yuliya Ardasheva[c]

[a]Faculty of Education, East China Normal University, Shanghai, China; [b]Department of Curriculum and Instruction, Faculty of Education, East China Normal University, Shanghai, China; [c]Department of Teaching and Learning, Washington State University, Tri-cities, USA

**ABSTRACT**
This meta-analysis reviewed research between 2012 and 2018 focused on technology-assisted second language (L2) vocabulary learning for English as a foreign language (EFL) learner. A total of 45 studies of 2,374 preschool-to-college EFL students contributed effect sizes to this meta-analysis. Compared with traditional instructional methods, the overall effect of technology-assisted L2 vocabulary learning was large ($g = .845$), suggesting that technology-assisted L2 vocabulary learning was more beneficial than non-technology-assisted instruction. Importantly, within-study comparison results indicated that technology could enhance learners' long-term vocabulary retention. Moderator analysis results highlight several variables—namely, device type, game condition, setting, test format, and reported reliability—affecting the effectiveness of vocabulary learning. Specifically, advantages were found for mobile devices and on-the-move learning, suggesting that L2 vocabulary learning may be most efficient when students use mobile phones and are not restricted by classroom settings. These variables should be considered when planning instruction in technology-assisted L2 vocabulary learning.

**ARTICLE HISTORY**
Received 22 April 2020
Revised 4 January 2021
Accepted 20 March 2021

**KEYWORDS**
Technology; language education; vocabulary; EFL; meta-analysis

## Introduction

The importance of vocabulary learning has been widely acknowledged and well documented in the field of second language acquisition (Ardasheva et al., 2019). Learning vocabulary is an essential part of mastering a second language (L2), contributing to enhancing L2 listening, speaking, reading, and writing skills (Gorjian et al., 2011). High-quality word knowledge—including the knowledge of forms (pronunciation, spelling, morphological and grammatical word properties) and the knowledge of multiple word meanings across different contexts—is associated with the understanding of rich and interrelated information communicated by that word and plays an essential role in vocabulary learning. Given the complexity of knowing a word (Schmitt, 2014), learning vocabulary is one of the biggest challenges that students face in their language studies in

CONTACT Tao Hao tao.hao@wsu.edu 3663 North Zhongshan Road, Shanghai, 200062 China.

particular due to the limited classroom time focused on L2 instruction and/or to the limited exposure to the L2 studied outside of the classroom. Not surprisingly, then, vocabulary acquisition is typically below expectations for students learning English as a second or foreign language (ESL/EFL; Du, 2004; Gibson, 2016) and the consequences of having a weak vocabulary capacity, such as poor reading comprehension or poor speaking or writing skills, may compromise students' motivation to learn.

As an important constituting component of L2 acquisition, vocabulary learning has been significantly impacted by the emergence of new technologies. Technological activities can elicit L2 learners' interest and provide students with more verbal and multimedia exposure to the target language as well as with more opportunities to interact with the target language through the use of various technological devices. The influence of technology on how learners access and learn L2 vocabulary can manifest itself in such inconspicuous ways as the use of computers or mobile phones increasing learning opportunities outside of classroom (Li et al., 2017; Stockwell, 2011). The application of technology may be particularly important in EFL contexts, the primary focus of the present study, where the target language may not be available on an everyday basis.

Not surprisingly, then, a total of seven meta-analyses to date synthesized studies focused on technology-assisted L2 vocabulary learning. Yet, with one notable exception, all previous meta-analyses focused on evaluating the efficacy of a single technology-assisted strategy (e.g., games, or mobile devices), a shortcoming the present meta-analysis intends to address by synthesizing findings regarding the effectiveness of a range of currently available and researched technological devices and delivery formats used in EFL contexts.

## Literature Review

### *Technology-Assisted Vocabulary Learning*

Vocabulary knowledge may be receptive, associated with a learner's ability to understand a word encountered when listening or reading, or productive, associated with a learner's ability to use the word when speaking or writing (Schmitt, 2014). "The two mastery types are often perceived as lying on a developmental continuum, with knowledge shifting from receptive to productive mastery over time" (Ardasheva et al., 2019, p. 127). In turn, vocabulary learning could be intentional or incidental (Hulstijn & Laufer, 2001; Nation, 2001). Intentional vocabulary learning refers to the learning activities explicitly focused on acquiring new vocabulary, such as learning selected target words using word lists. By contrast, incidental vocabulary learning refers to the learning activities not explicitly aimed at vocabulary learning per se, such as when learners acquire new vocabulary simply from watching L2 videos or playing online games, without a specific goal of learning new vocabulary.

Given the importance of vocabulary in L2 learning and the limited in-class time, the question of how technology can facilitate incidental vocabulary learning has been focal in a substantial number of technology-assisted language learning studies (Basal et al., 2016; Franciosi, 2017; Taj et al., 2017), with a number of studies considering the differential impacts on receptive versus productive vocabulary learning (Tsai & Tsai, 2018). However, the cutting-edge nature of technology does not guarantee an effective learning process, or a desirable learning outcome as suggested by a number of studies in which traditional teaching and learning methods yielded better L2 vocabulary learning outcomes than did technology-assisted methods (Basöz & Can, 2016; Taghizadeh & Porkar, 2018; Young & Wang, 2014).

Such discrepancies across individual studies as well as differences in effect sizes across meta-analyses attempting to resolve individual studies' discrepancies may be due to the constant technological changes, which calls for periodic systematic reviews and meta-analyses of most current literature, with a careful attention to key characteristics of individual studies. Conducted under the larger umbrella of computer-assisted language learning (CALL), technology-assisted vocabulary learning research is focused on such issues as the effectiveness of technology-assisted strategies, including devices and delivery formats. Of importance, is also research focused on individual differences, contextual factors, and intervention dimensions that may impact the effectiveness of technology-assisted strategies. Thus, below we first summarize key findings regarding the current knowledge base regarding technology-assisted vocabulary learning effectiveness focusing on key dimensions along which interventions vary and then spotlight key moderators of L2 vocabulary interventions.

## Evidence Regarding Technology-Assisted L2 Vocabulary Learning Effectiveness to Date

Meta-analysis as a research synthesis technique has been used extensively in the field of technology-assisted language learning (e.g., Chwo et al., 2018; Hassan Taj et al., 2016; Sung et al., 2015). Below, we briefly summarize key research areas and findings emanating from previous meta-analyses focused on technology-assisted L2 vocabulary learning.

### Overall Effectiveness of Technology-Assisted Instructional Strategies

Overall, the use of computer-assisted strategies (e.g., digital games, mobile devices) is typically associated with positive effects on L2 vocabulary learning, when compared to both no treatment and traditional methods (Abraham, 2008; Chen et al., 2018; Mahdi, 2018). The reported effect sizes, however, vary in magnitude and practical importance, highlighting a need to identify key dimensions along which intervention impacts vary. Comparing access versus no access to computer-mediated *glosses* (i.e., short explanations of words' meanings), for example, Abraham (2008) found large effects on incidental vocabulary learning, both on immediate and delayed posttests ($g = 1.40$ and 1.25, respectively). By contrast, Yun (2011) found that *hypertext glosses* (glosses linked to corresponding words by a point-and-click method) had a small effect ($g = .46$) on L2 vocabulary learning. It is important to note that Yun's study compared the effectiveness of two experimental conditions, text-only versus text + visual hypertext glosses. The results also indicated that low-proficiency learners benefited more from multiple glosses than did intermediate and advanced learners. These findings suggest the importance of considering both levels of comparison and individual differences when examining the effectiveness of technology-assisted strategies.

Another high interest topic in technology-assisted L2 vocabulary learning research is a focus on instructional effectiveness associated with a specific technological device.

### *Technological Devices: Mobile-Assisted versus Computer-Assisted Vocabulary Learning*

In his review of 16 studies, Mahdi (2018), for example, reported a medium effect ($g = 0.67$) of *mobile-assisted learning*, learning supported by handheld mobile devices such as cellphones, as compared with traditional pedagogy, for both receptive and productive L2 vocabulary learning. This study found that adult learners benefited more from mobile-mediated learning than young learners did. Lin and Lin (2019) synthesized 33 studies and found a large effect ($g = 1.005$) of mobile-assisted ESL/EFL vocabulary learning. This study discovered that messaging services (a short message service, a multimedia message service) were more beneficial for word retention than mobile applications (apps) were. However, Lin and Lin argued, this result should be interpreted prudently, since the intervention varied, in particular, in terms of duration. Similar to studies on computer-mediated glosses effectiveness reviewed earlier, these findings suggest the importance of considering levels of comparison and individual differences, but also the importance of considering contextual variables when examining the effectiveness of technology-assisted vocabulary learning strategies.

