Education and Information Technologies (2024) 29:4531–4563
https://doi.org/10.1007/s10639-023-11969-1

Second and Foreign Language Vocabulary Learning through Digital Reading: A Meta-Analysis

Tong Zhu[1] · **Yanhui Zhang**[1] · **Derek Irwin**[1]

Received: 3 March 2023 / Accepted: 8 June 2023 / Published online: 8 July 2023

Abstract
Digital reading facilitates L2 development by allowing anytime-anywhere learning with various digital resources. Although there has been increasing research exploring the role of digital reading on L2 vocabulary learning, synthesized evidence regarding the overall facilitating power of digital reading is still lacking. This meta-analysis aggregates 21 important empirical studies published within the last 20 years and indexed in scholarly recognized databases, so as to provide a comprehensive panoramic assessment of how effectively digital reading has enhanced second and foreign language vocabulary acquisition with diversified learner backgrounds and learning environments. A total of 77 effect sizes were yielded across different studies, and random-effect modeling was employed for analyzing the study-level heterogeneity and sub-group variability. Results demonstrate that digital reading had an overall significant effect ($d_{\text{immediate}}=1.45$, $p<.01$; $d_{\text{deleayed}}=2.98$, $p<.01$) on L2 vocabulary learning of between-subject studies. For within-subject studies, digital reading was found to have an upper-medium ($d=1.39$, $p<.01$) and medium ($d=.86$, $p<.01$) effect on immediate and delayed L2 vocabulary post-tests. Moderating factor analysis results show that L2 proficiency, vocabulary test formats, and digital resources could robustly explain the variance of effect sizes. The findings have strong pedagogical implications on the effective design for digital reading tasks, including the development of adaptive learning algorithms and personalized lexical glosses. Recommendations for future research in the field are provided by pinpointing where to improve in terms of experimental design and the focus of the learner group.

Keywords Second and Foreign Language · L2 Vocabulary Learning · Digital Reading · Lexical Glosses · Meta-Analysis

1 Introduction

Second and foreign language (L2) reading offers pedagogical advantages, such as providing examples of creative language use and abundant cultural information in authentic contexts. In both second- and foreign-language environments, reading has long been

Extended author information available on the last page of the article

recognized as the primary source of L2 vocabulary learning (Boers, 2022; Huckin & Coady, 1999; Krashen, 1989; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb & Chang, 2015). With the increasing popularity of digital technology in language education (Golonka et al., 2014), digital reading has become ubiquitous for L2 reading. This study broadly defines 'digital reading' as onscreen reading with or without access to digital resources outside the reading texts. Given the crucial role of vocabulary in all language use (Schmitt et al., 2017, 2021) and successful language learning (Devine, 1988, p. 49; Laufer, 2003), the effectiveness of digital reading for L2 vocabulary learning has been the subject of much research interest for the last twenty years (Akbulut, 2007; AkbuSeileek, 2011; Hsieh et al., 2012; Hsu et al., 2013; Lee et al., 2016; Ruiz et al., 2021; Wang, 2016). Constantly evolving e-learning and artificial intelligence (AI) tools present new opportunities and challenges for effective L2 vocabulary learning through digital reading, which allows not only anytime-and-anywhere reading but also access to various digital resources, including but not limited to lexical priming, quizzes, audio narrations, lexical glosses, and personalized reading systems.

Although L2 vocabulary learning through reading can be enhanced by different digital resources, it remains unknown which resource is most facilitative to that learning. The present study rigorously synthesizes the overall effect of digital reading on L2 vocabulary learning. Twenty-one prominent empirical studies in this meta-analysis create a panoramic view of the cumulative effect and address the moderating effect regarding the potential variables: 1) L2 proficiency, 2) test formats, and 3) digital resources. These variables are identified as potential moderators based on previous L2 reading and vocabulary learning research findings. Understanding the relationship between learners' individual differences, research designs, and digital affordances has important implications for maximizing the effectiveness of digital reading for L2 vocabulary learning. In addition, the results provide critical insights for future research directions.

2 Literature Review

2.1 Previous Meta-Analyses

Several attempts have been made to synthesize studies comparing reading from paper and digital devices. Delgado et al. (2018) include 38 between-subject studies and 16 within-subject studies that were published in seven databases between 2000 and 2017. The mean effect size ($g = -0.21$) suggested that paper-based reading outperformed digital reading. Clinton (2019) also conducted a meta-analysis comparing reading from paper and screens, which includes 29 studies that were published in seven databases between 2008 and 2018. These studies overall generate 33 independent effect sizes for onscreen and paper-based reading performance. Results showed that onscreen reading had a negative effect on reading performance compared to paper-based reading ($g = -0.25$). Although previous meta-analyses have revealed an advantage of paper-based reading over

digital reading, the generalizability of this finding is limited by measurements and resources. These meta-analyses primarily focused on measuring reading comprehension, whereas other learning outcomes, such as vocabulary acquisition, were under-explored. Without sufficient vocabulary skills, the concepts underlying given words cannot be known, leaving the knowledge of syntax and discourse almost useless (Schmitt et al., 2021). Receiving abundant and adequate input during reading is one of the most efficient ways to improve vocabulary skills (Krashen, 1989). Therefore, it is necessary to explore the facilitative potential of digital reading for vocabulary learning. The learning-enhancing power of digital reading was further underestimated by the fact that digital resources outside the reading texts were largely neglected.

To date, only a few meta-analyses have been conducted on L2 vocabulary learning through digital reading, with a specific focus on digital resources outside the reading texts. For example, Abraham (2008) conducted a meta-analysis to investigate the overall effects of digital reading with lexical glosses on L2 vocabulary learning. The study analyzed 11 published papers up to 2007. Digital reading showed a large mean weighted effect on immediate ($d = 1.40$) and delayed ($d = 1.25$) vocabulary post-tests compared to control groups without access to digital glosses. Later, Yun (2011) ran a similar meta-analysis of 10 papers published between 1990 and 2009, and the study found a medium positive effect ($g = 0.46$) of digital reading with lexical glosses on L2 vocabulary learning.

Although revealing, these previous meta-analytic studies have several drawbacks needing to be addressed. First, these studies were conducted more than ten years ago and reported inconsistent results (e.g., Abraham, 2008; Yun, 2011). Hence, the research area calls for up-to-date systematic reviews with more recent empirical studies focusing on vocabulary learning through digital reading. Second, despite recent review studies on L2 vocabulary learning and lexical glosses (Boers, 2022; Ramezanali et al., 2021; Vahedi et al., 2016; Yanagisawa et al., 2020), none were specifically conducted in the digital context. Moreover, L2 vocabulary learning through reading can be facilitated by not only lexical glosses but also other digital resources such as personalized reading systems (Wang, 2016). Thus, the effect of digital reading on L2 vocabulary learning should be synthesized and compared from a more holistic perspective. To fill these gaps, the current study consolidates the literature on the accuracy and robustness of L2 vocabulary learning through digital reading. Meanwhile, this meta-analysis focuses on potential moderating factors for a more thorough understanding of the role of digital reading in L2 vocabulary learning. The following section provides a detailed discussion and justification for the selection of the influential moderator variables, including differences in learner variables, research designs, and digital affordances.

2.2 Potential Moderator Variables

2.2.1 Learner Variables: Individual Differences in L2 Proficiency

Successful L2 vocabulary learning through reading primarily relies on text comprehension and word inference (Boers, 2022; Krashen, 1989). Whether a text can be sufficiently

understood and whether novel words can be successfully inferred are influenced by the learner and textual factors. Learner factors mainly refer to learners' individual differences in L2 proficiency. When learners notice an unknown word, they usually need to pause to make inferences or find references to the word. Those with lower L2 proficiency may have difficulty inferring and identifying the correct word meaning within the reading context (Bengeleil & Paribakht, 2004). It seems that more proficient learners acquire L2 vocabulary more efficiently through reading. For example, Vahedi et al. (2016) found that L2 proficiency was a statistically significant moderator variable ($Q=6.53$, $p<0.05$). In particular, compared to beginners ($g=0.75$), intermediate ($g=0.85$) and advanced learners ($g=0.82$) more efficiently learned L2 vocabulary through reading with lexical glosses. However, Yun (2011) found that onscreen reading with computerized lexical glosses was most effective for beginners ($g=0.70$) and least effective for intermediate learners ($g=0.23$). It appears that the relationship between language proficiency and learning efficiency varies with reading settings and designs. As suggested by Yanagisawa et al. (2020), more research needs to be conducted to investigate the interaction between different reading designs and L2 proficiency. To this end, we include learners' individual differences in language proficiency as a moderator variable.

