Effects of Mobile-Assisted Language Learning on EFL/ESL Reading Comprehension

Rui Li
School of Foreign Languages, Hunan University, Hunan, China // liruidianzi@hotmail.com

(Submitted July 2, 2021; Revised September 5, 2021; Accepted October 25, 2021)

ABSTRACT: While an increasing number of studies have cast light on the effectiveness of MALL (mobile-assisted language learning) on English as a foreign/second language (EFL/ESL) reading comprehension, there is still a lack of comprehensive meta-analysis regarding the effect sizes of these studies. To fill the gap, this study reported results based on a meta-analysis of 20 effect sizes from 17 experimental and quasi-experimental studies published during 2000–2020. The results showed that the overall effect size was significantly large, suggesting the use of MALL applications for EFL/ESL reading comprehension is more effective than traditional methods. The moderating effects of eight moderators were analyzed. The intervention settings and intervention durations were found to be significant moderators, while others did not find a significant moderating effect. Implications of the findings were discussed.

Keywords: Meta-analysis, Mobile-assisted language learning (MALL), Reading comprehension

1. Introduction

Compared with other language skills, the development of reading ability is the foundation of foreign/second language (FL/L2) learning (Li, 2021a). Motivated by the possibility of creating portable, connective, context-sensitive, location-aware, multifunctional and ubiquitous learning environments, the use of mobile-assisted language learning (MALL) applications, e.g., smartphones, tablets and e-readers, has to date been proven to be useful to develop L2 learners' language skills (Burston, 2014; Burston, 2015; Hwang & Fu, 2019; Li & Hafner, 2022; Shadiev et al., 2020), L2 reading comprehension in particular (e.g., Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020; Moon et al., 2021). Research into the use of MALL for L2 reading comprehension is crucial because a better understanding of the effects and the related moderators will inform reading pedagogy.

While researchers (e.g., Keezhatta & Omar, 2019; Mays et al., 2020; Sofiana & Mubarok, 2020) have paid much attention to empirically examine the pedagogical affordances of MALL for L2 reading comprehension, and numerous studies have obtained the facilitative effects in "increasing reading frequency, allowing for peer interaction, achieving higher sustained attention and acquiring better decoding skills" (Lin et al., 2020, p. 851), synthesized empirical evidence of its impact on reading comprehension is still lacking. Compared with the empirical studies, meta-analysis results are more reliable and generalizable, as they are based on results of multiple studies and increased sample sizes (Yanagisawa et al., 2020). In order to narrow the gap, this study synthesized various literature on MALL for L2 reading comprehension and conducted a meta-analysis to provide a more up-to-date vision on this issue. Specifically, it aims to (a) empirically generalize findings of previous MALL for L2 reading research while (b) dealing with the variability of the aggregated effects from a meta-analytic perspective.

2. Literature review

2.1. Related studies on MALL for L2 reading comprehension

Reading is an active, dynamic and complex cognitive process that involves the selection of relevant information, the mapping of information into a mental representation and the integration of information using existing knowledge (Lin et al., 2020). Previous studies (e.g., Davis & Lyman-Hager, 1997; Singhal, 1998; Whitford & Joanisse, 2018) have examined readers' decoding and metacognitive processes, L2 lexical accessibility, L2 reading strategies, cultural differences and L2 subskills of reading comprehension. In recent decades, the introduction of MALL technologies, such as cellphones (Chen et al., 2011; Sofiana & Mubarok, 2020), tablet PCs (Lin, 2017; Lan et al., 2013) and PDAs (Hsu et al., 2013; Wu et al., 2011), has reshaped the traditional L2 reading pedagogical paradigm, and researchers were having mixed and inconclusive findings towards the shift.

On the one hand, some researchers (Alemi & Lari, 2012; Lin, 2014; Mays et al., 2020) have found the facilitative effects of MALL for L2 reading comprehension. For instance, Alemi and Lari (2012) adopted a quasi-experiment to investigate the effect of vocabulary learning with SMS (short messaging service) on L2 reading comprehension. Results indicated that the experimental group outperformed the control group in L2 reading performance. In another quasi-experiment, Mays and colleagues (2020) explored the use of mobile ARS (Audience Response Systems) with student-generated questioning on EFL learners' reading comprehension. The results indicated that the quality of questions provided by the experimental group improved at a greater rate over time compared with the control group. Participants of the experimental group also have a higher level of collaboration and engagement than those of the control group.