The final key topic in technology-assisted L2 vocabulary learning research of interest to the present study is a focus on the effectiveness of delivery formats.

### *Delivery Formats: Game-Based versus Non-Game-Based Vocabulary Learning*

Comparing *digital game-based learning* (DGBL), learning integrating games to maximize students' engagement and learning potential, versus traditional pedagogy, Chen et al. (2018), for example, found a large effect of DGBL on vocabulary learning ($d = 1.027$). In contrast to other studies (Mahdi, 2018; Yousefi & Biria, 2018), no significant difference was found when age was examined as a moderator. A similarly large overall DGBL effect on L2 vocabulary learning ($d = 0.986$) was reported in Tsai and Tsai (2018) meta-analysis of 26 studies. Similar, Chiu (2013) found that game-based learning ($n = 9$, $d = 1.113$) was superior to non-game-based learning ($n = 7$, $d = 0.495$). However, it should be noted that the non-game-based learning studies in Chiu's meta-analysis still used computer-assisted interventions rather than traditional pedagogy with no computer-assisted components.

Overall, existing synthesis research on the effectiveness of technology-assisted L2 vocabulary learning is typically focused on a single technology-assisted strategy. The findings across these studies, however, suggest the importance of considering levels of comparison and moderator variables at different levels, which we discuss next.

## Moderator Variables

### *The Role of Age and Proficiency in Technology-Assisted L2 Vocabulary Learning*

Past meta-analytic studies across educational research and language learning fields have demonstrated that participant characteristics may serve as sources of variation in the estimated outcomes. Mahdi (2018) and Yousefi and Biria (2018), for example, found that age was a significant moderator of technology-assisted L2 vocabulary learning. Learners' L2 proficiency (Yun, 2011) and educational level (Chiu, 2013) have also been reported as important moderators in technology-assisted language learning. For instance, Chiu found that in a computer-assisted intervention, high school/college students' L2 vocabulary gains were substantially larger than those of elementary school students (effect sizes of 1.032 and .321, respectfully). It is important to note that in EFL settings educational level may be perceived as a proxy of L2 proficiency. That is, while elementary EFL students with only a few years of L2 study may be often only beginners, with additional years of L2 study at the secondary and higher education levels EFL learners may progress to intermediate or advanced proficiency. Considering individual differences, however, we cannot consider age and proficiency as one and the same moderator as in many studies participants at the same educational level may be further categorized as beginner, intermediate, and advanced. Thus, guided by prior research reviewed above, the current study investigated L2 proficiency and educational levels (age proxy) as potential moderators.

## The Role of Context in Technology-Assisted L2 Vocabulary Learning

Researchers have long acknowledged that contextual features may affect the effectiveness of an intervention. When it comes to technology-assisted L2 learning, literature suggests that the effectiveness of vocabulary learning might be moderated by the device, delivery format, settings, study duration, and test format. Regarding device and delivery format, for example, Chiu (2013) found that gains in the computer game-assisted vocabulary learning condition were significantly lower ($d$ = .495) than those in other computer-assisted (but not game-assisted) vocabulary learning conditions ($d$ = 1.113). Studies included in Chiu's meta-analysis, however, were completed by 2011 when smartphones had not yet been as widely used as today. Thus, no comparisons between computer- and mobile-assisted technologies were made. This is of concern, as even though game-based language learning can either be computer- or mobile-mediated, CALL researchers tend to separate game-assisted language learning from other interventions due to its unique characteristics (e.g., high engagement) and complexity.

Implementation settings and study duration have also been investigated as moderators in previous meta-analyses. For instance, research setting was a significant moderator for mobile-assisted word retention in ESL/EFL contexts (Lin & Lin, 2019). Sung et al. (2015) found that 1–6 month interventions had the largest effect size, followed by 2–4 weeks and > 6 months interventions ($g$ = .772, .622, and .130, respectively). By contrast, computer-assisted language learning interventions lasting less than 1 month yielded larger effect ($d$ = 1.574) than those lasting longer ($d$ = 0.499; Chiu, 2013). The author attributed this unexpected finding to initial learner interest in the applications of new technology, followed by fatigue.

Lastly, the test format also serves as a moderator for vocabulary learning. Corresponding to two vocabulary knowledge types discussed earlier, vocabulary assessments can be roughly grouped under two categories, namely, receptive (i.e., participants select correct answers) and productive (i.e., participants write answers in their own words) and can be administered immediately or sometime after the intervention. For instance, examining game-assisted L2 interventions, Tsai and Tsai (2018) found that the effect size was larger when students were tested on a productive ($d$ =.839) rather than on a receptive ($d$ = .332) vocabulary knowledge test. Synthesizing computer-mediated glosses intervention research, Abraham (2008) found a significant difference in effects on learning between immediate receptive ($g = 1.81$) and productive ($g$ = .60) tests. A statistically significant difference was also found between delayed receptive ($g = 1.43$) and productive ($g$ = .21) vocabulary tests. It should be noted, however, that this finding needs to be interpreted with caution, since only one study included in Abraham's meta-analysis administered delayed vocabulary measures.

Thus, in terms of contextual moderators the present study separately examined device (computer versus mobile), delivery format (game- versus non-game based), intervention setting (classroom on the move) and duration (less or more than two weeks; Sung et al., 2015), as well as test format (receptive versus productive) as potential moderators of technology-assisted vocabulary learning.

### *The Role of Methodological Characteristics in Technology-Assisted L2 Vocabulary Learning Studies*

Literature (Marley & Levin, 2011; Plonsky, 2011) suggests that methodological characteristics of a study can affect the robustness of the results regarding the causality of interventions and outcomes. Adesope et al. (2010), for example, discovered that reliability-unreported studies yielded a statistically detectable effect size while reliability-reported studies did not. Cho et al. (2018), found significant difference between standardized ($d$ = .70) and researcher-developed ($d$ = .19) tests for mobile-assisted language learning. In order to ensure that recommendations for educational practices are built upon credible evidence, methodological moderators were also explored in the current study.

Therefore, building upon previous work, this study investigated the following methodological characteristics as potential moderators: reliability (reported or not), test origin (standardized or researcher-developed), and study design (quasi-experimental or experimental).

## Theoretical Framework

In grounding this study on technology-assisted vocabulary learning, we drew on two theoretical frameworks for, namely Paivio's dual-coding and Long's social interaction theories.

### *Dual Coding Theory*

The main premise of dual-coding theory (DCT; Clark & Paivio, 1991) is that learners processes information through two mental codes, namely, verbal and nonverbal. While the verbal code deals with processing and representing language, the nonverbal code deals with processing nonlinguistic information (Moody et al., 2018). The two codes are simultaneously independent and interconnected. In other words, although functionally independent, the two codes may have an additive effect on learning through, what Paivio (2006) calls, "a dual verbal-nonverbal memory trace" (p. 4). Such dual verbal-nonverbal memory trace may be particularly productive for creating and expanding cognitive schemas, constructed through activating prior knowledge and elaboration (Woloshyn et al., 1994).

Specific to vocabulary learning, verbal-only associations may produce shallow understandings, resulting from a learner's failure to make the abstract concrete (Moody et al., 2018). Thus, practices emphasizing the imageability (concreteness) of words are essential in vocabulary instruction. Technology-assisted learning has a tremendous potential to enhance such practices by providing an audio-visual (and, as we discuss below, interactive) input. This is particularly important as many teachers assume that students will naturally connect visuals and words (Metros, 2008). Teachers who understand DCT, Moody et al. (2018) argued, "recognize that instruction must include a purposeful focus on contextual referents so that all students will understand and internalize new words" (p. 4).