2.2.2 Research Designs: Vocabulary Test Formats

Another factor that influences text comprehension and vocabulary learning is the textual factor, which mainly refers to the density of novel words and the number of word recurrences. Comprehensible reading texts require learners to recognize and decode a minimum of 95% of the words in a text (Laufer, 1997; Nation, 2013). If the ratio of novel words exceeds 5%, learners may not be able to obtain sufficient contextual clues to infer their meaning (Laufer, 2020). However, due to the fact that word learning is an incremental process (Milton, 2009; Schmitt, 2010) and that word knowledge is a multifaceted construct (Fitzpatrick & Clenton, 2017; Nation, 2013), correct inference of word meaning is usually inadequate for efficient vocabulary learning. For one thing, a novel word must reappear several times before being learned incidentally through reading (Chen & Truscott, 2010; Pellicer-Sánchez, 2016; Waring & Takaki, 2003). A new meeting with a word could extend or consolidate the lexical knowledge gained from previous meetings (Webb & Chang, 2015). For another, the acquisition of receptive vocabulary knowledge usually comes before productive vocabulary knowledge, rendering word recall more difficult to be acquired than word recognition (Laufer & Goldstein, 2004; Laufer & Paribakht, 1998; González-Fernández & Schmitt, 2020). Therefore, different test formats may have a moderating effect on vocabulary gains through reading. In a meta-analysis on L2 vocabulary learning with lexical glosses, Ramezanali et al. (2021) found no significant difference across the four test formats ($Q=0.29$, $p>0.05$), i.e., form recall ($g=0.44$), form recognition ($g=0.61$), meaning recall ($g=0.35$), and meaning recognition ($g=0.49$), though other formats such as vocabulary knowledge scale were not explored. These results suggest a need for further investigation of the relationship between different aspects of word knowledge and vocabulary test formats. This meta-analysis sets out to explore the potential moderating effect of test formats on L2 vocabulary learning through digital reading.

2.2.3 Digital Affordances: Facilitating Digital Resources

Vocabulary learning through reading can be facilitated by digital resources, as they enable anytime-and-anywhere learning. More importantly, digital resources cater to the learner and textual factors. For example, digital reading can present reading texts with lexical glosses, which compensate for limited vocabulary size by providing first language (L1) or L2 explanations of word meanings embedded in or hyperlinked to texts, often in bold or colored forms (Chen & Yen, 2013; Huang, 2018). Based on Schmidt's (1990) noticing hypothesis, lexical glosses enhance vocabulary learning as the bold or colored forms draw learners' attention to the glossed words (Rouhi & Mohebbi, 2012; Yanguas, 2009). With the explanations of word meanings, learners can easily obtain correct form-meaning connections and better comprehend texts (Yanagisawa et al., 2020). Digital reading also can present reading texts with lexical priming, which momentarily exposes learners to a formal priming stimulus before displaying the target word (Liu & Leveridge, 2017). For example, the word 'fake' is formally similar to the target word 'fate' and thus can be selected as the formal priming stimulus. Briefly presenting the stimulus before the target word can pre-activate learners' lexical knowledge and enhance word recognition (Liu & Leveridge, 2017). With recent advances in e-learning tools, digital reading allows learners to access multiple digital resources simultaneously. For example, learners can use a reading system to access e-dictionaries, lexical glosses, quizzes, and/or audio narration while reading onscreen. Further, natural language processing tools and adaptive algorithms afford AI reading systems, such as personalized reading systems, which can analyze and accumulate learner profiles while recommending the most appropriate reading materials. The recommended reading materials are designed to contain new words they have encountered lately to reinforce word learning and suit learners' language proficiency. In other words, personalized reading systems provide comprehensible texts while increasing the number of word recurrences. Although empirical research respectively showed the effectiveness of L2 vocabulary learning through onscreen reading with access to lexical glosses (AkbuSeileek, 2011; Khezrlou, 2019; Khezrlou & Ellis, 2017), multiple digital resources (Gorjian et al., 2011; Johnson & Heffernan, 2006; Proctor et al., 2007), or personalized reading systems (Hsieh et al., 2012; Hsu et al., 2013; Wang, 2016), the most effective design of digital reading remains unknown. This is one major gap that this meta-analysis aims to address by conducting a moderator analysis to compare the effectiveness of onscreen reading with access to different digital resources for L2 vocabulary learning.

2.3 Research Questions

The current meta-analysis intends to investigate L2 vocabulary learning through digital reading from a holistic perspective. In addition to the overall effect of digital reading, three moderator variables were included and analyzed. The following two research questions guided this study:

1) What is the overall effect of digital reading on L2 vocabulary learning?
2) How is the effect of digital reading on L2 vocabulary learning moderated by L2 proficiency, test formats, and digital resources?

3 Methodology

3.1 Literature Search and Retrieval

We conducted an exhaustive literature search to retrieve quantitative studies with a within-subject or between-subject design that explored the effects of digital reading on L2 vocabulary learning. We aimed to search and retrieve high-quality academic literature written in English and published in the last twenty years. Eight scholarly recognized databases were used for potential literature search: Academic Search Premier, ERIC, JSTOR, LLBA, PsycArticles, PsycINFO, Scopus, and Web of Science. Various combinations of keywords were used for each database: *computerized* OR *mobile* OR *digital* OR *electronic* OR *personalized* OR *adaptive* OR *intelligent* OR *hypertext* OR *hypermedia* AND *reading* AND *vocabulary* AND *learning*. To ensure that all relevant literature has been included, we consulted the reference sections of previous review studies. Bibliographies from retrieved studies were also cross-referenced.

3.2 Literature Screening and Reviewing

The title and abstract of all records were initially screened independently by all authors. Afterward, full texts of these initially screened records were retrieved and uploaded at Endnote with duplicates manually removed. Each text was further examined by all authors. Records meeting the inclusion and exclusion criteria were finally identified as eligible for data extraction and coding. The inclusion and exclusion criteria are presented below. Any discrepancies during literature screening and reviewing were resolved by discussion and consensus.

Studies published in the last twenty years were included.
Studies written in English were included.
Studies with vocabulary measurements were included.
Studies provided with means, sample sizes, and standard deviations for calculating Cohen's *d* were included.
Within-subject studies investigating L2 vocabulary learning through digital reading with a pretest–posttest design were included.
Between-subject studies were included when L2 vocabulary gains through onscreen reading without access to digital resources outside the actual texts were evaluated against the gains of a comparison group, in which learners read on papers with the same instruction.
Between-subject studies were included when L2 vocabulary gains through onscreen reading with access to an extra digital resource outside the actual texts were evaluated against the gains of a comparison group, in which learners receive the same instruction but read onscreen without access to the digital resource.
Case or review studies were excluded.
Studies without accessible full texts were excluded.

Studies which focused on participants who were diagnosed with learning disabilities, such as dyslexia, were excluded.

Following the *Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement* (Page et al., 2021), Fig. 1 shows detailed information about what the authors did and found during the process of literature screening and reviewing. As shown in Fig. 1, 21 studies, asterisked in the References section, were finally included in this meta-analysis.

3.3 Data Coding

Data coding is essential for meta-analysis, whereby various information from studies is translated into a standardized format on a coding scheme table (Plonsky & Oswald, 2015, p. 246). Microsoft Excel was used for data recording and coding. As shown in Table 1, the present study coded all eligible studies with three main coding categories. First, the information on the study context (i.e., publication characteristics and learner variables) was collected and coded if applicable. Secondly, the information on the research design (i.e., between-subject or within-subject research design, text readability, digital resources outside the reading texts, and vocabulary measurements) was also collected and coded if applicable. Finally, descriptive statistics of sample sizes (N), mean scores (M), and standard deviations (SD) were utilized for calculating the effect sizes.

Two independent coders were involved in the coding process to ensure the reliability of the research. To identify the strength of agreement between the two coders, Cohen's Kappa statistic (k) was calculated using SPSS software and used to determine the inter-coder reliability. Table 2 presents the criteria for interpreting the magnitudes of k values. Kappa is often 'presented along with the agreement rate, which

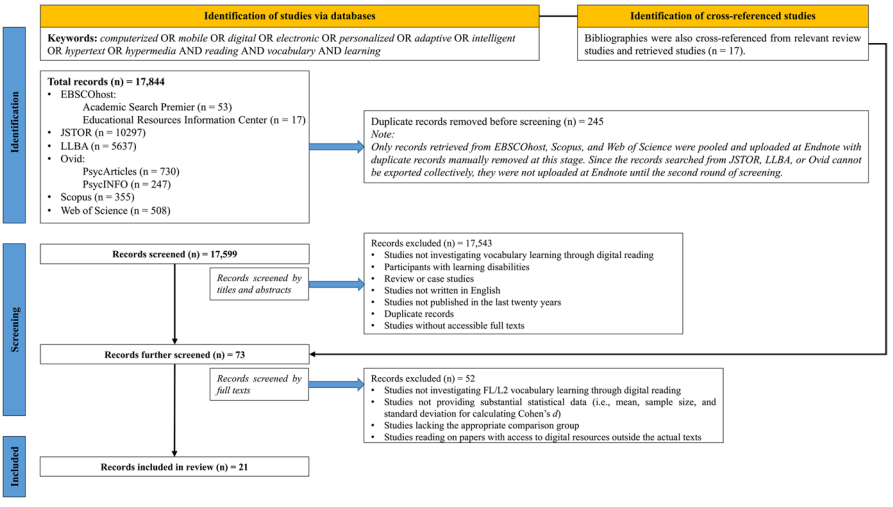

Fig. 1 PRISMA Flow Chart for Literature Searching and Reviewing

Table 1 Coding Book

Major Category	Subcategory	Coding	
Publication Characteristics	Author/ Publication year/ Journal/ Title		
Learner Variable	L2 proficiency	1 = not reported; 2 = beginning; 3 = intermediate; 4 = advanced; 5 = various	
Research Design	Between-subject or within-subject study		
	Comprehensible or challenging reading texts		
	Test format(s)	1 = not reported; 2 = meaning recognition; 3 = meaning recall; 4 = form recognition; 5 = form recall; 6 = VKS (vocabulary knowledge scale); 7 = mixed formats of form recognition, form recall, and meaning recognition	
	Digital resource(s)	1 = onscreen reading without access to digital resources outside the actual texts; 2 = lexical priming; 3 = personalized reading system; 4 = lexical gloss; 5 = multiple resources, including lexical glosses plus quizzes or lexical glosses plus quizzes and audio narrations	
Descriptive Statistics		Within-subject study	Between-subject study
	Sample size (N)	N	$N_{control\ group}$ and $N_{experimental\ group}$
	Mean score (M)	$M_{pretest}$, $M_{immediate\ posttest}$, and $M_{delayed\ posttest}$	$M_{immediate\ posttest}$ and $M_{delayed\ posttest}$ of the experimental group and the control group
	Standard deviation (SD)	$SD_{pretest}$, $SD_{immediate\ posttest}$, and $SD_{delayed\ posttest}$	$SD_{immediate\ posttest}$ and $SD_{delayed\ posttest}$ of the experimental group and the control group