On the other hand, other researchers (Chen et al., 2011; Lin, 2017) have obtained the limited effects of MALL for L2 reading comprehension. For instance, Chen and colleagues (2011) conducted a quasi-experiment to compare the effectiveness of direct access to digital materials with QR (quick response) codes and that of scaffolded questioning in improving EFL learners' reading comprehension. Results suggest that the MALL technology did not influence EFL learners' reading comprehension, rather the traditional approach with scaffolded questioning improved their reading performance. Similarly, Lin (2017) also conducted a quasi-experiment to examine the effectiveness of a MALL technology on EFL learners' reading performance. Results did not find any significant difference between both groups.

Taken together, although the aforementioned studies have been helpful in shedding some light on the use of MALL for L2 reading comprehension, the discrepancy among them might be explained by a number of moderators, such as proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types, according to the existing studies (Gutiérrez-Colón et al., 2020; Sung et al., 2015). The present study, therefore, was promoted by a need to revisit the moderators that may moderate the effects of MALL for L2 reading comprehension.

2.2. Related reviews of MALL for L2 reading comprehension

To date, several reviews (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020; Reiber-Kuijpers et al., 2021) on MALL for L2 reading comprehension have been conducted. For instance, Gutiérrez-Colón and colleagues (2020) provided a thorough review of MALL for L2 reading research between 2012 to 2017. Results of their study indicated that future study should focus on the use of appropriate mobile device types, the use of appropriate screen sizes for mobile reading and the application of mobile device in the informal settings. Lin and colleagues (2020) reviewed the literature on MALL for L2 reading comprehension during 2008 to 2018. Based on the review results, they provided design-related, strategy-related and learner-related guidelines. In another review, Klimova and Zamborova (2020) conducted a literature search of 21 articles on MALL for L2 reading comprehension and coded them based on research objective, participants, MALL technology, intervention durations, outcomes, main results and limitations. In a more recent review, Reiber-Kuijpers and colleagues (2021) systematically synthesized digital reading in FL/L2 in relation to digital reading environments, tasks, readers, and strategy use during 2008 to 2020. Results of their review suggest that future attempt should be extended to informal settings, and researchers should explore reading in more authentic environments and consider the important role of teachers. Although these studies have afforded insights into trends of MALL for L2 reading comprehension, studies published to date have neither directly calculated the aggregated effects of MALL for L2 reading comprehension, nor dealt with the variability of the aggregated effects with moderator analyses, which suggests an urgent need to meta-analyze the effects of MALL for L2 reading comprehension and examine whether the calculated effects were moderated by a series of moderators, including proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types.

2.3. Research purposes and questions

Two research purposes should be achieved regarding the effects of MALL for L2 reading comprehension and its related moderators. First, drawing on the data collected from the primary studies, a meta-analysis was conducted on the aggregated overall effect sizes of MALL for L2 reading comprehension. It should be pointed out here that the experimental group used MALL for L2 reading comprehension, e.g., PDAs, e-reader, tablet PCs and cellphones, while the control group used non-MALL for L2 reading comprehension, including traditional pencil and paper (Lin, 2017; Lan et al., 2013), traditional teacher-centered lectures (Priyanti et al., 2019; Wang, 2017), PowerPoint (Mays et al., 2020), paper-based materials (Wu et al., 2011), traditional classroom instructions (Wu et al., 2010) and non-personalized instructions (Hsu et al., 2013), among others. Second, informed by several

existing meta-analyses (Chen et al., 2020; Sung et al., 2015), the moderating effects were conducted in relation to proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types. Consequently, two research questions to be addressed are as follows.

- Research question 1 (RQ1): What is the overall effect size of MALL for L2 reading comprehension vs. non-MALL for L2 reading comprehension?
- Research question 2 (RQ2): How do moderators, such as proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types, affect the aggregated effect sizes?