### *Long's Interaction Hypothesis*

According to Long (1996), social interaction plays a critical role in L2 acquisition by enhancing meaning negotiation and thus the comprehensibility of the L2 input. Social interaction holds special significance in language learning since the primary function of language is communication and effective learning is more likely when learners can actively interact with other people. In particular, meaningful interaction can assist learners in noticing and using new words for communicative purposes and thus enhance their L2 learning. Technology-assisted learning environments can provide social interaction either by proxy, through an audio-visual input, which can help learners pay particular attention to to-be-learned words and interact with them on computer screens (Mohsen, 2016), or directly, through interactive environments in which learners can exchange information and collaborate with others at any place due to technology portability (e.g., SMS, email). In other words, according to interaction hypothesis, technology-assisted vocabulary learning may help learners experience L2 more directly and naturally through meaning negotiation with the interactive technology itself or with others.

## Purpose of the Present Study

As noted earlier, seven meta-analyses to date synthesized studies focused on technology-assisted L2 vocabulary learning. Most of these studies, however, either combined multiple second/foreign languages or only focused on the effects of a single technology-assisted strategy (e.g., computer-mediated glosses). To inform the field of EFL education, there is a need for larger scale investigations on the effectiveness of technology in EFL vocabulary learning, especially considering the ever-increasing versatility of technology applications in education and in everyday life.

In an earlier meta-analysis on this topic, Chiu (2013) reviewed studies employing a variety of computer-assisted strategies, including, for example, computer glosses and game-assisted vocabulary learning. Chiu's study, however, did not explore mobile-assisted L2 vocabulary learning and only included studies completed by 2011. It is worth noting that studies conducted before 2011 used very different technological tools than the ones we have access to today and that there have been many empirical studies examining the effectiveness of computer-assisted vocabulary instruction published since 2011 due to the fast-changing nature of technology.

Therefore, the current meta-analysis extended on previous work by synthesizing the most recent EFL technology-assisted vocabulary learning research conducted between 2012 and 2018 to examine the effectiveness of more current technological tools on both vocabulary learning and retention over time.

## Methods

### *Research Questions*

Two primary research questions guided this meta-analysis:

1. What are the overall effects of currently available technology-assisted strategies on EFL vocabulary learning in comparison with traditional pedagogies?
2. What are the individual difference (age, proficiency), contextual (device, delivery format, duration, setting, assessment format), and methodological (reliability, test origin, study design) moderators that have significant influences on the technology-assisted vocabulary learning?

### *Study Inclusion Criteria*

The current meta-analysis extended on Chiu's (2013) study by focusing on most recent, 2012–2018, research. In addition, the studies had to meet the following requirements in order to be included in this meta-analysis:

- must use a computer- or mobile-based intervention;
- must recruit EFL students (ESL students were not included because of potential differences in technology needs between these two groups, considering lack vs. access to the English language in their learning and everyday environments);
- must recruit mainstream students (cannot be those with disability);
- must have at least one experimental group and one control group (the control group must be using traditional teaching or learning method without technology);
- must report vocabulary related learning outcomes (cannot be students' technology-assisted L2 vocabulary learning attitudes).

### *Search Procedures*

The primary search strategies included an electronic literature search as well as complementary literature searches. The following electronic databases were included: Eric, Education Full Text, Educational Administration Abstracts, Social Sciences Abstracts, Linguistics and Language Behavior Abstracts, PsychINFO, ProQuest Dissertations and Theses, and Google Scholar. To conduct a comprehensive and systematic search on the above databases the following search terms and logic were used: (ab (language learn*))

AND (ab(vocabulary)) AND (ab (computer* OR technolog* OR mobile* OR game*)) AND (ab (experiment* OR treat* OR intervention*)). Additional studies were retrieved from references in the documents located through computer searches.

From Eric, Education Full Text, Educational Administration Abstracts, and Social Sciences Abstracts, initial electronic literature searches yielded a total of 171 potentially relevant articles. After eliminating reports, 51 papers retained for further screening. From PsychINFO and ProQuest, the initial electronic literature searches yielded a total of 76 potentially relevant articles, with 42 articles and 34 dissertations retained for further screening. Lastly, 31 articles and 2 dissertations were retrieved for screening from the Linguistics and Language Behavior Abstracts database, yielding a total of 160 studies for further screening.

After initial screening by reading study titles and abstracts, 120 articles were removed. Final decisions to include or exclude studies were based on the reading of the full texts. A full text screening of the remaining 40 studies was conducted independently by two authors. Most studies excluded during the full-text screening process were those focused on ESL rather than EFL students; a number of studies were also excluded due to insufficient reporting of statistics. Finally, 24 articles and two dissertations were retained from electronic searcher; seven additional studies were retrieved from the references. Because some articles reported on multiple studies, the 33 articles yielded 45 studies. Figure 1 provides a visual depiction of study selection procedures.

### Data Coding

Studies were coded by two independent raters using a pre-established coding protocol, which was developed based on the Valentine and Cooper (2003, 2008) criteria for quality review. The coding protocol included the following categories:

- Study citation, author affiliation, type of publication, region, technology strategy.
- Student characteristics (e.g., educational level, L2 fluency).
- Contextual features (e.g., device, game- or non-game based, study duration, setting, test format).
- Methodology characteristics (e.g., research design, reliability, measurement tool).
- Result (e.g., number of participants, means, standard deviations).

There were two outcomes of data extraction: (1) posttest information for experimental and control group to generate a set of effect sizes (ESs) and (2) a list of categorical codes for subsequent moderator analyses. After an initial training, two raters independently coded 50% of randomly selected studies to insure consistency in coding. The inter-rater reliability was estimated using Cohen's Kappa for categorical variables (range: .83−1.0) and Pearson's r for continuous variables (range: .92−1.0).

### Data Analyses

Hedges'*g* served as the measure of effect size in this study. Hedges'*g* was calculated as the difference between the experimental (technology-assisted) and control (traditional

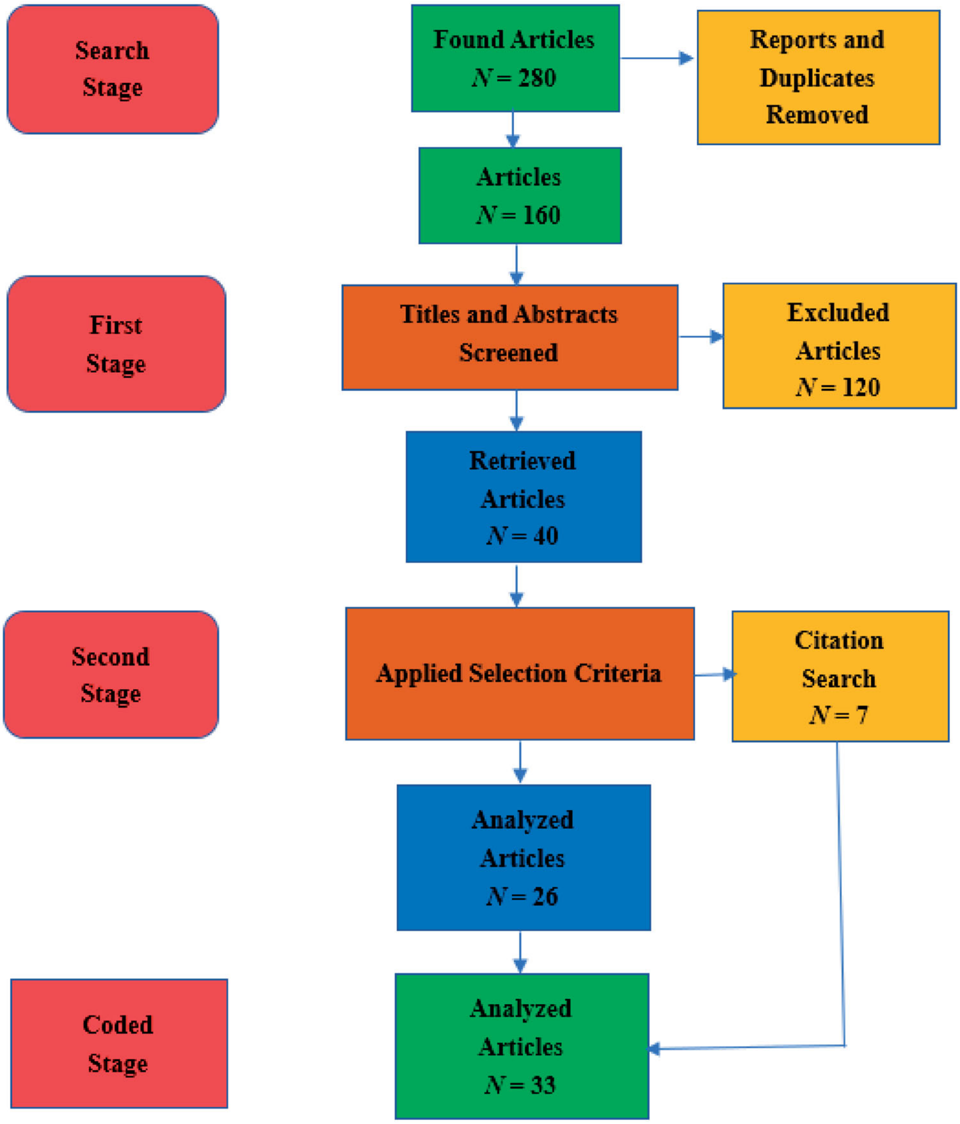


**Figure 1.** Visual depiction of study selection procedures.

methods) mean scores divided by the pooled standard deviation of the two groups. Hedges'*g* was computed using and reported as an unbiased estimate of the standardized mean difference effect size. As pointed out by Hedges and Olkin (1985), the effect size obtained by Cohen's *d* is likely to be biased by different sample sizes across studies.