Table 2 Interpretation of Cohen's Kappa Statistic (k) for the Strength of Agreement (Landis & Koch, 1977)

Cohen's Kappa statistic (*k*)	Strength of agreement
< 0.00	Poor
0.00–0.20	Slight
0.20–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Table 3 The Interpretation of Effect Size (Plonsky & Oswald, 2014)

Effect Size	Within-Subject Studies	Between-Subject Studies
Small effect	Less than 0.60	Less than 0.40
Medium effect	Between 0.60 and 1.00	Between 0.40 and 0.70
Upper-medium effect	Between 1.00 and 1.40	Between 0.70 and 1.00
Large effect	Larger than 1.40	Larger than 1.00

is the number of agreed-on codes divided by the total number of coding opportunities' (Cooper, 2017, p. 136). The agreement rate and the Kappa coding statistic were 99.82% and 0.99, respectively, indicating an almost perfect agreement between the two coders. After reaching the agreement rate and Kappa statistics, the two coders discussed and resolved the one discrepancy.

3.4 Meta-Analysis Procedures and Statistics

The software used to conduct the meta-analytic procedures was StataSE Version 17.0. As the effect sizes of within-subject studies tend to be larger than between-subject studies (Plonsky & Oswald, 2014), unweighted and weighted effect sizes were calculated respectively for within-subject and between-subject studies. Table 3 presents Plonsky and Oswald's (2014) criteria for interpreting the magnitudes of *d*-type effect size. To estimate the range of effects across the participant population, a 95% confidence interval was constructed. In addition, weighted effect sizes were respectively calculated by sample size for within-subject and between-subject studies.

The heterogeneity of the effect sizes was examined by employing Q statistics and I-squared (I^2) statistics. A statistically significant Q-test rejects the null hypothesis and indicates that the variance of effect sizes is more than the sampling error. In addition, I^2 statistics indicate the total variance rate. Specifically, $I^2 > 25\%$ indicates heterogeneity among the eligible studies included in the meta-analysis (Talan et al., 2020; Upadhyay et al., 2022). $I^2 > 75\%$ indicates considerable heterogeneity (Cooper, 2017). Since the Q statistics and I^2 statistics for within-subject ($I^2 = 91.55\% > 25\%$; $Q = 71.71$, $p < 0.01$) and between-subject ($I^2 = 98.44\% > 75\%$; $Q = 1229.89$, $p < 0.01$) studies indicated considerable heterogeneity, this meta-analysis employed the random-effects model, which calculates the variability in effect sizes due to study-level variance (Cooper, 2017). Further, both sampling error and moderator variables may contribute to heterogeneity. In particular, the heterogeneity due

to moderator variables can be further identified by subgroup analysis when more than two eligible studies are included in a subgroup (Upadhyay et al., 2022).

Publication bias was assessed for the reliability of the study. Academic publications have drawn attention to studies reporting statistically significant and positive results with larger effect sizes, which could result in publication bias (Lin & Chu, 2018; Merino-Armero et al., 2021; Talan et al., 2020). Publication bias may also be caused by the researcher's expectation of a good performance, the name of some authors, and some degree of subjectivity in assessing outcomes (Merino-Armero et al., 2021). As this meta-analysis adopted the random-effects model, greater precision power was given to studies with larger sample sizes. Therefore, instead of using the funnel plot, a nonparametric trim-and-fill analysis of publication bias was conducted to 'estimate the number of missing studies that might exist in a meta-analysis' (Duval & Tweedie, 2000, p. 456). The results showed that no new studies would be necessary for immediate effect sizes of both within-subject (observed=7, imputed=0) and between-subject studies (observed=46, imputed=0). This implied that the meta-analysis for immediate effect sizes was not influenced by publication bias. However, the results showed that one imputed study was respectively necessary for delayed effect sizes of within-subject (observed=4, imputed=1) and between-subject studies (observed=20, imputed=1). As a result, delayed effect sizes were not weighted but reported as aggregated mean effect sizes for both observed and observed-plus-imputed studies.

4 Results

In terms of publication characteristics, the eligible studies included in this meta-analysis were published from 2006 to 2021 in 15 journals and two conference proceedings. All included studies were published in different journals or conference proceedings except for eight studies, two of which were each published in *Computer-Assisted Language Learning*, *Computers & Education*, *Language Learning and Technology*, or *Language Teaching Research*. As for learner variables, most participants reached an intermediate level of L2 proficiency. In terms of research designs, five studies had within-subject designs, while the other 16 had between-subject designs. Lexical glosses were the most frequently used digital resources outside the reading texts. Among all eligible studies, fifteen used lexical glosses while one used a personalized reading system, one used lexical priming, one read onscreen without access to any digital resources outside the reading texts, and three used multiple resources. The most frequently used test format was form recall, followed by meaning recognition. Finally, in terms of descriptive statistics, the sample sizes of the 21 included studies ranged from 16 to 282. A total of 77 effect sizes were yielded on immediate and delayed vocabulary post-tests. Five within-subject studies yielded 11 effect sizes, seven of which were immediate effect sizes ranging from 0.14 to 3.76, and the other four were delayed effect sizes ranging from 0.53 to 1.3. Sixteen between-subject studies yielded 66 effect sizes, 46 of which were immediate effect sizes ranging from 0.03 to 11.18. The other 20 delayed effect sizes ranged from -0.12 to 8.17.

Table 4 presents the immediate effect of digital reading on L2 vocabulary learning. The third column reports the aggregated effect size, and the fourth column reports the effect size weighted by sample size. As can be seen in Table 4, digital

Table 4 Immediate Effect Sizes

Study	*k*	Cohen's *d*		*p*	95% CI	
		Aggregated Mean	Weighted Mean		Lower	Upper
Within-Subject Studies	7	1.27	**1.39**	<0.01	0.39	2.01
Between-Subject Studies	46	2.12	**1.45**	<0.01	1.53	2.71

K is the number of studies, and CI is the confidence interval around the aggregated effect sizes.

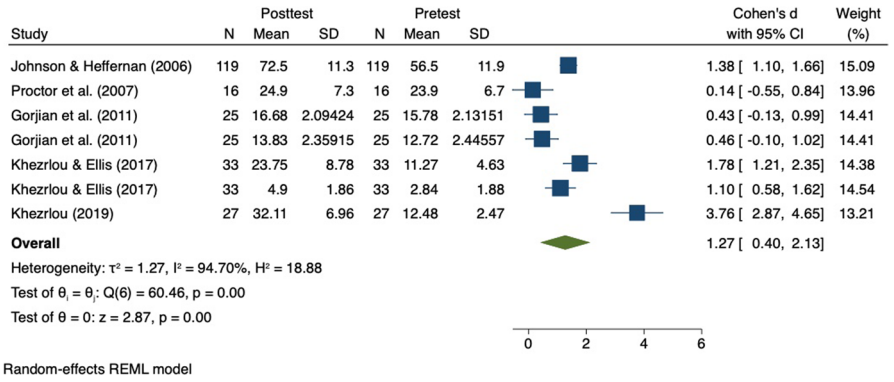

Fig. 2 Forest Plot of the Immediate Effects of Within-Subject Studies

reading had a statistically significant effect on L2 vocabulary learning. To be specific, digital reading respectively had an upper-medium effect and a large immediate effect on L2 vocabulary for within-subject studies ($d_{weighted}=1.39$, $p<0.01$) and between-subject studies ($d_{weighted}=1.45$, $p<0.01$). The immediate effect sizes and confidence intervals for each within-subject and between-subject studies are shown, respectively, by Fig. 2 and Fig. 3.

Table 5 reports the delayed effect of digital reading on L2 vocabulary learning. As previously analyzed, there was one imputed study for within-subject delayed effect sizes and one imputed study for between-subject delayed effect sizes. The third column of Table 5 presents the aggregated effect size for observed studies, and the fourth column presents the aggregated effect size for observed and imputed studies. As shown in Table 5, digital reading had a medium effect on delayed vocabulary post-tests of within-subject studies ($d_{observed+imputed}=0.86$, $p<0.01$), indicating that the overall effect decreased over time. In terms of between-subject studies, digital reading was found to have an accumulated large effect on L2 vocabulary learning as the delayed effect size ($d_{observed+imputed}=2.98$, $p<0.01$) was larger than the immediate effect size ($d_{weighted}=1.45$, $p<0.01$). The delayed effect sizes and confidence intervals for each within-subject and between-subject study can be respectively found in Fig. 4 and Fig. 5.