3. Research design

3.1. Literature retrieval

We adopted a careful and exhaustive literature retrieval approach to investigate the effectiveness of MALL applications on EFL/ESL learners' reading development. Drawing on the insights of recently published reviews (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020), the potential keywords and/or keyword combinations used in those review articles were consulted to promote a comprehensive search. Related primary studies were searched from several electronic online databases (e.g., web of science, ScienceDirect, Springer, ProQuest, Scopus, Wiley, ERIC) and search engines (Google Scholar and Baidu Scholar) by using a combination of the following MALL-related and reading-related keywords integrated with Boolean operators. The following Boolean expressions of keywords, i.e., (mobile-assisted language learning OR MALL OR mobile applications OR portal devices OR handheld devices OR mobile technologies OR mobile learning OR m-learning OR ubiquitous learning OR u-learning OR mobile phones OR cellphones OR smartphones OR e-reader OR tablets OR personal digital assistants OR PDAs OR gamification) AND (reading OR reading competence OR reading skills OR reading comprehension OR reading abilities OR reading performance), were executed. Second, to further avoid the insufficient search of a significant portion of the relevant literature in the first-round, we conducted a second-round backward and forward citation search based on the review articles (Gutiérrez-Colón et al., 2020; Klimova & Zamborova, 2020; Lin et al., 2020), along with snowballing technique (Biernacki & Waldorf, 1981) by scanning references in the identified articles (e.g., Gheytasi et al., 2015; Keezhatta & Omar, 2019; Naderi & Akrami, 2018). Third, informed by Sung and colleagues (2015), we also manually searched publications in the following major CALL journals (e.g., *Computer Assisted Language Learning*, *Language Learning & Technology*, *ReCALL*, *System* and *CALICO Journal*) and educational technology journals (e.g., *Educational Technology & Society*, *Computers & Education*, *Internet and Higher Education*, *Computers in Human Behavior*, *British Journal of Educational Technology*, *Educational Technology Research and Development*, *Journal of Computing in Higher Education*, *Journal of Educational Computing Research*, *Journal of Computer Assisted Learning*, *Australian Journal of Educational Technology*, *Interactive Learning Environments*, and *The Asia-Pacific Education Researcher*, among others) to further avoid the incomplete inclusion.

3.2. Inclusion and exclusion criteria

A total of 81 studies pertinent to MALL for L2 reading comprehension were identified via the initial literature retrieval. The following inclusion and/or exclusion criteria were proposed to ensure whether the retrieved studies were eligible for the meta-analysis. In what follows, a second-round manual inclusion and/or exclusion was executed.

(1) Publications that were written in English should be confined to 2000–2020. This time range was chosen because MALL technologies remained few in number before 2000 (Duman et al., 2014). Second, to obtain a more comprehensive view, we intend to expand the time span of the recently published narrative reviews (e.g., ranging from 2008–2018 in Lin et al., 2020; from 2018–2020 in Klimova & Zamborova, 2020; and from 2012–2017 in Gutiérrez-Colón et al., 2020).
(2) The study should adopt a form of MALL technologies (e.g., mobile phones, PDAs, computer tablets or e-readers) for EFL/ESL reading comprehension. Those studies that failed to use technologies or used MALL technologies on first or other foreign language (not EFL or ESL) reading comprehension were excluded.
(3) The publications should contain sufficient statistics for data calculation or transformation of aggregated overall effect sizes. As such, only the experimental or quasi-experimental studies that examined the effectiveness of a mobile-assisted device on foreign or second language reading comprehension should be

included. More specifically, the independent variables should include different interventional modes (e.g., traditional learning method as a control group vs. mobile-assisted learning method as a treatment group), and the dependent variable should include a measure of the researcher-designed or standardized pre- and post-tests on reading comprehension between different modes. Those publications that investigated L2 learners' attitudes or perceptions, pedagogical or theoretical recommendations regarding mobile L2 reading comprehension were excluded. Furthermore, other review publications (e.g., review articles, book reviews, and editorial materials) were excluded as well. As a result, 21 eligible publications were finalized for meta-analysis.

3.3. Variables coded for each study

A well-designed coding scheme should "capture the pertinent information suitable for meta-analysis, including the identification of the publications, the characteristics of the participants, theoretically relevant features of the study and measured variables" (Wilson, 2019, p. 154). Explicit coding scheme was thus proposed in Table 1.

After the code scheme was developed, coding procedures were observed as follows: First, given the recommended practice for data dependencies (Plonsky, 2011; Plonsky & Oswald, 2014), multiple effect sizes reported in a single publication involved different participants or different types of measurement were coded separately to ensure the reliability of the analyses, resulting in a total of 21 eligible primary studies that yielded 24 effect sizes as independent studies. For instance, Hsu and colleagues (2013) conducted a quasi-experimental design that compared two experimental groups and one control group. The effect sizes were coded separately, since EFL learners of the experimental groups that involved different participants used a mobile language management system, while those of the control group adopted the traditional approach. Further, two experienced coders who are senior CALL researchers skilled at meta-analysis coding protocols negotiated with each other to ensure the consistent understanding of each variable and item. Then, they independently coded the items, and the interrater reliability (percentage of agreement) was 97%. The discrepancies were resolved by consensus through discussions.