If means and standard deviations were not available, other statistics provided in the article were converted into Cohen's *d* using effect size conversion calculations. For example, some continuous outcomes may be presented as point-biserial correlations, as *t*-tests, or as F-ratios from a one-way ANOVA. These statistics were converted to the appropriate effect size metric. The formula by Hedges and Olkin (1985) provided below served to convert Cohen's $d$ to Hedges'$g$:

$$g \cong d \times \left(1 - \frac{3}{4(n1 + n2) - 9}\right)$$

### *Moderator Analysis*

We examined educational level, L2 fluency, device, game, study duration, setting, test format, research design, reliability, and measurement tool as moderators of technology-assisted L2 vocabulary learning. All moderator variables in this study were categorical. Comprehensive Meta-Analysis (CMA), Version 3.0, was used to conduct moderator analysis. We also adopted the $Q_B$-test (i.e., between-group test) for each moderator variable to examine whether differences among subgroup means were statistically significant. As advocated by Borenstein et al. (2009) for combining studies within individual subgroups, we report the mixed-effects rather than fixed-effect analysis results generated by CMA for moderator analyses. We also computed $\tau^2$ and $I^2$ to provide estimates of the amount of heterogeneity (these data are available upon request).

### *Within-Study Comparisons of Subgroup Mean Effects*

Although moderator analysis provides valuable information about study features explaining the differences in mean effects, internal validity concerns may threaten the results. For example, a positive effect may largely stem from the immediacy of the posttest administered right after the technology-assisted intervention. To date, only one meta-analytic review involving technology-assisted vocabulary learning investigated the effectiveness of delayed post-tests. Specifically, Abraham (2008) found that the effect of computer-mediated glosses on vocabulary learning remained robust over time, with seemingly similar weighted mean effective sizes for immediate ($g = 1.40$) and delayed ($g = 1.25$) posttests. However, Abraham's meta-analysis did not explore whether the effect sizes associated with immediate and delayed post-tests were significantly different from each other. Given limited synthesis research on the effects of within-study comparisons of technology-assisted vocabulary learning, there is an urgent need to understand the effectiveness of long-term retention of vocabulary. Hence, a comparison of immediate and delayed posttests from a subsample of studies, including a corresponding test of statistical significance, is needed. In this study, a new effect size was created: $g_2$ - $g_1$, which is the difference between the mean effect size of delayed effect and the immediate effect. The formula used was provided in Borenstein et al. (2009):

$$\mathrm{var}(d_2 - d_1) = V_1 + V_2 - (2r\sqrt{V_1}\sqrt{V_2})$$

where $V_1$ and $V_2$ denote the variance of each outcome measure, $r$ represents the estimated correlation between the immediate and delayed outcomes. We used the average value of correlation ($r$ = .5) to calculate variance in this study; we also measured the outcomes using lower and higher correlations ($r$ = .25 and $r$ = .75) to test the robustness of the results.

## Results

### Description of Included Studies

The meta-analysis included 33 unique research reports (31 journal articles, one master thesis, and one doctoral dissertation) from the year 2012 to 2018. The EFL participants presented a variety of backgrounds, cultures, and ages. Studies were conducted from preschool to university in China, Cyprus, India, Iran, Japan, Saudi Arabia, Taiwan, Thailand, and Turkey. Most studies lasted more than two weeks, utilized random assignment experimental designs, and administered researcher-developed tests. Almost half of the studies used computers and almost half used mobile phones as their technology device. Only one article (Taghizadeh & Porkar, 2018) used tablets as one of the examined technology delivery devices (along with cellphones), thus, this study was categorized under mobile devices and contributed two effect sizes to the mobile versus computer moderator analysis. Specific technology applications employed by each individual study are listed in Table 1.

### Overall Effect Size Estimation

A total of 33 research reports yielding 45 studies ($N = 2374$) were analyzed using random-effects model. We used a random-effects model because, as noted earlier, included studies differed in terms of participant, intervention, etc. characteristics, so we expected significant variability across effect sizes underlying different studies (Borenstein et al., 2009). Overall, technology-assisted L2 vocabulary learning had a large positive effect, $g = .845$, 95% [CI: .625; 1.064] (see Table 2). $Q$ statistic is used to conduct homogeneity analysis and to determine if the heterogeneity is statistically significant. If $Q$ value exceeded the critical value set at $a = .05$ of the $X^2$ distribution, the mean effect size was considered to be statistically heterogeneous. $I^2$, "a measure of inconsistency across the findings of the studies" (Borenstein et al., 2009, p. 118), served to interpret variability percentage attributed to heterogeneity.

Using these criteria, the overall sample in the present study was heterogeneous, $Q(44) = 308.065$, $p < .001$, $I^2 = .86$. The total between-studies variability (i.e., due to heterogeneity among studies) was 86%; 14% of the variance was within-study (i.e., due to sampling error). These results justify the need for moderator effects' analyses. Figure 2 shows the forest plot for the effect size distribution, using random-effects methods.

### Analysis of Moderator Effects

Moderator analyses were organized as follows: (1) participant characteristics, (2) contextual features, and (3) methodology characteristics.

### Participant Characteristics

Table 3 describes the weighted mean effect sizes for participants' educational level and L2 fluency. Grade level was coded as pre/elementary school, secondary, and college. The between-levels difference was not statistically significant, $Q(2) = 1.004$, $p = .605$, which suggests that educational level was a not significant moderator. While 32 studies provided learner's L2 proficiency level as "beginner," "intermediate," or "advanced," 13 studies did not provide learners' L2 proficiency level or L2 placement test results. Thus,