Subgroup analysis was conducted to investigate the sources of heterogeneity and their moderating effects. This meta-analysis selected three potential moderator variables, namely L2 proficiency, test formats, and digital resources, to explore

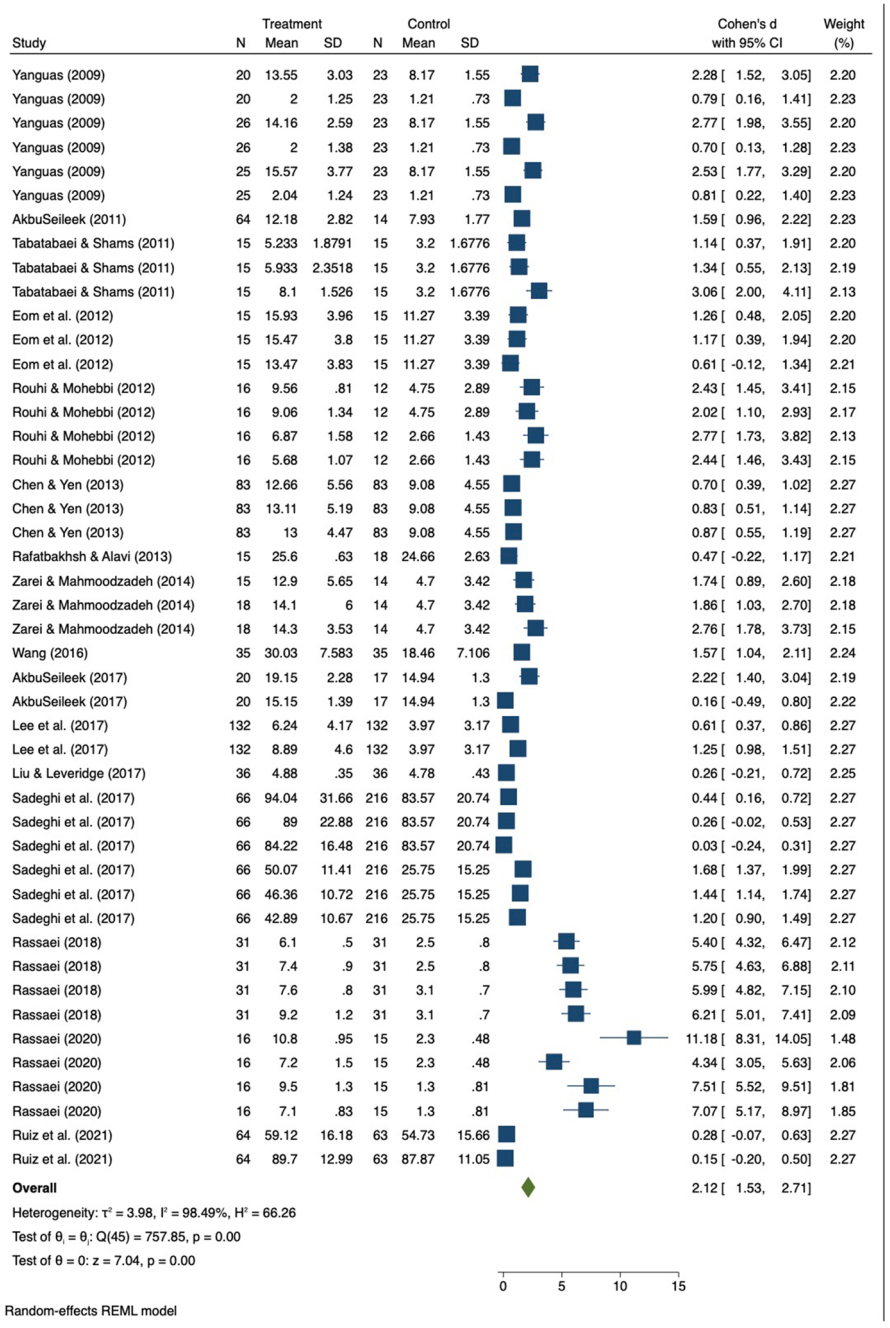

Fig. 3 Forest Plot of the Immediate Effects of Between-Subject Studies

their moderating effects on L2 vocabulary learning through reading. Table 6 and Table 7, respectively, show the moderating effects on effect sizes of within-subject and between-subject studies. To visualize the results and add transparency, forest plots are additionally illustrated in Fig. 6 and Fig. 7.

L2 proficiency was found to have a statistically significant moderating effect on effect sizes of both within-subject ($Q=19.35$, $p<0.01$) and between-subject ($Q=78.63$, $p<0.01$) studies. In within-subject comparisons, digital reading had an upper-medium effect on intermediate L2 learners' vocabulary learning ($d=1.35$, $p<0.01$), and the effect was statistically significant. For L2 beginners, digital reading only had a statistically significant and small effect ($d=0.40$, $p=0.02$) on their vocabulary learning. As for between-subject comparisons, digital reading had a statistically significant and large effect on learners at both the beginning ($d=1.42$, $p<0.01$) and intermediate ($d=3.73$, $p<0.01$) levels of L2 proficiency.

Further, the effect sizes of within-subject studies ($Q=6.77$, $p=0.03$) and between-subject studies ($Q=21.21$, $p<0.01$) were both significantly varied by test formats. For within-subject studies, meaning recognition had a statistically significant and medium effect ($d=0.64$, $p=0.02$) on L2 vocabulary learning through digital reading. Other test formats, including vocabulary knowledge scale ($d=2.24$, $p<0.01$) and mixed formats ($d=1.38$, $p<0.01$), had statistically significant and large effects on L2 vocabulary learning through digital reading. For between-subject studies, all test formats had a large moderating effect, and their effects were statistically significant.

Finally, the inspection of the effect sizes indicated that digital resources only had a statistically significant moderating effect on between-subject studies ($Q=55.80$, $p<0.01$). Among all accessible digital resources, personalized reading systems ($d=1.58$, $p<0.01$) and lexical glosses ($d=2.59$, $p<0.01$) showed statistically significant and large moderating effects on L2 vocabulary learning through digital reading.

5 Discussion

The main objectives of this meta-analysis were to synthesize research results and examine moderator variables to understand the effect of digital reading on L2 vocabulary learning. Mean effect sizes were calculated, and potential moderator variables were examined to capture the complex relationship between digital reading and L2 vocabulary learning. Regarding the mean effect sizes, a significant effect of digital reading was found for immediate L2 vocabulary tests, with an upper-medium effect of within-subject studies ($d_{weighted}=1.39$, $p<0.01$) and a large effect of between-subject studies ($d_{weighted}=1.45$, $p<0.01$). These positive results advocate for learning L2 vocabulary through digital reading, although the positive effect decreased over time for within-subject studies ($d_{observed+imputed}=0.86$, $p<0.01$). Conversely, the effect for between-subject studies accumulated over time ($d_{observed+imputed}=2.98$, $p<0.01$). Compared with the effect sizes recently calculated by Ramezanali et al. (2021), both the immediate ($g=0.46$) and delayed ($g=0.28$) effect sizes were smaller than this meta-analysis found. This may be explained by the fact that some studies included in Ramezanali et al. (2021) were not conducted in

Table 5 Delayed Effect Sizes

Study	k	Cohen's d		*p*
		Aggregated Mean		
		Observed (CI)	Observed + Imputed (CI)	
Within-Subject Studies	4	0.78 (0.43—1.14)	**0.86** (0.55—1.18)	< 0.01
Between-Subject Studies	20	3.22 (2.13 – 4.32)	**2.98** (1.83 – 4.13)	< 0.01

K is the number of studies, and CI is the confidence interval around the aggregated effect sizes.

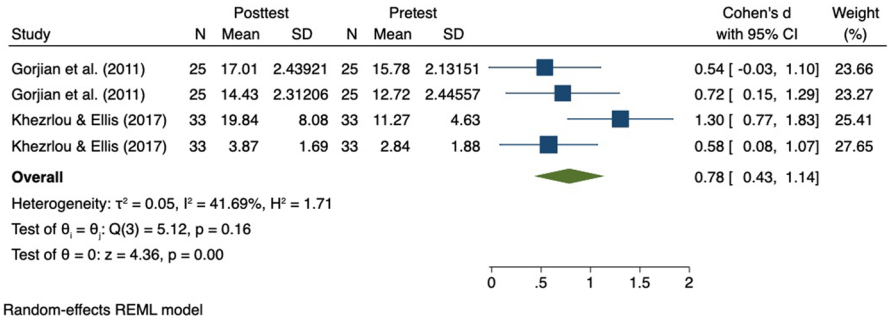

Fig. 4 Forest Plot of the Delayed Effects of Within-Subject Studies

digital contexts but focused on paper reading with lexical glosses, while this meta-analysis only included onscreen reading with lexical glosses. In digital contexts, lexical glosses can be provided through hyperlinks, allowing learners to access or not access glosses, avoiding their attention being split between the text and glosses, and mitigating their cognitive load (AkbuSeileek, 2017; Chen & Yen, 2013). In line with this discussion, Abraham's (2008) meta-analysis on L2 vocabulary learning with computerized lexical glosses reported larger effect sizes ($d_{immediate}=1.4$, $d_{delayed}=1.25$) than Ramezanali et al. (2021). Nevertheless, the weighted effect sizes reported by this meta-analysis were still larger. This may be explained by the fact that Abraham (2008) and Ramezanali et al. (2021) only investigated the facilitative potential of lexical glosses while this meta-analysis included other resources such as personalized reading systems. Different digital resources accessible to onscreen reading have moderating effects on L2 vocabulary learning through digital reading, which is further discussed in the following analysis of moderator variables.