Table 1. The descriptive information of coding scheme

Coding types	Subtypes	Operational definitions	References
Proficiency levels	1. low	Studies that reported learners as foreign language beginners.	Li (2021a)
	2. intermediate	Studies that reported learners as intermediate learners.	
Educational levels	1. primary education	Kindergarten or primary school students.	Li (2022)
	2. secondary education	Junior middle school or senior high school students.	
	3. tertiary education	College students.	
Screen sizes	1. small	Mobile phones and handheld PDAs.	Researcher-designed
	2. larger	Tablet PCs.	
Software types	1. general purposes	Applications that were NOT designed for educational purposes, e.g., WeChat, WhatsApp, telegram and QR codes.	Chen et al. (2020)
	2. educational purposes	Applications that were designed for educational purposes, e.g., language management system.	
Intervention settings	1. formal/classroom	Formal learning activities that occurred in the classroom.	Chen et al. (2020)
	2. informal/outdoor	Informal outside-of-the-classroom learning activities.	
Intervention durations	1. one session, ≤1 week	Durations fewer than one week or only one session.	Chen et al. (2020)
	2. >1 week, ≤4 weeks	Durations over one week, but fewer than four weeks.	
	3. >4 weeks, ≤one semester	Durations over four weeks, but within one semester.	
Durations	Number of weeks	Studies that reported the specific number of weeks regarding intervention durations.	Xu et al. (2019)
Instructional	1. drill and practice	Studies that used MALL to practice L2	Researcher-designed

approaches		reading ability.	
	2. personalized learning	Studies that used MALL to appropriately provide reading materials to learners based on their reading abilities.	
	3. game-based learning	Studies that adopted games as scaffolds to facilitate L2 reading.	
	4. multimedia learning	Studies that adopted multimedia resources or glosses to facilitate L2 reading.	
	5. collaborative learning	Studies that supported peer-peer and/or tutor-learner feedbacks or collaborations.	
Measured outcome types	1.standardized	Standardized TOEIC/IELTS/OPT reading comprehension test scores.	Li (2021a)
	2.researcher-designed	Reading comprehension test scores made by researchers.	

Note. L2 = second language; TOEIC = Test of English for International Communication; IELTS = International English Language Testing System; OPT = Oxford Placement Test.

3.4. Calculation and analysis of the effect sizes

As the small sample sizes might bias the aggregated effect sizes, Hedges' *g* that "provided a simple correct for the bias" was taken as the effect size index for the eligible studies (Lipsey & Wilson, 2001, p. 48). When the original data reported in the primary studies did not include means and standard deviations, we used other statistical values, e.g., *t*-value, to calculate effect sizes. For instance, we used pre- and post-test means, sample size, and *t*-value to compute the effect sizes in several eligible studies (e.g., Ataee et al., 2015; Chen & Hsu, 3008; Gheytasi et al., 2015; Lan et al., 2013). The interpretations of the magnitude of an effect size were based on Plonsky and Oswald (2014): 0.200, 0.500, and 0.800 for small, moderate, and large effects, respectively.