**Table 1.** Characteristics of the 45 studies included in the meta-analysis.

| Study (Year) | N | Grade | Region | Technology |
|---|---|---|---|---|
| *Alemi et al. (2012) | 45 | College | Iran | SMS Messaging |
| Alfadil (2017) | 64 | Secondary | Saudi Arabia | Virtual Reality Game |
| Azabdaftari and Mozaheb (2012) | 80 | College | Iran | SRS and SMS |
| Basal et al. (2016) | 67 | College | Turkey | Computer software |
| Basal et al. (2016) | 54 | College | Turkey | WhatsApp |
| Basöz and Can (2016) | 35 | Preschool | Turkey | Computer software |
| Cavus and Ibrahim (2017) | 37 | Secondary | Cyprus | NEU-CST |
| Franciosi (2017) exp1 | 84 | College | Japan | Computer software |
| Franciosi (2017) exp2 | 97 | College | Japan | Computer software |
| Ghorbani and Jahandar (2015) | 40 | College | Iran | Computer software |
| Hayati et al. (2013) ex1 | 30 | College | Iran | SMS messaging |
| Hayati et al. (2013) ex2 | 30 | College | Iran | SMS messaging |
| *Hirschel and Fritz (2013) exp1 | 78 | College | Japan | Praxised.com |
| *Hirschel and Fritz (2013) exp2 | 114 | College | Japan | Praxised.com |
| Hwang and Chen (2013) | 60 | Elementary | Taiwan | PDAs |
| Jafari and Chalak (2016) | 60 | Secondary | Iran | WhatsApp |
| Khodaparast and Ghafournia (2015) exp1 | 60 | College | Iran | Skype software |
| Khodaparast and Ghafournia (2015) exp2 | 60 | College | Iran | CD |
| Khodaparast and Ghafournia (2015) exp3 | 60 | College | Iran | Skype software and CD |
| Lai (2014) | 40 | Secondary | China | Instant message |
| *Lin et al. (2014) exp1 | 49 | Secondary | Taiwan | Computer software |
| *Lin et al. (2014) exp2 | 55 | Secondary | Taiwan | Computer software |
| Maftoon et al. (2015) | 44 | Secondary | Iran | VTS.S |
| Mirzaei et al. (2016) | 50 | Secondary | Iran | LexisBoard |
| Nejati et al. (2018) exp1 | 20 | Secondary | Iran | CAVI Software Vocaboly |
| Nejati et al. (2018) exp2 | 20 | Secondary | Iran | CAVI Software Vocaboly |
| Nikoopour and Kazemi (2014) exp1 | 72 | College | Iran | Mobile Flashcards |
| Nikoopour and Kazemi (2014) exp2 | 72 | College | Iran | Online Flashcards |
| Ono et al. (2015) | 99 | College | Japan | Mobile Tools |
| Praveen & Rajan,(2013) | 70 | Secondary | India | Graphic Organizer |
| Sharifi et al. (2017) | 66 | College | Iran | e-portfolio |
| *Salavati and Salehi (2016) | 60 | College | Iran | CD |
| *Saran et al. (2012) exp1 | 35 | College | Turkey | MMS |
| *Saran et al. (2012) exp2 | 35 | College | Turkey | Microsoft .NET software |
| *Saran et al. (2012) exp3 | 34 | College | Turkey | MMS |
| *Saran et al. (2012) exp4 | 33 | College | Turkey | Microsoft .NET software |
| Suwantarathip and Orawiwatnakul (2015) | 80 | College | Thailand | SMS Messaging |
| Tabatabaei and Goojani (2012) | 60 | Secondary | Iran | SMS Messaging |
| Taghizadeh and Porkar (2018) exp1 | 30 | College | Iran | SMS Messaging |
| Taghizadeh and Porkar (2018) exp2 | 30 | College | Iran | Tablet |
| Taj et al. (2017) | 122 | College | Saudi Arabia | WhatsApp |
| Young and Wang (2014) | 52 | Elementary | Taiwan | GeCALL System |
| Wu (2014) | 50 | College | China | Java Application |
| Wu (2015a) | 70 | College | China | CET 6 APP |
| Wu (2015b) | 199 | College | China | CET 4 APP |

*Note*: *Indicates a study that included within-study (immediate-delayed posttest) comparison.

**Table 2.** Overall effect sizes of the 33 studies.

| Model | Effect size | | | | | Test of Heterogeneity | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | *k* | *g* | *SE* | 95% CI | *Q* | *df* | *p* | $I^2$ |
| Random | 2374 | 45 | .845* | .112 | [0.63,1.06] | 308.065 | 44 | 0.000 | 85.72 |

*Note. N*: number of participants; *k*: number of independent comparisons; *SE*: standard error; CI: confidence interval; *df*: degrees of freedom. *$p < .05$.

L2 fluency in our study was coded as “beginner,” “intermediate,” “advanced,” or “unreported.” The corresponding between-levels difference was not statistically significant, $Q\ (3) = .354$, $p = .95$, indicating that L2 fluency was not a significant moderator.

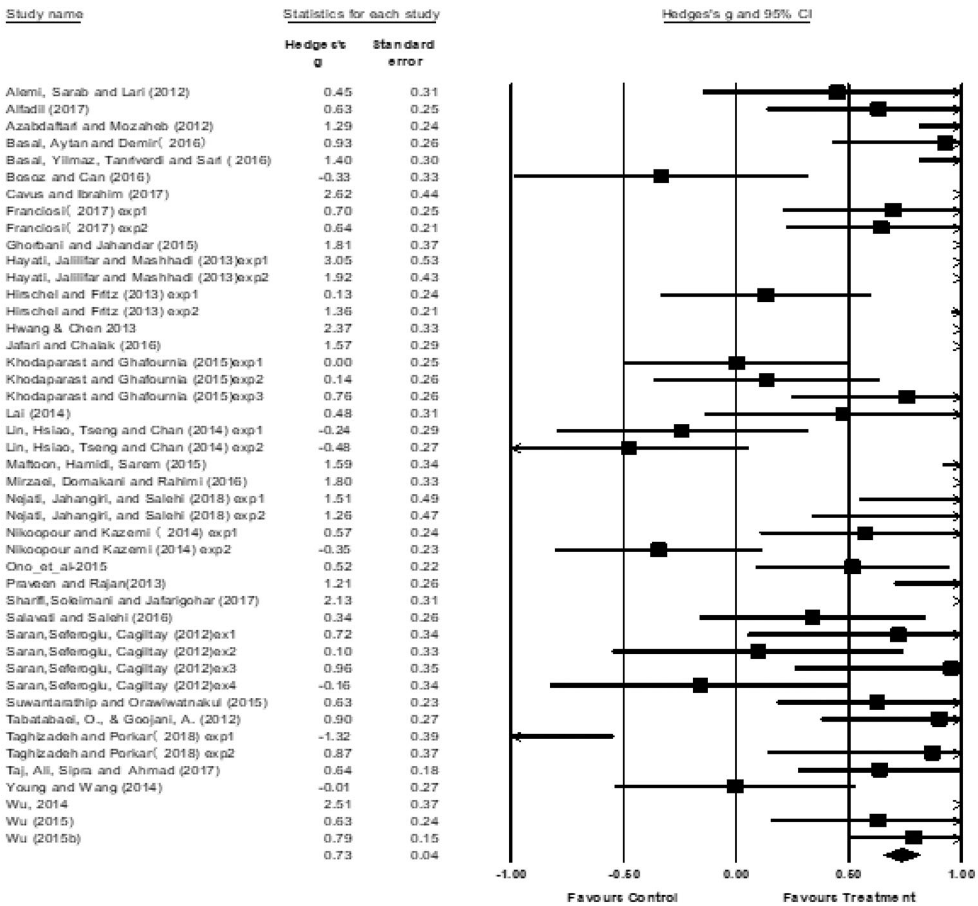


**Figure 2.** Forest plot of effect sizes.

**Table 3.** Weighted mean effect sizes for participant characteristics.

| | | | Effect size | | | Test of heterogeneity | | |
|---|---|---|---|---|---|---|---|---|
| Variable | *N* | *k* | *g* | *SE* | 95% CI | $Q_B$ | *df* | *p* |
| Educational level | | | | | | | | |
| Pre/Elementary | 147 | 3 | 0.68 | 0.83 | [−0.94, 2.30] | | | |
| Secondary | 531 | 12 | 1.06* | 0.25 | [0.56, 1.55] | | | |
| College | 1696 | 30 | 0.78* | 0.12 | [0.54, 1.02] | | | |
| Between levels (QB) | | | | | | 1.004 | 2 | .605 |
| L2 fluency | | | | | | | | |
| Beginner | 441 | 10 | .854* | 0.32 | [0.23, 1.48] | | | |
| Intermediate | 843 | 17 | .786* | 0.15 | [0.50, 1.07] | | | |
| Advanced | 170 | 5 | 1.165 | 0.64 | [−0.09, 2.42] | | | |
| Unreported | 920 | 13 | .813* | 0.18 | [0.46, 1.17] | | | |
| Between levels (QB) | | | | | | 0.354 | 3 | .95 |

*$p < .05$.