Subgroup analysis suggested that the digital resource was a statistically significant moderator variable for between-subject studies. Onscreen reading without access to any digital resources outside the reading texts ($d=0.17$, $p=0.57$) or onscreen reading with lexical priming ($d=0.26$, $p=0.28$) only had small effects on L2 vocabulary learning. Onscreen reading with access to multiple digital resources outside the reading texts ($d=1.17$, $p=0.26$), on the other hand, had a large effect on L2 vocabulary learning. However, their effects were not statistically significant, which may be due to the small number of eligible studies. Therefore, these results must be interpreted with caution. Personalized reading systems ($d=1.58$, $p<0.01$) and lexical glosses ($d=2.59$, $p<0.01$) had statistically

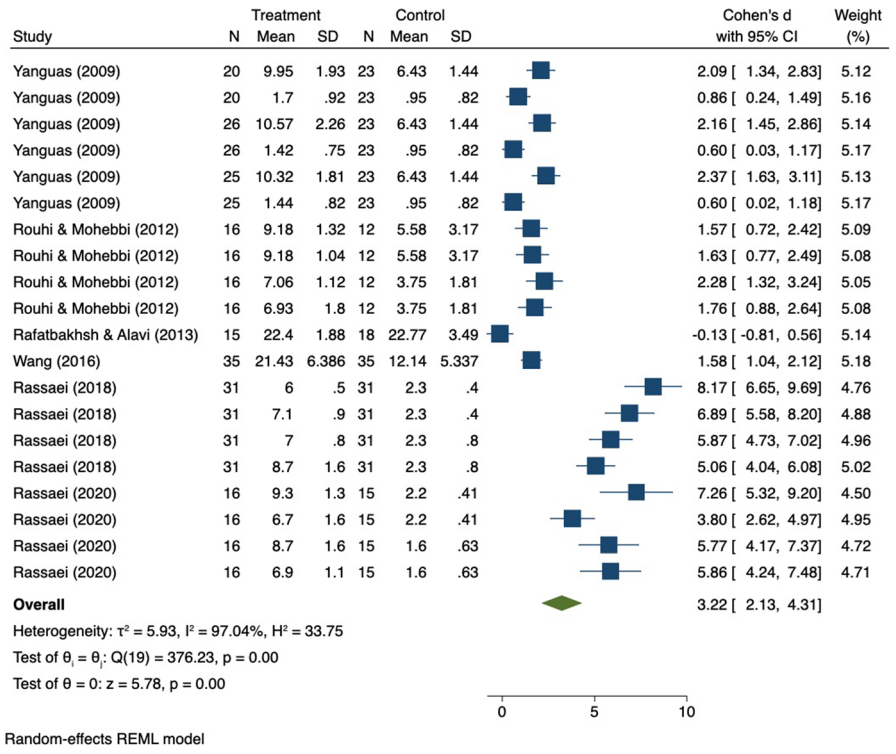

Fig. 5 Forest Plot of the Delayed Effects of Between-Subject Studies

Table 6 Moderating Effects on Effect Sizes of Within-Subject Studies

Moderator Variables	*K*	Cohen's *d*	Confidence Interval		*p*
			Lower	Upper	
L2 Proficiency (Q = 19.35, $p < 0.01$)					
Beginning	3	**0.40**	0.05	0.74	0.02
Intermediate	7	**1.35**	0.55	2.14	< 0.01
Various	1	**1.38**	1.10	1.66	< 0.01
Test Format(s) (Q = 6.77, $p = 0.03$)					
Not reported	4	**0.53**	0.25	0.82	< 0.01
Meaning recognition	3	**0.64**	0.12	1.16	0.02
VKS	3	**2.24**	0.80	3.68	< 0.01
Mixed	1	**1.38**	1.10	1.66	< 0.01
Digital Resource(s) (Q = 3.18, p = 0.07)					
Lexical glosses	5	1.66	0.64	2.69	< 0.01
Multiple resources	6	0.66	0.28	1.05	< 0.01

K is the number of studies

Table 7 Moderating Effects on Effect Sizes of Between-Subject Studies

Moderator Variables	K	Cohen's d	Confidence Interval		*p*
			Lower	Upper	
L2 Proficiency (Q = 78.63, $p < 0.01$)					
Not reported	25	**1.60**	1.26	1.95	< 0.01
Beginning	13	**1.42**	0.98	1.85	< 0.01
Intermediate	30	**3.73**	2.67	4.78	< 0.01
Various	2	0.21	-0.03	0.46	0.09
Test Format(s) (Q = 21.21, $p < 0.01$)					
Not reported	4	0.66	-0.35	1.67	0.20
Meaning recognition	17	**1.44**	1.00	1.89	< 0.01
Meaning recall	12	**3.07**	1.54	4.61	< 0.01
Form recognition	5	**4.61**	2.32	6.91	< 0.01
Form recall	29	**2.72**	1.84	3.60	< 0.01
VKS	3	**1.44**	1.16	1.71	< 0.01
Digital Resource(s) (Q = 55.80, p < 0.01)					
Onscreen reading without access to digital resources outside the actual texts	2	0.17	-0.42	0.76	0.57
Lexical priming	1	0.26	-0.21	0.72	0.28
Personalized reading system	2	**1.58**	1.20	1.96	< 0.01
Lexical glosses	63	**2.59**	2.04	3.14	< 0.01
Multiple resources	2	1.17	-0.85	3.20	0.26

K is the number of studies.

significant and large moderating effects while lexical glosses appeared to be the most effective resource. Since only one eligible study used personalized reading systems, comparing its effect to the effect of 15 eligible studies of lexical glosses might be risky. To interpret the result, one advantage of personalized reading systems is that they adapt reading texts to learners' language proficiency. Although personalized reading systems afford reading comprehensible texts, which is one of the most efficient ways for L2 learning (Boers, 2022; Krashen, 1989), all eligible studies on lexical glosses used comprehensible texts and counterbalanced the adaptive advantage of personalized reading systems. Another advantage of personalized reading systems is that they recommend texts containing unfamiliar words that learners have previously encountered so as to increase the number of word recurrences and enhance vocabulary learning. Nevertheless, learners may not correctly infer word meanings, as no definitions were provided for unfamiliar words in the recommended texts. Further, personalized reading systems do not highlight these words, so learners may skip novel words during reading instead of noticing them and inferring their meanings. According to the noticing hypothesis (Schmidt, 1990), a lexical item must be noticed before being processed and learned. In other words, any lexical item that is not noticed is unlikely to be learned.

In addition to digital resources, L2 proficiency was found to have a statistically significant moderating effect on both within-subject and between-subject studies. In both cases, although intermediate L2 learners produced wider confidence intervals,

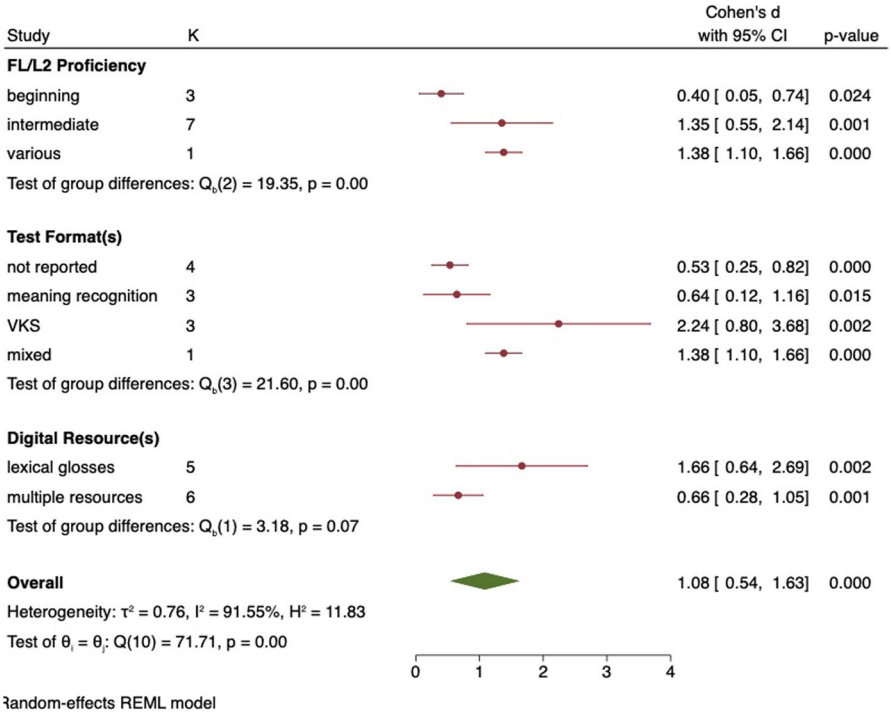

Fig. 6 The Forest Plot for Moderating Effects on Effect Sizes of Within-Subject Studies

they benefited more from digital reading than L2 beginners. The larger effect for intermediate L2 learners may be explained by the concerns raised about L2 beginners lacking enough vocabulary base to infer word meanings and retain words in the reading context (AkbuSeileek, 2011; Bengeleil & Paribakht, 2004; Laufer, 1997). Intermediate L2 learners, on the other hand, can make inferences about novel words based on contextual cues. Therefore, intermediate L2 learners can better learn L2 vocabulary through onscreen reading without access to any digital resources outside the reading texts, digital reading with lexical priming, or digital reading via personalized reading systems. As for digital reading with lexical glosses or multiple resources, although learners do not need to infer novel words as word meanings are provided, less proficient learners tend to be less efficient in allocating attentional resources than higher proficient learners (AkbuSeileek, 2008; Payne & Ross, 2005; Liu & Leveridge, 2017; Ruiz et al., 2021). To be specific, L2 beginners split more attention between the reading text and lexical gloss than L2 intermediate learners.