4. Results

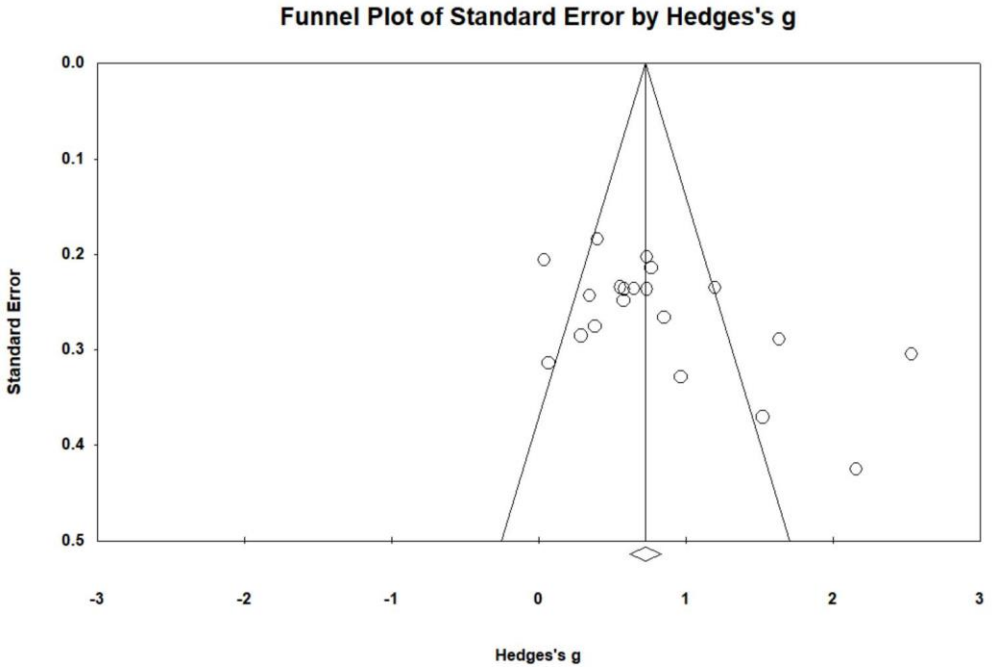
4.1. Outlier diagnosis results

According to Lipsey and Wilson (2001, p. 108), potential outliers with the extreme effect sizes that were "more than 3 standard deviations from the mean of all the effect sizes" should be excluded from the analysis. In doing so, four studies (g = 4.243, Grami & Hashemian, 2017; g = 4.632, Hazaea & Alzubi, 2016; g = 8.371, Keezhatta & Omar, 2019; g = 3.175, Motallebzadeh & Ganjali, 2011) out of 21 eligible studies that yielded extremely large effect sizes were excluded, resulting in a total of 17 remaining primary studies that yielded 20 independent studies (effect sizes) for the final analysis.

4.2. Publication bias analysis results

Since researchers normally did not publish nonsignificant results, the publication bias refers to the phenomenon that unpublished studies might differ from the published studies (Borenstein et al., 2009). The results of publication bias were often inspected via funnel plot and a fail-safe *N* method (Li, 2021a). As studies were distributed symmetrically (Figure 1), the probability of having a publication bias is rare. In addition, by evaluating how many unpublished studies with nonsignificant results would change the meta-analytic results from significant to nonsignificant, Rosenthal (1991) proposed a classical fail-safe N method to avoid the file-drawer problem. It was found that there existed no publication bias, since the result of fail-safe *N* was 950, which was significantly higher than the respective observed number 20 (z = 13.644, p < .001), that Rosenthal (1991) suggested for the file-drawer problem.

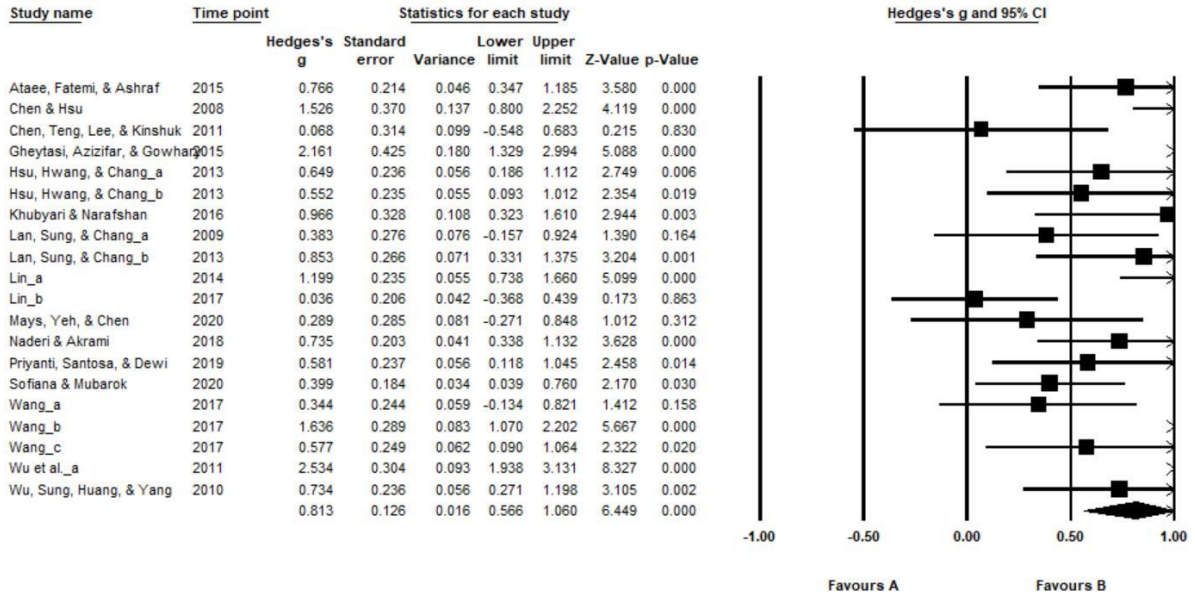
Figure 1. Funnel plot of the selected studies