## *Contextual Features*

Table 4 describes the weighted mean effect sizes for contextual features. Device was coded as computer or mobile. The between-levels difference for device was statistically significant, $Q$ (1) = 4.31, $p$ = .038, indicating that the device was a significant moderator. Post hoc analysis showed that mobile-assisted vocabulary learning had higher weighted mean effect

**Table 4.** Weighted mean effect sizes for contextual features.

| Variable | $N$ | Effect size $k$ | $g$ | $SE$ | 95% CI | Test of heterogeneity $Q_B$ | $df$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| Device | | | | | | | | |
| Computer | 1157 | 24 | 0.63* | 0.15 | [0.34, 0.93] | | | |
| Mobile | 1217 | 21 | 1.09* | 0.16 | [0.77, 1.40] | | | |
| Between levels (QB) | | | | | | 4.31 | 1 | 0.038 |
| Game | | | | | | | | |
| Game-based | 296 | 5 | 0.49* | 0.13 | [0.24, 0.74] | | | |
| Non game-based | 2078 | 40 | 0.90* | 0.13 | [0.65, 1.15] | | | |
| Between levels (QB) | | | | | | 5.25 | 1 | 0.022 |
| Study duration | | | | | | | | |
| $\leq$ Two weeks | 368 | 9 | 0.95* | 0.22 | [0.52, 1.39] | | | |
| $>$ two weeks | 2006 | 36 | 0.82* | 0.13 | [0.56, 1.07] | | | |
| Between levels (QB) | | | | | | 0.281 | 1 | 0.596 |
| Setting | | | | | | | | |
| Classroom | 1348 | 28 | 0.53* | 0.13 | [0.27, 0.78] | | | |
| On the move | 1026 | 17 | 1.37* | 0.18 | [1.02, 1.71] | | | |
| Between levels (QB) | | | | | | 14.57 | 1 | 0 |
| Test format | | | | | | | | |
| Receptive | 1300 | 28 | 0.62* | 0.15 | [0.33, 0.91] | | | |
| Productive | 645 | 9 | 1.24* | 0.21 | [0.82, 1.66] | | | |
| Both | 429 | 8 | 1.18* | 0.23 | [0.72, 1.64] | | | |
| Between levels (QB) | | | | | | 7.55 | 2 | 0.023 |

*$p < .05$.

size ($g = 1.09$) than did computer-assisted learning ($g = .63$). Further, the effect of game-based technology assisted learning was found to be statistically significant, $Q$ (1) = 5.25, $p$ = .022, which suggests that game was a significant moderator. Non game-based technology ($g$ =.90) had higher effect size than did game-based technology ($g = .49$).

Study duration was coded as less than/equal to two weeks versus more than two weeks, based on prior L2 intervention research (Plonsky, 2011). While there were statistically detectable effect sizes for two types of settings ($g = .95$ and $g = .82$, respectively), the between-levels difference was not statistically significant, $Q$ (1) = .281, $p = .60$, indicating that study duration was not a significant moderator. Studies with no "fixed" educational setting for learning ($N = 17$) were coded as "on the move." The between-levels difference was statistically significant, $Q$ (1) = 14.57, $p < .001$, which suggests that setting was a significant moderator. Post hoc analysis revealed that learning on the move had higher weighted mean effect size ($g = 1.37$) than did learning in the classroom setting ($g = .53$).

The test format was coded as receptive, productive, or both. For example, studies using only multiple-choice questions were coded as "receptive" since they did not ask learners to generate original responses; fill-in-the-blank or translation questions, on the other hand, were coded as "productive" since both asked learners to use L2 knowledge to generate original responses in L2 or L1. Applying both assessment formats in a single study was coded as "both." The between-levels difference was statistically significant, $Q$ (2) = 7.55, $p$ = .023, indicating that the type of test format was a significant moderator for technology-assisted L2 vocabulary learning.

### *Methodological Characteristics*

For methodological characteristics (see Table 5), only reported reliability moderator was statistically significant, $Q$ (1) = 6.317, $p$ = .012. Differences between studies using

**Table 5.** Weighted mean effect sizes for methodology characteristics.

| Variable | $N$ | $k$ | Effect size $g$ | $SE$ | 95% CI | Test of heterogeneity $Q_B$ | $df$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| Research design | | | | | | | | |
| Experimental | 1729 | 32 | 0.91* | 0.16 | [0.38, 0.99] | | | |
| Quasi-experimental | 645 | 13 | 0.69* | 0.15 | [0.63, 1.20] | | | |
| Between levels (QB) | | | | | | 1.123 | 1 | .289 |
| Reliability | | | | | | | | |
| Reported | 1046 | 21 | 0.56* | 0.15 | [0.26, 0.86] | | | |
| Unreported | 1328 | 24 | 1.106* | 0.16 | [0.80, 1.41] | | | |
| Between levels (QB) | | | | | | 6.317 | 1 | .012 |
| Measurement tool | | | | | | | | |
| Research-developed | 2148 | 40 | 0.83* | 0.12 | [0.60, 1.06] | | | |
| Standardized test | 226 | 5 | 0.96* | 0.42 | [0.13, 1.79] | | | |
| Between levels (QB) | | | | | | 0.088 | 1 | .77 |

*$p < .05$.

experimental versus quasi-experimental designs, $Q$ (1) = 1.123, $p$ = .289, and researcher-developed versus standardized tests, $Q$ (1) = .088, $p$ = .77, were not statistically significant.

### *Effects of within-Study Comparisons*

A within-study comparison of mean effects was conducted in the present review to examine whether the invention effects would last over time. There were a total of 10 studies that administered an immediate and a delayed posttest, in which most participants were college students (8 studies) and the rest were secondary students (2 studies). Those 10 studies are marked with an asterisk* in Table 1. Seven studies used computers and three used mobile phones as their technology device, no studies using games conducted within-study comparisons. The average difference between the effect size for delayed and immediate posttest was .22 with variance .008 and stand error .09. The $z$-value is 2.38 with a two-tailed $p$-value of .017. (95% CI [.04, .39]). The combined weighted mean effect size is significantly larger on delayed test ($g$ = .593) than on immediate posttest ($g$ = .377). If the correlation falls into .25 and .75, and the two-tailed $p$-value would be .0516 and .00075. These results suggest that the positive effects of technology-assisted vocabulary learning did not stem from the immediate posttest only and that the positive effects on learning not only did not diminish over time, but on the contrary, significantly increased over time.

## Publication Bias

The articles easy to locate tend to be the ones usually with higher effect sizes than those studies with lower effect sizes (Rosenthal, 1979)—a phenomenon known as the drawer file effect. Specifically, the studies with significant results are easier to find their ways to be published than studies with no significant results. This tendency will lead to the bias that is the Type I error in the published literature and may carry over to the conclusions of the meta-analysis. Even though there are no perfect solutions to this problem, several steps decreasing the bias were taken in this study. First, a vigorous search for studies including both published and unpublished ones was conducted. Second, the potential existence of publication bias was examined via computing with CMA software. The Classic Fail-Safe *N* test was conducted to assess the number of missing studies required to make the effect nonsignificant (set at $a \geq .05$) in the analysis. This test indicated that 4036 studies would need to be found before the cumulative effect would become trivial (set $a = .05$). Moreover, Orwin's fail-safe *N* test is 625, suggesting that 625 studies would be required to nullify the existing overall mean effect size (set $g = .05$). Lastly, the results of Eggers's regression test also showed the absence of publication bias ($p = .06$).

## Discussion

The overall meta-analysis results showed that learners who had access to technology-assisted learning performed better on measures of L2 vocabulary than did learners with no access to technology. The results suggest that if teachers and learners could take advantage of technology, L2 vocabulary could be learned in a more efficient and, arguably, more enjoyable way. This finding is consistent with dual-coding theory (Clark & Paivio, 1991) and interaction hypothesis (Long, 1996) predictions and attests to the explanatory value of these theoretical frameworks in technology-assisted vocabulary learning research. That is, DCT and interaction hypothesis may provide viable explanations as to why students benefit more from computer-mobile software and digital games in their vocabulary learning when learners have access to information through both visual and verbal formats and when their learning is supported by meaning negotiation with the interactive technology. The dual verbal-nonverbal memory and enhanced access to interaction are critical to L2 learners who might find difficulty in internalizing the meaning of new vocabulary given their limited exposure to L2 and account for technology-assisted positive effects on immediate learning and long-term vocabulary retention in our findings.

### *Participant Characteristics*

Moderator analyses indicated that educational level and L2 fluency were not significant moderators. Regarding educational level, although technology-assisted learning was found to be beneficial for secondary and college students, there was no statistically detectable effect associated with pre/elementary school level, suggesting that using technology at this educational level was not associated with any particular benefits for L2 vocabulary learning. This finding, in part consistent with prior research (Abraham, 2008; Chiu, 2013; Mahdi, 2018), suggests that merely relying on technology as a means for educating younger learners without other instructional supports is unlikely to achieve desirable outcomes, at least when it comes to L2 vocabulary learning. It is plausible that preschool and elementary school students may be too young to use technology effectively to support their learning. Alternatively, L2 words to be learned at the pre/elementary level may represent frequently used words already present across multiple contexts (TV shows, cartoons) and thus could be mastered without the aid of additional technology. However, caution is needed in interpreting and generalizing these results since only three studies included pre/elementary-school-aged students. The three studies involved computer-, mobile-, and game-mediated interventions. Thus, additional studies are needed to examine the effects of technology-assisted vocabulary learning for younger students, ideally examining the potentially moderating effects of word frequency and exposure (Puimège & Peters, 2019).