This finding is consistent with Abraham's (2008), which reported that more proficient learners ($d_{\text{beginning}} = 0.57$, $d_{\text{intermediate}} = 1.34$, $d_{\text{advanced}} = 2.06$) could better connect vocabulary in the glosses to their pre-existing vocabulary network and semantic system. Similar conclusions have been drawn in recent empirical research focusing on the development of other aspects of foreign and L2 skills. For instance, Zhang and MacWhinney (2023) demonstrated that, compared to beginning learners, increasing unfamiliar training stimuli will more effectively help intermediate

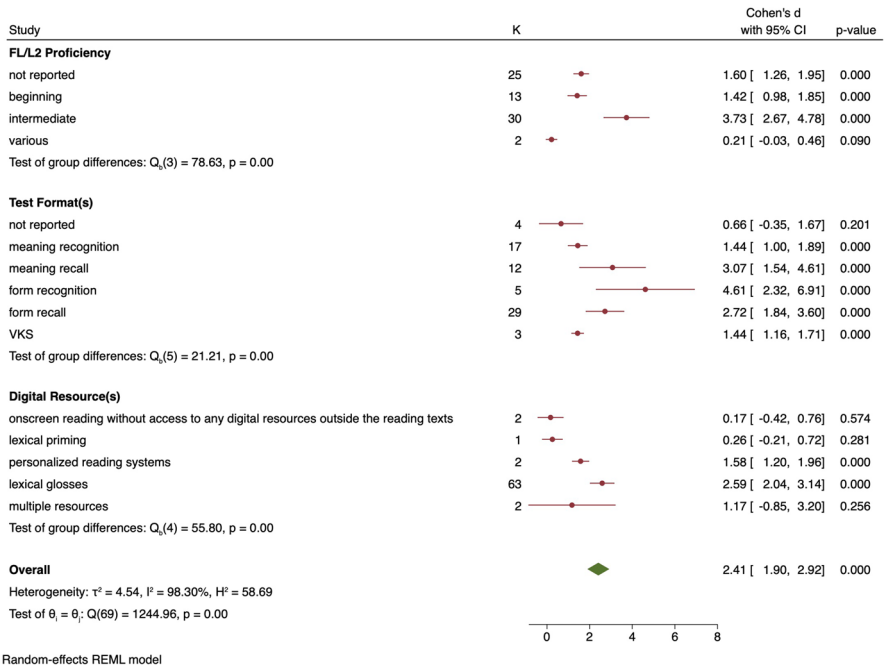

Fig. 7 The Forest Plot for Moderating Effects on Effect Sizes of Between-Subject Studies

learners acquire the phonetic knowledge of an L2. Lantz-Andersson (2018) indicated that language activities on social platforms provide diverse linguistic repertoires for L2 learners to develop their L2 socio-pragmatic competence, but advanced L2 proficiencies are needed to better exploit such skills of language-in-use on social media. Pedagogically, the findings tend to endorse a graduated increase of more novelty and diversified instructional designs and learning materials and strategies for L2 learners tailored to their proficiency levels and individual differences. As more specifically shown in the current study, this approach entails broadening exposure to various digital resources as well as providing more personalized reading systems as the learners' L2 proficiency advances.

Finally, when it comes to the test formats, we also found statistically significant moderating effects on within-subject and between-subject studies. For between-subject studies, although researchers suggested that word recall was more difficult to acquire than word recognition (González-Fernández and Schmitt, 2020; Laufer & Goldstein, 2004; Laufer & Paribakht, 1998), statistically significant and large effects were found for recognition tests, recall tests and VKS tests that involved measuring word use. It appears that digital reading can effectively enhance learning all aspects of L2 vocabulary knowledge. For within-subject studies, statistically significant and large effects were found for VKS tests and mixed tests. Meaning recognition tests appeared to have a medium moderating effect on L2 vocabulary learning through digital reading. Given the wide confidence intervals and the small number of within-subject

studies, this result must be interpreted with caution and awaits confirmation from future replication studies.

6 Conclusion

To summarize, this meta-analysis found that digital reading effectively enhanced L2 vocabulary learning. L2 proficiency, test formats, and digital resources were found to be statistically significant moderators. Subgroup analytic results suggested that intermediate learners benefited more from digital reading than L2 beginners. However, only a few eligible studies (i.e., AkbuSeileek, 2011, 2017; Eom et al., 2012; Gorjian, et al., 2011; Khezrlou, 2019; Lee, et al., 2017; Liu & Leveridge, 2017; Rassaei, 2020; Ruiz et al. 2021) clarified the tests and/or criteria for the level of proficiency. We thus suggest future research be more transparent on proficiency assessment and more rigorous about defining L2 proficiency. Further, no eligible studies investigated advanced learners' L2 vocabulary learning through digital reading, which is a gap to be addressed by future empirical studies. Results also suggested that all aspects of L2 vocabulary knowledge, including meaning recognition, meaning recall, form recognition, form recall, and vocabulary use, were facilitated by digital reading. Digital reading with access to lexical glosses appears to be the most efficient design for L2 vocabulary learning, followed by personalized reading systems. Hence, pedagogically, we suggest that teachers and learners may wish to increase the use of personalized reading systems and lexical glosses for digital reading, so as to enhance L2 intermediate and advanced learners' vocabulary learning. As L2 proficiency has been shown as a prominent moderating factor, another pedagogical insight is that increased exposure to various digital resources with comparable difficulty ladders should be offered to L2 intermediate and advanced learners to optimally enhance their vocabulary learning.

Although the statistical analysis is generally reliable, these conclusions should be interpreted as suggestive instead of definitive, due to the small number of within-subject studies included in the current meta-analysis, and more importantly, the limited number of studies on computerized lexical priming, personalized reading systems, and multiple digital resources. Future research is recommended to further explore the effectiveness of digital reading with access to digital resources outside the reading texts and lexical glosses. Along with the continuous development of natural language processing techniques and the fast update of adaptive learning algorithms, personalized reading systems have great potential in facilitating L2 vocabulary learning through digital reading and await further exploration. For example, future studies may apply personalized learning technology to lexical glosses. In addition to adapting the reading texts to learners' individual differences in L2 proficiency, personalized lexical glosses can adapt the glossed words to learners' individual differences in L2 vocabulary knowledge. Finally, as this meta-analysis seems to be the first and the only study comparing the effect of various resources on L2 vocabulary learning through digital reading, future studies may replicate systematic reviews to confirm our findings and sequence the effect of digital reading on other aspects of L2 learning.

Appendix 1

Table 8 Study Characteristics

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
Johnson and Heffernan (2006)	Computer Assisted Language Learning	The short readings project: A CALL reading activity utilizing vocabulary recycling	Within-subject design	Comprehensible	A2—B1 (Intermediate)	Other	Multiple resources
Proctor et al. (2007)	Journal of Literacy research	Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support	Within-subject design	Challenging	A1 (Beginning)	Meaning recognition	Multiple resources
Gorjian et al. (2011)	Computer Assisted Language Learning	The impact of asynchronous computer-assisted language learning approaches on English as a foreign language high and low achievers' vocabulary retention and recall	Within-subject design	Comprehensible	A2; B1 (Beginning; Intermediate)	Not reported	Multiple resources

Table 8 (continued)

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
Khezrlou and Ellis (2017)	System	Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions	Within-subject design	Comprehensible	B2 (Intermediate)	Other; meaning recognition	Lexical glosses
Khezrlou (2019)	RELC Journal	Form-focused instruction in CALL: What do learners think?	Within-subject design	Comprehensible	B2 (Intermediate)	Other	Lexical glosses
Yanguas (2009)	Language Learning and Technology	Multimedia glosses and their effect on L2 text comprehension and vocabulary learning	Between-subject design	Comprehensible	Not reported	Meaning recognition; form recall	Lexical glosses
AkbuSeileek (2011)	Computers & Education	Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary	Between-subject design	Comprehensible	A1 – A2 (Beginning)	Meaning recognition	Lexical glosses

Table 8 (continued)

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
Tabatabaei and Shams (2011)	Journal of Language Teaching and Research	The effect of multimedia glosses on online computerized L2 text comprehension and vocabulary learning of Iranian EFL learners	Between-subject design	Comprehensible	Not reported	Form recall	Lexical glosses
Eom et al. (2012)	Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications	Sense-Specific Lexical Information for Reading Assistance	Between-subject design	Comprehensible	B1 (Intermediate)	Form recall	Lexical glosses
Rouhi and Mohebbi (2012)	Journal of Asia TEFL	The effect of computer assisted L1 and L2 glosses on L2 vocabulary learning	Between-subject design	Comprehensible	Not reported	Meaning recall; form recall	Lexical glosses
Chen and Yen (2013)	Computers & Education	Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages	Between-subject design	Comprehensible	Not reported	Meaning recognition	Lexical glosses