4.3. Overall analysis results

The overall effect size was estimated using a random effect model, which "assumes that each observed effect size differs from the sampling error plus a value that represents other sources of variability" (Lipsey & Wilson, 2001, p. 119). As shown in Figure 2, the aggregated effect size computed from the 20 independent studies is large, g = 0.813, 95% CI = [0.566, 1.060] and significant, $z(19)$ = 6.449, p < .001, indicating a positive and large effect for the use of MALL applications for L2 reading comprehension.

Figure 2. Forest plot of the selected studies



4.4. Homogeneity analysis

As apparent in Table 2, Q-value was 92.713 with p < .001, indicating that there were between-group differences among the effect sizes resulting from factors other than subject-level sampling error. The I^2 for the overall model

showed high heterogeneity (I^2 = 79.507), indicating that one or more moderators could account for this heterogeneity (Borenstein et al., 2005; Borenstein et al., 2009).

Table 2. Heterogeneity analysis results

Q	*df*	*p*	I^2
92.713	19	.000	79.507

4.5. Moderator analysis results

The moderator analysis was carried out to examine the moderating effect of eight moderators, including proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types. The moderator analysis results have been summarized in Table 3. Two moderators were found to have a moderating effect on the overall effect sizes: intervention settings and intervention durations, whereas the other moderators did not find a significant moderating effect.

Table 3. Moderator analysis results

Moderators	*k*	*g*	*z*	95% CI	$Q_{between}$	*p*
Proficiency levels					1.586	0.208
1. low	11	0.659	4.993***	[0.400, 0.917]		
2. intermediate	9	0.998	5.003***	[0.538, 1.457]		
Educational levels					2.461	0.292
1. primary education	3	0.519	2.938**	[0.173, 0.866]		
2. secondary education	9	0.740	4.710***	[0.432, 1.048]		
3. tertiary education	8	1.003	3.848***	[0.492, 1.514]		
Screen sizes					0.048	0.826
1. small	12	0.792	5.048***	[0.485, 1.100]		
2. larger	8	0.853	3.777***	[0.410, 1.295]		
Software types					0.080	0.778
1. educational	15	0.794	5.280***	[0.499, 1.089]		
2. general	5	0.877	3.492***	[0.385, 1.369]		
Intervention settings					4.201*	0.040
1. formal/classroom	17	0.667	6.348***	[0.461, 0.873]		
2. informal/outdoor	3	1.684	3.471**	[0.733, 2.635]		
Intervention durations					6.626*	0.036
1. one session, ≤1 week	3	0.397	3.014**	[0.139, 0.656]		
2. >1 week, ≤4 weeks	4	0.948	2.342*	[0.155, 1.742]		
3. >4 weeks, ≤one semester	13	0.895	5.752***	[0.590, 1.200]		
Instructional approaches					2.145	0.709
1. drill and practice	3	1.079	2.890**	[0.347, 1.811]		
2. personalized learning	4	0.785	4.537***	[0.446, 1.124]		
3. game-based learning	2	0.786	1.966*	[0.002, 1.569]		
4. multimedia learning	7	0.869	2.607**	[0.216, 1.522]		
5. collaborative learning	4	0.601	4.671***	[0.349, 0.853]		
Measured outcome types					0.219	0.640
1. standardized	8	0.734	3.545***	[0.328, 1.139]		
2. researcher-designed	12	0.868	5.348***	[0.550, 1.186]		

Note. *k* is the number of independent studies available for the certain variable; Hedges' *g* is effect size; CI is short for confidence interval; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

4.5.1. Proficiency levels

Proficiency levels included two categories: low (*k* = 11, 55%) and intermediate (*k* = 9, 45%). According to Table 3, intermediate proficiency learners achieved large effect size (*g* = 0.998, 95% CI = [0.538, 1.457]), and low proficiency learners obtained a moderate effect size (*g* = 0.659, 95% CI = [0.400, 0.917]). However, there was no statistical difference between the two proficiency levels, $Q_{between}$ = 1.586, *p* = 0.208.