Regarding L2 fluency, although technology-assisted learning was found to be beneficial for beginner and intermediate proficiency learners, no statistically detectable effect was found for advanced learners. Technology failing to facilitate L2 vocabulary learning for advanced learners might be attributed to higher proficiency of these learners, which partially aligns with the expertise reversal effect (Kalyuga et al., 2003). This finding, however, is only tentative in nature as 13 of the identified studies did not report participants' English proficiency level.

### Contextual Features

Among contextual features, device type, setting, game condition, and test format were statistically significant moderators. Specifically, advantages were found for mobile devices and on-the-move learning, suggesting that L2 vocabulary learning may be most efficient when students use mobile phones and are not restricted by classroom settings. These findings also provide support for Long's (1996) interaction hypothesis in language learning. Compared with computers, mobile devices provide students with a more personalized interface and interaction with peers and teachers. Further, the portability and interconnectivity of mobile devices enhance the integration of formal and informal learning, which can promote learners' interest and thus foster comprehension and retention. Consistent with Chiu's (2013) findings, game condition was also a significant moderator, with students learning under non-game-based technology condition significantly outperforming those learning under game-based technology condition. That said, as suggested by statistically detectable effect sizes, both game-based and non-game-based technology-assisted interventions are still more beneficial for L2 vocabulary learning than the traditional methods are. Test format was the last statistically significant moderator for technology assisted L2 vocabulary learning four in the present study. In particular, productive tasks tests and tests integrating receptive and productive tasks and were associated with larger effect sizes than were receptive tasks tests. This finding suggests that the full potential of technology-assisted L2 vocabulary learning may be better captured by more comprehensive rather than multiple-choice assessments. No evidence of statistically significant moderation due to study duration was found. In other words, interventions of less than or more than two weeks were associated with statistically detectable effect sizes, indicating that technology enhanced L2 vocabulary learning regardless of how long the study intervention lasted, highlighting the efficacy of technology-assisted learning.

### Effects of within-Study Comparisons

The results from the current study extended Abraham's (2008) research by providing further evidence of long-term effectiveness of technology-assisted strategies on vocabulary learning and retention. We were able to examine longitudinal associations between computer- and mobile-assisted interventions and students' vocabulary acquisition across

10 independent study samples and found significant gains from immediate to delayed posttest. The results have important implications, including supporting the usefulness of incorporating computer- and mobile-assisted learning into EFL courses. Given that participants from eight studies were college students, we may conclude that students at this educational level have the needed autonomy to use computer or mobile software to enhance their vocabulary learning and retention, thus transforming traditional language learning into a more student-centered environment. Since there has been no studies in the current review investigating the effects of games on students' long-term vocabulary retention, future research on game-assisted vocabulary learning should examine not only immediate learning, but also long-term retention of new vocabulary.

### *Methodological Characteristics*

Methodological characteristics of studies contributed to some additional variation in the overall findings. That is, although across methodological characteristics, all studies included in the present meta-analysis yielded statistically detectable effects, regardless of strengths or weaknesses of their designs, only reliability reported was a statistically significant moderator. The statistically significantly higher effect size associated with the weaker design studies not reporting reliabilities suggests a need for practitioners to be cautious when interpreting individual study findings and for researchers to take greater care in designing and reporting their work.

## Conclusion

Teaching and learning L2 vocabulary could be a great challenge facing EFL teachers and their students. As students may feel frustrated with the vast numbers of L2 words to be learned and remembered to enable their effective comprehension and communication in a new language, teachers may experience difficulties in generating and maintaining student motivation to learn L2 vocabulary. Thus, vocabulary learning and long-term retention would always be important goals for language learners and their educators. Findings from this meta-analysis showed that, overall, technology-assisted L2 vocabulary learning is more beneficial than non-technology-assisted instruction. Further, this meta-analysis showcases the advantages of L2 vocabulary learning through a host of different technologies and indicates that technology can enhance learners' long-term retention of new words. More pronounced advantages were found for mobile devices and on-the-move learning, suggesting that L2 vocabulary learning may be most efficient when students use mobile phones and are not restricted by classroom settings. The results also highlight several important variables—device type, game condition, setting, test format, and reported reliability—as moderators of the technology-assisted vocabulary learning effectiveness. Therefore, these variables are suggested to be considered when planning instruction in technology-assisted L2 vocabulary learning.

## ORCID

Zhe Wang http://orcid.org/0000-0001-6476-9538
Yuliya Ardasheva http://orcid.org/0000-0001-9494-0316

## References

*Indicates study included in this 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. https://doi.org/10.1080/09588220802090246

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

*Alemi, M., Sarab, M. R. A., & Lari, Z. (2012). Successful learning of academic word list via MALL: Mobile assisted language learning. *International Education Studies*, *5*(6), 99–109. https://doi.org/10.5539/ies.v5n6p99

*Alfadil, M. M. (2017). Virtual *reality game classroom implementation: Teacher perspectives and student learning outcomes* [Unpublished PhD dissertation]. University of Northern Colorado.

Ardasheva, Y., Hao, T., & Zhang, X. (2019). Pedagogical implications of current SLA research for vocabulary skills. In N. Polat, P. MacIntyre, & T. Gregersen (Eds.), *Research-driven pedagogy: Implications of L2A theory and research for the teaching of language skills* (pp. 125–144). Routledge.

*Azabdaftari, B., & Mozaheb, M. A. (2012). Comparing vocabulary learning of EFL learners by using two different strategies: Mobile learning vs. flashcards. *The EuroCALL Review*, *20*(2), 47–59. https://doi.org/10.4995/eurocall.2012.11377

*Basal, A., Aytan, T., & Demir, I. (2016). Teaching vocabulary with graphic novel. *English Language Teaching*, *9*(9), 95–109. https://doi.org/10.5539/elt.v9n9p95

*Basal, A., Yilmaz, S., Tanriverdi, A., & Sari, L. (2016). Effectiveness of mobile applications in vocabulary teaching. *Contemporary Educational Technology*, *7*(1), 47–59. https://doi.org/10.30935/cedtech/6162

*Basöz, T., & Can, D. T. (2016). The effectiveness of computers on vocabulary learning among preschool children: A semiotic approach. *Cypriot Journal of Educational Sciences*, *11*(1), 02–08. https://doi.org/10.18844/cjes.v11i1.266

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

*Cavus, N., & Ibrahim, D. (2017). Learning English using children's stories in mobile devices. *British Journal of Educational Technology*, *48*(2), 625–641. https://doi.org/10.1111/bjet.12427

Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology*, *49*(1), 69–77. https://doi.org/10.1111/bjet.12526

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

Cho, K., Lee, S., Joo, M. H., & Becker, B. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences*, *8*(3), 105. https://doi.org/10.3390/educsci8030105

Chwo, G. S. M., Marek, M. W., & Wu, W. C. V. (2018). Meta-analysis of MALL research and design. *System*, *74*, 62–72. https://doi.org/10.1016/j.system.2018.02.009

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF013200

Du, H. (2004). Reflections on vocabulary size of Chinese university studies. *International Education Journal*, *5*(4), 571–581.

*Franciosi, S. J. (2017). The effect of computer game-based learning on FL vocabulary transferability. *Educational Technology & Society*, *20*(1), 123–133.

*Ghorbani, T., & Jahandar, S. (2015). The effect of E-learning on Iranian intermediate EFL learners word retention. *International Research Journal of Applied and Basic Sciences*, *9*(7), 103–106.

Gibson, C. (2016). Bridging English language learner achievement gaps through effective vocabulary development strategies. *English Language Teaching*, *9*(9), 134–138. https://doi.org/10.5539/elt.v9n9p134

Gorjian, B., Moosavinia, S. R., Kavari, K. E., Asgari, P., & Hydarei, A. (2011). The impact of asynchronous computer-assisted language learning approaches on English as a foreign language high and low achievers' vocabulary retention and recall. *Computer Assisted Language Learning*, *24*(5), 383–391. https://doi.org/10.1080/09588221.2011.552186

*Hayati, A., Jalilifar, A., & Mashhadi, A. (2013). Using short message service (SMS) to teach English idioms to EFL students. *British Journal of Educational Technology*, *44*(1), 66–81. https://doi.org/10.1111/j.1467-8535.2011.01260.x

Hassan Taj, I., Sulan, N., Sipra, M., & 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.