Table 8 (continued)

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
Rafatbakhsh and Alavi (2013)	the Proceedings of the Second World Congress on Extensive Reading	The impact of computerized and paper-based 'Whodunits' as extensive reading materials on Iranian EFL learners' incidental vocabulary learning	Between-subject design	Comprehensible	Not reported	Not reported	Onscreen reading without digital resources outside the reading texts
Zarei and Mahmoodzadeh (2014)	Journal of English Language and Literature	The Effect of multimedia glosses on L2 reading comprehension and vocabulary production	Between-subject design	Comprehensible	A2 (Beginning)	Form recall	Lexical glosses
Wang (2016)	Journal of Computer Assisted Learning	Promoting contextual vocabulary learning through an adaptive computer-assisted EFL reading system	Between-subject design	Comprehensible	B1 (Intermediate)	Meaning recall	Personalized reading system

Table 8 (continued)

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
AkbuSeileek (2017)	International Journal of Learning Technology	An integrative approach based on using annotations and computer-mediated feedback to support learners' proficiency in the English language	Between-subject design	Comprehensible	A2 (Beginning)	Not reported	Multiple resources
Lee et al. (2017)	Language Learning and Technology	The effects of concordance-based electronic glosses on L2 vocabulary learning	Between-subject design	Comprehensible	B1 – B2 (Intermediate)	Meaning recall	Lexical glosses
Liu and Leveridge (2017)	British Journal of Educational Technology	Enhancing L2 vocabulary acquisition through implicit reading support cues in e-books	Between-subject design	Comprehensible	B1 – B2 (Intermediate)	Form recall	Lexical priming
Sadeghi et al. (2017)	Journal of Research in Reading	Calling Iranian learners of L2 English: Effect of gloss type on lexical retention and reading performance under different learning conditions	Between-subject design	Comprehensible	B2 (Intermediate)	Meaning recognition; other	Lexical glosses

Table 8 (continued)

Study	Journal	Title	Research Design	Text Readability	L2 Proficiency	Test Format(s)	Digital Resource(s)
Rassaei (2018)	Language Teaching Research	Computer-mediated textual and audio glosses, perceptual style and L2 vocabulary learning	Between-subject design	Comprehensible	B2 (Intermediate)	Form recall; form recognition	Lexical glosses
Rassaei (2020)	Modern Language Journal	Effects of mobile-mediated dynamic and nondynamic glosses on L2 vocabulary learning: A sociocultural perspective	Between-subject design	Comprehensible	B1 (Intermediate)	Meaning recall; form recall	Lexical glosses
Ruiz et al. (2021)	Language Teaching Research	The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL	Between-subject design	Challenging	A1 – C1 (Various)	Form recall; form recognition	Lexical glosses

Appendix 2

Table 9 Individual Effect Sizes for Immediate and Delayed Vocabulary Post-tests of Within-Subject Studies

Study	Immediate Vocabulary Post-tests							Delayed Vocabulary Post-tests						
	Npost	Mpost	SDpost	Npre	Mpre	SDpre	Cohen's *d*	Npost	Mpost	SDpost	Npre	Mpre	SDpre	Cohen's *d*
Johnson and Heffernan (2006)	119	72.5	11.3	119	56.5	11.9	1.3788493							
Proctor et al. (2007)	16	24.9	7.3	16	23.9	6.7	0.14272613							
Gorjian et al. (2011)	25	16.68	2.09424	25	15.78	2.13151	0.4259436	25	17.01	2.43921	25	15.78	2.13151	0.53699321
	25	13.83	2.35915	25	12.72	2.44557	0.4619708	25	14.43	2.31206	25	12.72	2.44557	0.71856243
Khezrlou and Ellis (2017)	33	23.75	8.78	33	11.27	4.63	1.778098	33	19.84	8.08	33	11.27	4.63	1.3014513
	33	4.9	1.86	33	2.84	1.88	1.1015886	33	3.87	1.69	33	2.84	1.88	0.5762153
Khezrlou (2019)	27	32.11	6.96	27	12.48	2.47	3.7589608							

Appendix 3

Table 10 Individual Effect Sizes for Immediate and Delayed Vocabulary Post-tests of Between-Subject Studies

Study	Immediate Vocabulary Post-tests							Delayed Vocabulary Post-tests						
	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*
Yanguas (2009)	20	13.55	3.03	23	8.17	1.55	2.2849781	20	9.95	1.93	23	6.43	1.44	2.0891637
	20	2	1.25	23	1.21	0.73	0.786067	20	1.7	0.92	23	0.95	0.82	0.864279
	26	14.16	2.59	23	8.17	1.55	2.7651254	26	10.57	2.26	23	6.43	1.44	2.1559488
	26	2	1.38	23	1.21	0.73	0.70311266	26	1.42	0.75	23	0.95	0.82	0.59983798
	25	15.57	3.77	23	8.17	1.55	2.528611	25	10.32	1.81	23	6.43	1.44	2.3669441
	25	2.04	1.24	23	1.21	0.73	0.80727555	25	1.44	0.82	23	0.95	0.82	0.59756107
AkbuSeileek (2011)	64	12.18	2.82	14	7.93	1.77	1.5918597							
Tabatabaei and Shams (2011)	15	5.233	1.8791	15	3.2	1.6776	1.1413642							
	15	5.933	2.3518	15	3.2	1.6776	1.3379304							
	15	8.1	1.526	15	3.2	1.6776	3.0556393							
Eom et al. (2012)	15	15.93	3.96	15	11.27	3.39	1.2642312							
	15	15.47	3.8	15	11.27	3.39	1.1663944							
	15	13.47	3.83	15	11.27	3.39	0.60828971							
Rouhi and Mohebbi (2012)	16	9.56	0.81	12	4.75	2.89	2.4318687	16	9.18	1.32	12	5.58	3.17	1.5701691
	16	9.06	1.34	12	4.75	2.89	2.0162425	16	9.18	1.04	12	5.58	3.17	1.6304025
	16	6.87	1.58	12	2.66	1.43	2.7727502	16	7.06	1.12	12	3.75	1.81	2.2788413
	16	5.68	1.07	12	2.66	1.43	2.4449879	16	6.93	1.8	12	3.75	1.81	1.7625173
Chen and Yen (2013)	83	12.66	5.56	83	9.08	4.55	0.70470186							
	83	13.11	5.19	83	9.08	4.55	0.82573465							
	83	13	4.47	83	9.08	4.55	0.86914544							

Table 10 (continued)

Study	Immediate Vocabulary Post-tests							Delayed Vocabulary Post-tests						
	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*
Rafatbakhsh and Alavi (2013)	15	25.6	0.63	18	24.66	2.63	0.47163115	15	22.4	1.88	18	22.77	3.49	-0.1286184
Zarei and Mahmoodzadeh (2014)	15	12.9	5.65	14	4.7	3.42	1.7409802							
	18	14.1	6	14	4.7	3.42	1.8626305							
	18	14.3	3.53	14	4.7	3.42	2.7564347							
Wang (2016)	35	30.03	7.583	35	18.46	7.106	1.5744987	35	21.43	6.386	35	12.14	5.337	1.5786111
AkbuSeileek (2017)	20	19.15	2.28	17	14.94	1.3	2.2205433							
	20	15.15	1.39	17	14.94	1.3	0.15560145							
Lee et al. (2017)	132	6.24	4.17	132	3.97	3.17	0.61286688							
	132	8.89	4.6	132	3.97	3.17	1.2454919							
Liu and Leveridge (2017)	36	4.88	0.35	36	4.78	0.43	0.25507192							
Sadeghi et al. (2017)	66	94.04	31.66	216	83.57	20.74	0.44126516							
	66	89	22.88	216	83.57	20.74	0.25545734							
	66	84.22	16.48	216	83.57	20.74	0.03277408							
	66	50.07	11.41	216	25.75	15.25	1.6830671							
	66	46.36	10.72	216	25.75	15.25	1.43858							
	66	42.89	10.67	216	25.75	15.25	1.1970978							
Rassaei (2018)	31	6.1	0.5	31	2.5	0.8	5.396628	31	6	0.5	31	2.3	0.4	8.171933
	31	7.4	0.9	31	2.5	0.8	5.7547583	31	7.1	0.9	31	2.3	0.4	6.8923984
	31	7.6	0.8	31	3.1	0.7	5.9867109	31	7	0.8	31	2.3	0.8	5.875
	31	9.2	1.2	31	3.1	0.7	6.2096362	31	8.7	1.6	31	2.3	0.8	5.0596441

Table 10 (continued)

Study	Immediate Vocabulary Post-tests							Delayed Vocabulary Post-tests						
	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*	Nexp	Mexp	SDexp	Nctrl	Mctrl	SDctrl	Cohen's *d*
Rassaei (2020)	16	10.8	0.95	15	2.3	0.48	11.179974	16	9.3	1.3	15	2.2	0.41	7.2642509
	16	7.2	1.5	15	2.3	0.48	4.3394776	16	6.7	1.6	15	2.2	0.41	3.7960293
	16	9.5	1.3	15	1.3	0.81	7.5141631	16	8.7	1.6	15	1.6	0.63	5.766938
	16	7.1	0.83	15	1.3	0.81	7.069673	16	6.9	1.1	15	1.6	0.63	5.8619247
Ruiz et al. (2021)	64	59.12	16.18	63	54.73	15.66	0.27568095							
	64	89.7	12.99	63	87.87	11.05	0.1516552							

.