4.5.2. Educational levels

Three categories involved in educational levels: primary education (including pre-school, kindergarten and primary school education, k = 3, 15%), secondary education (junior middle school or senior high school education, k = 9, 45%) and tertiary education (college education and beyond, k = 8, 40%). As indicated in Table 3, EFL learners of tertiary education had large effect size (g = 1.003, 95% CI = [0.492, 1.514]), larger than those of the primary (g = 0.519, 95% CI = [0.173, 0.866]) and secondary education (g = 0.740, 95% CI = [0.432, 1.048]) that reported moderate effect sizes. However, no significant difference was found among the three educational levels, Q_{between} = 2.461, p = .292.

4.5.3. Screen sizes

Screen sizes contained two categories: small (normally display sized from 3.5 to 7'', e.g., handheld cellphones and PDAs, k = 12, 60%) and larger (normally sized from 7 to 10.5'', e.g., tablets, k = 8, 40%). According to Table 3, compared with the small screen size (g = 0.792, 95% CI = [0.485, 1.100]), MALL applications with larger screen size (g = 0.853, 95% CI = [0.410, 1.295]) achieved much higher moderating effect. However, the difference did not reach a significance level, Q_{between} = 0.048, p = .826.

4.5.4. Software types

Software types included educational purposes (k = 15, 75%) and general purposes (k = 5, 25%). Table 3 showed using different types of MALL software resulted in moderate-to-large effect sizes. The effect size of educational purposes is 0.794, 95% CI = [0.499, 1.089], slightly lower than that of general purposes (g = 0.877, 95% CI = [0.385, 1.369]). Between-group comparison indicated no statistically significant difference, Q_{between} = 0.080, p = .778.

4.5.5. Intervention settings

Two categories of intervention settings were involved: formal/classroom (k = 17, 85%) and informal/outdoor (k = 3, 15%). According to Table 3, significantly larger effect size was reported for using MALL applications in informal/outdoor setting (g = 1.684, 95% CI = [0.733, 2.635]) than that in formal/classroom setting (g = 0.667, 95% CI = [0.461, 0.873]), Q_{between} = 4.201, p = .040.

4.5.6. Intervention durations

Intervention durations that were divided into three categories: "one session, ≤1 week" (k = 3, 15%), ">1 week, ≤4 weeks" (k = 4, 20%) and ">4 weeks, ≤one semester" (k = 13, 65%) were found to have a significant moderating effect on the overall effect size, Q_{between} = 6.626, p = 0.036. Post-hoc comparison was computed to locate the source of the moderator effect. "One session, ≤1 week" (g = 0.397, 95% CI = [0.139, 0.656]) had the weakest moderating effect, which was statistically lower as compared ">4 weeks, ≤one semester" (g = 0.895, 95% CI = [0.590, 1.299]), Q_{between} = 5.950, p = .015. Neither significant difference existed between ">1 week, ≤4 weeks" and "one session, ≤1 week" (Q_{between} = 1.674, p = .196), nor between ">1 week, ≤4 weeks" and ">4 weeks, ≤one semester" (Q_{between} = 0.015, p = .902), indicating the longer the intervention durations, the larger the effect sizes.

4.5.7. Instructional approaches

Instructional approaches could be categorized into five approaches: drill and practice (k = 3, 15%), personalized learning (k = 4, 20%), game-based learning (k = 2, 10%), multimedia learning (k = 7, 35%) and collaborative learning (k = 4, 20%). Table 3 indicated drill and practice (g = 1.079, 95% CI = [0.347, 1.811]) and multimedia learning (g = 0.869, 95% CI = [0.216, 1.522]) had significantly high effect sizes, while game-based learning (g = 0.786, 95% CI = [0.002, 1.569]), personalized learning (g = 0.785, 95% CI = [0.446, 1.124]) and collaborative learning (g = 0.601, 95% CI = [0.349, 0.853]) had significantly moderate effect sizes. No between-group difference was observed, Q_{between} = 2.145, p = .709.

4.5.8. Measured outcome types

Two measured outcome types could be categorized: standardized test (k = 8, 40%) and researcher-designed test (k = 12, 60%). Table 3 indicated no significant difference was found, $Q_{between}$ = 0.219, p = .640, with the moderate-to-large effect size of standardized test being 0.734, 95% CI = [0.328, 1.139], and researcher-designed test being 0.868, 95% CI = [0.550, 1.186].