*Hwang, W. Y., & Chen, H. S. (2013). Users' familiar situational contexts facilitate the practice of EFL in elementary schools with mobile devices. *Computer Assisted Language Learning*, *26*(2), 101–125. https://doi.org/10.1080/09588221.2011.639783

*Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, *41*(3), 639–653. https://doi.org/10.1016/j.system.2013.07.016

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load Hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558. https://doi.org/10.1111/0023-8333.00164

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

*Jafari, S., & Chalak, A. (2016). The role of WhatsApp in teaching vocabulary to Iranian EFL learners at junior high school. *English Language Teaching*, *9*(8), 85–92. https://doi.org/10.5539/elt.v9n8p85

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

*Khodaparast, F., & Ghafournia, N. (2015). The effect of asynchronous/synchronous approaches on English vocabulary achievement: A study of Iranian EFL learners. *English Language Teaching*, *8*(4), 117–127. https://doi.org/10.5539/elt.v8n4p117

*Lai, W. H. (2014). *Using mobile instant messenger (WhatsApp) to support second language learning* [Unpublished Master Thesis]. Hong Kong University, Hongkong, China.

Li, J., Cummins, J., & Deng, Q. (2017). The effectiveness of texting to enhance academic vocabulary learning: English language learners' perspective. *Computer Assisted Language Learning*, *30*(8), 816–843. https://doi.org/10.1080/09588221.2017.1366923

*Lin, C.-C., Hsiao, H.-S., Tseng, S.-P., & Chan, H.-J. (2014). Learning English vocabulary collaboration in a technology-supported classroom. *The Turkish Online Journal of Educational Technology*, *13*(1), 162–173.

Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Long, M.H. (1996). The role of linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research*, *56*(1), 134–154. https://doi.org/10.1177/0735633117698826

*Maftoon, P., Hamidi, H., & Sarem, S. (2015). The effects of CALL on vocabulary learning: A case of Iranian intermediate EFL learners. *Brain: Broad Research in Artificial Intelligence and Neuroscience*, *3*(4), 19–30.

Marley, S. C., & Levin, J. R. (2011). When are prescriptive statements in educational research justified? *Educational Psychology Review*, *23*(2), 197–206. https://doi.org/10.1007/s10648-011-9154-y

Metros, S. E. (2008). The educator's role in preparing visually literate learners. *Theory into Practice*, *47*(2), 102–109. https://doi.org/10.1080/00405840801992264

*Mirzaei, A., Domakani, M. R., & Rahimi, S. (2016). Computerized lexis-based instruction in EFL classrooms: Using multi-purpose LexisBOARD to teach L2 vocabulary. *ReCALL*, *28*(1), 22–43. https://doi.org/10.1017/S0958344015000129

Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning*, *29*(7), 1220–1237. https://doi.org/10.1080/09588221.2016.1210645

Moody, S., Hu, X., Kuo, L. J., Jouhar, M., Xu, Z., & Lee, S. (2018). Vocabulary instruction: A critical analysis of theories, research, and practice. *Education Sciences*, *8*(4), 180–122. https://doi.org/10.3390/educsci8040180

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

*Nejati, E., Jahangiri, A., & Salehi, M. R. (2018). The effect of using computer-assisted language learning (CALL) on Iranian EFL learners' vocabulary learning: An experimental study. *Cypriot Journal of Educational Sciences*, *13*(2), 113–362. https://doi.org/10.18844/cjes.v13i2.752

*Nikoopour, J., & Kazemi, A. (2014). Vocabulary learning through digitized & non-digitized flashcards delivery. *Procedia - Social and Behavioral Sciences*, *98*, 1366–1373. https://doi.org/10.1016/j.sbspro.2014.03.554

*Ono, Y., Ishihara, M., & Yamashiro, M. (2015). Blended instruction utilizing mobile tools in English teaching at colleges of technology. *Electrical Engineering in Japan*, *192*(2), 1–11. https://doi.org/10.1002/eej.22717

Paivio, A. (2006, September). Dual coding theory and education [Paper presentation]. Pathways to Literacy Achievement for High Poverty Children Conference, University of Michigan.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, *61*(4), 993–1038. https://doi.org/10.1111/j.1467-9922.2011.00663.x

*Praveen, S. D., & Rajan, P. (2013). Using graphic organizers to improve reading comprehension skills for the middle school ESL students. *English Language Teaching*, *6*(2), 155–170.

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, *69*(4), 943–977. https://doi.org/10.1111/lang.12364

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641

*Salavati, M., & Salehi, H. (2016). Impact of using instructional video games as an on EFL learners' vocabulary retention. *Universal Journal of Educational Research*, *4*(12), 2724–2728.

*Saran, M., Seferoglu, G., & Cagiltay, K. (2012). Mobile language learning: Contribution of multimedia messages via mobile phones in consolidating vocabulary. *The Asia-Pacific Education Researcher*, *21*(1), 181–190.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

*Sharifi, M., Soleimani, H., & Jafarigohar, M. (2017). E-portfolio evaluation and vocabulary learning: Moving from pedagogy to andragogy. *British Journal of Educational Technology*, *48*(6), 1441–1450. https://doi.org/10.1111/bjet.12479

Stockwell, G. (2011). Online approaches to learning vocabulary: Teacher-centered or learner-centered? *International Journal of Computer-Assisted Language Learning and Teaching*, *1*(1), 33–44. https://doi.org/10.4018/ijcallt.2011010103

Sung, Y. T., Chang, K. E., & Yang, J. M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review*, *16*, 68–84. https://doi.org/10.1016/j.edurev.2015.09.001

*Suwantarathip, O., & Orawiwatnakul, W. (2015). Using mobile-assisted exercises to support students' vocabulary skill development. *Turkish Online Journal of Educational Technology-TOJET*, *14*(1), 163–171.

*Tabatabaei, O., & Goojani, A. H. (2012). The impact of text-messaging on vocabulary learning of Iranian EFL learners. *Cross-Cultural Communication*, *8*(2), 47.

*Taghizadeh, M., & Porkar, P. (2018). Tablet, flashcard and SMS and their effects on EFL learners' attitudes and vocabulary Knowledge. *International Journal of English Language and Translation Studies*, *6*(1), 105–118.

*Taj, I., Ali, F., Sipra, M., & Ahmad, W. (2017). Effect of technology enhanced language learning on EFL reading comprehension at tertiary level. *Arab World English Journal*, *8*(1), 108–129. https://doi.org/10.24093/awej/vol8no1.9

Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education*, *125*, 345–357. https://doi.org/10.1016/j.compedu.2018.06.020

Woloshyn, V. E., Paivio, A., & Pressley, M. (1994). Use of elaborative interrogation to help students acquire information consistent with prior knowledge an information inconsistent with prior knowledge. *Journal of Educational Psychology*, *86*(1), 79–89. https://doi.org/10.1037/0022-0663.86.1.79

*Wu, Q. (2014). Learning ESL vocabulary with smartphones. *Procedia - Social and Behavioral Sciences*, *143*, 302–307. https://doi.org/10.1016/j.sbspro.2014.07.409

*Wu, Q. (2015a). Designing a smartphone app to teach English (L2) vocabulary. *Computers & Education*, *85*, 170–179. https://doi.org/10.1016/j.compedu.2015.02.013

*Wu, Q. (2015b). Pulling mobile assisted language learning (MALL) into the mainstream: MALL in broad practice. *PloS One*, *10*(5), e0128762. https://doi.org/10.1371/journal.pone.0128762

Valentine, J. C., & Cooper, H. (2003). *What works clearinghouse study design and implementation assessment device (Version 1.1)*. U.S. Department of Education.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods*, *13*(2), 130–149. https://doi.org/10.1037/1082-989X.13.2.130

Yousefi, M. H., & Biria, R. (2018). The effectiveness of L2 vocabulary instruction: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, *3*(1), 21. https://doi.org/10.1186/s40862-018-0062-2

*Young, S. S. C., & Wang, Y. H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Journal of Educational Technology & Society*, *17*(3), 239–251.

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