Funding The study was partially supported by Li Dak Sum Innovation Fellowship at University of Nottingham Ningbo China.

Data availability Data used for the current study are available upon reasonable request for academic research..

Declarations

Competing interests The authors declare no competing interests.

Informed consent Not applicable.

References

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

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499–517. https://doi.org/10.1007/s11251-007-9016-7

AkbuSeileek, A. F. (2008). Hypermedia annotation presentation: Learners' preferences and effect on EFL reading comprehension and vocabulary acquisition. *CALICO Journal, 25*(2), 1–15.

AkbuSeileek, A. F. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education, 57*, 1281–1291.

AkbuSeileek, A. F. (2017). An integrative approach based on using annotations and computer-mediated feedback to support learners' proficiency in the English language. *International Journal of Learning Technology, 12*(2), 165–185.

Bengeleil, N., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *Canadian Modern Language Review, 61*, 225–249.

Boers, F. (2022). Glossing and vocabulary learning. *Language Teaching, 55*(1), 1–23. https://doi.org/10.1017/s0261444821000252

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

Chen, I. J., & Yen, J. C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers and Education, 63*, 416–423. https://doi.org/10.1016/j.compedu.2013.01.005

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading, 42*(2), 288–325. https://doi.org/10.1111/1467-9817.12269

Cooper, H. (2017). *Research Synthesis and Meta-Analysis: A Step–by–Step Approach (5th ed.)*. Sage.

Delgado, P., Vargas, C., Ackerman, R., & Salmeron, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Devine, J. (1988). The relationship between general language competence and second language reading proficiency: Implications for teaching. In J. D. P. L. Carrell & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 260–277). Cambridge University Press.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

*Eom, S.,Dickinson, M., & Sachs, R. (2012). *Sense-Specific Lexical Information for Reading Assistance.* Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications, BEA 2012 at the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Fisher, T., Sharples, M., Pemberton, R., Ogata, H., Uosaki, N., Edmonds, P., Hull, A., & Tschorn, P. (2012). Incidental second language vocabulary learning from reading novels: A comparison of three mobile modes. *International Journal of Mobile and Blended Learning, 4*(4), 47–61.

Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly, 51*, 844–867.

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

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics, 41*(4), 481–505.

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.

Hsieh, T. C., Wang, T. I., Su, C. Y., & Lee, M. C. (2012). A fuzzy logic-based personalized learning system for supporting adaptive English learning. *Educational Technology and Society, 15*(1), 273–288.

Hsu, C. K., Hwang, G. J., & Chang, C. K. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers and Education, 63*, 327–336. https://doi.org/10.1016/j.compedu.2012.12.004

*Huang, H. B. (2018). Computer multimedia aided word annotation for incidental vocabulary acquisition in English reading. *Educational Sciences: Theory & Practice*. https://doi.org/10.12738/estp.2018.6.248

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*, 181–193.

Johnson, A., & Heffernan, N. (2006). The short readings project: A CALL reading activity utilizing vocabulary recycling. *Computer Assisted Language Learning, 19*(1), 63–77.

Khezrlou, S. (2019). Form-focussed instruction in CALL: What do learners think? *Relc Journal, 50*(2), 235–251. https://doi.org/10.1177/0033688217738820

Khezrlou, S., & Ellis, R. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System, 65*, 104–116. https://doi.org/10.1016/j.system.2017.01.009

Krashen, S. (1989). We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *The Modern Language Journal, 73*(4), 440–464.

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lantz-Andersson, A. (2018). Language plays in a second language: Social media as contexts for emerging sociopragmatic competence. *Education and Information Technologies, 23*, 705–724.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. C. T. Huckin (Ed.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* (pp. 20–34). Cambridge University Press.

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

Laufer, B. (2020). Lexical Coverages, Inferencing Unknown Words and Reading Comprehension: How Are They Related? *TESOL Quarterly, 54*(4), 1076–1085. https://doi.org/10.1002/tesq.3004

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 53*(3), 399–436.

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365–391.

Lee, H., Lee, H., & Lee, J. H. (2016). Evaluation of Electronic and Paper Textual Glosses on Second Language Vocabulary Learning and Reading Comprehension. *Asia-Pacific Education Researcher, 25*(4), 499–507. https://doi.org/10.1007/s40299-015-0270-1

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning and Technology, 21*(2), 32–51.

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics, 74*(3), 785–794. https://doi.org/10.1111/biom.12817

Liu, Y. T., & Leveridge, A. N. (2017). Enhancing L2 vocabulary acquisition through implicit reading support cues in e-books. *British Journal of Educational Technology, 48*(1), 43–56.

Merino-Armero, J. M., González-Calero, J. A., & Cózar-Gutiérrez, R. (2021). Computational thinking in K-12 education. An insight through meta-analysis. *Journal of Research on Technology in Education*, 1–26. https://doi.org/10.1080/15391523.2020.1870250

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaf, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, 105906.

Payne, J., & Ross, B. (2005). Synchoronous CMC, working memory, andL2 oral proficiency development. *Language Learning and Technology, 9*(3), 35–54.

Pellicer-Sánchez, A. (2016). IncidentalL2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition, 38*, 97–130.

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

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878–912.

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In P. Luke (Ed.), *Advancing Quantitative Methods in Second Language Research* (pp. 106–128). Routledge.

Proctor, C. P., Dalton, B., & Grisham, D. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research, 39*(1), 71–93.

*Rafatbakhsh, E., & Alavi, S. (2013). *The impact of computerized and paper-based 'Whodunits' as extensive reading materials on Iranian EFL learners' incidental vocabulary learning*. Paper presented at the Proceedings of the Second World Congress on Extensive Reading, Seoul, Korea. Retrieved January 25, 2023, from: https://erfoundation.org/ERWC2-Proceedings.pdf#page=284

Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *Tesol Quarterly, 55*(1), 105–133.

Rassaei, E. (2018). Computer-mediated textual and audio glosses, perceptual style and L2 vocabulary learning. *Language Teaching Research, 22*(6), 1–19.

Rassaei, E. (2020). Effects of mobile-mediated dynamic and nondynamic glosses on L2 vocabulary learning: A sociocultural perspective. *Modern Language Journal, 104*(1), 284–303.

Rouhi, A., & Mohebbi, H. (2012). The effect of computer assisted L1 and L2 glosses on L2 vocabulary learning. *Journal of Asia TEFL, 9*(2), 1–19.

*Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, *25*(4), 510–539. https://doi.org/10.1177/1362168819872859

Sadeghi, K., Khezrlou, S., & Modirkhameneh, S. (2017). Calling Iranian learners of L2 English: Effect of gloss type on lexical retention and reading performance under different learning conditions. *Journal of Research in Reading, 40*, 66–86.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Macmillan.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212–226.

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal, 12*(4), 1–10.

Tabatabaei,O., & Shams, N. (2011). The effect of multimedia glosses on online computerized L2 text comprehension and vocabulary learning of Iranian EFL learners. *Journal of Language Teaching and Research*, *2*(3), 714-725

Talan, T., Dogan, Y., & Batdi, V. (2020). Efficiency of digital and non-digital educational games: A comparative meta-analysis and a meta-thematic analysis. *Journal of Research on Technology in Education, 52*(4), 474–514.

Upadhyay, K., Viramgami, A., Bagepally, B. S., & Balachandar, R. (2022). Association between blood lead levels and markers of calcium homeostasis: A systematic review and meta-analysis. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-05976-4

Vahedi, V. S., Ghonsooly, B., & Pishghadam, R. (2016). Vocabulary glossing: A meta-analysis of the relative effectiveness of different gloss types on L2 vocabulary acquisition. *Teaching English with Technology, 16*(1), 3–25.

Wang, Y. H. (2016). Promoting contextual vocabulary learning through an adaptive computer-assisted EFL reading system. *Journal of Computer Assisted Learning, 32*(4), 291–303. https://doi.org/10.1111/jcal.12132

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

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

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

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. Language Learning and Technology, 13(2), 48-67

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

Zarei, A. A., & Mahmoodzadeh, P. (2014). The effect of multimedia glosses on L2 reading comprehension and vocabulary production. *Journal of English Language and Literature*, 1(1), 1-7.

Zhang, Y., & MacWhinney, B. (2023). The role of novelty stimuli in second language acquisition: evidence from the optimized training by the Pinyin Tutor at TalkBank. *Smart Learning Environments*, 10(3). https://doi.org/10.1186/s40561-023-00223-3

Authors and Affiliations

Tong Zhu[1] · Yanhui Zhang[1] · Derek Irwin[1]

✉ Yanhui Zhang
Yanhui.Zhang@nottingham.edu.cn

Tong Zhu
hvxtz4@nottingham.edu.cn

Derek Irwin
Derek.Irwin@nottingham.edu.cn

[1] University of Nottingham Ningbo China, Ningbo, China