5. Discussion

The present study endeavored to quantitatively meta-analyze the overall effect size of MALL for L2 reading comprehension identified in the primary literature. Simultaneously, moderator analyses were also conducted to examine the moderating effects of proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types for the effect size. The meta-analytical findings regarding two RQs were discussed in the remainder of this section.

5.1. Overall effect size of MALL for L2 reading comprehension

RQ1 dealt with the overall effect size of MALL for L2 reading comprehension vs. non-MALL for L2 reading comprehension. A total of 20 effect sizes generated an overall aggregated effect size of 0.813 (95% CI = [0.566, 1.060]). The meta-analysis result indicated a positive and large effect for the use of MALL applications for L2 reading comprehension, suggesting that the use of MALL for L2 reading comprehension is more effective than traditional methods without MALL applications for L2 reading comprehension. A more informative interpretation of the results could be achieved by comparing the effect sizes with similar meta-analyses on MALL (e.g., Chen et al., 2020; Cho et al., 2018; Lin & Lin, 2019; Sung et al., 2015). The effect size (ES = 0.813) of this study is much larger than those meta-analyses on MALL for language learning (Chen et al., 2020; Cho et al., 2018; Sung et al., 2015), but comparable with the results reported in a meta-analysis conducted by Lin and Lin (2019). Among those similar studies, Sung and colleagues (2015) conducted a meta-analysis of 44 MALL studies published between 1993 and 2003 on mobile devices for language learning, and found a moderate effect size of 0.55 for the use of MALL in language learning. Likewise, Cho and colleagues (2018) meta-analyzed 20 MALL studies published between 2005 and 2017, presenting a similar overall effect size of 0.51. In a more recent study, Chen et al. (2020) performed a meta-analysis of MALL in language learning based on 84 studies published during 2008 to 2018, and obtained a moderate-to-large effect size of 0.722. A plausible explanation for the discrepancy might be due to the different domains of investigation: domain-general vs. domain-specific. In other words, different from those domain-general meta-analyses (Chen et al., 2020; Cho et al., 2018; Sung et al., 2015) that dealt with MALL for language learning in general, the current study meta-analyzed the use of MALL for L2 reading comprehension in a domain-specific way. This explanation also lends support in another domain-specific meta-analysis (Lin & Lin, 2019), which systematically synthesized findings from 33 eligible studies published during 2005 to 2018, and also obtained a large effect size of 1.005 regarding MALL applications for L2 vocabulary learning.

5.2. Findings from testing for moderators

The moderating effects of proficiency levels, educational levels, screen sizes, software types, intervention settings, intervention durations, instructional approaches and measured outcome types were considered in RQ2.

Intervention settings. There were two settings involved: informal/outdoor setting vs. formal/classroom setting. In this study, significantly larger effect size was reported for using MALL applications in informal/outdoor setting (g = 1.684) than that in formal/classroom setting (g = 0.667), which suggests that learners who used MALL applications for informal L2 reading would outperform those did formally. This result is partly consistent with the findings of existing MALL studies (Chen et al., 2020; Sung et al., 2015), which claimed the stronger effect of learning with MALL applications in informal/outdoor setting than in formal/classroom setting. Intriguingly, while significant moderating effect of MALL applications for L2 reading comprehension was found in the current study, significant between-group differences were not reported in both studies that meta-analyzed MALL applications for language learning. Reasons might be that, unlike other domain-general language learning activities, the domain-specific MALL for L2 reading comprehension emphasizes more on the "increased self-practice outside class, independent learning and self-paced learning" (Lin et al., 2020, p. 853). Chen and Lin (2016) also asserted that EFL learners prefer reading in a self-paced informal way. In other words, "mobile

readers infrequently spend long time in formal reading, but they frequently spend short time in reading utilizing their spare time" (Chen & Lin, 2016, p. 568). Meanwhile, it should also be cautioned here that, the result of higher effect size for the informal/outdoor settings over formal/classroom settings did not mean formal classroom learning was not important at all. Rather, the optimal learning outcome could be achieved only when the formal classroom learning could be complemented by the informal outside-of-the-classroom learning (Sung et al., 2015).

Intervention durations. The moderator analysis results tended to support longer durations, especially the intermediate-term durations (">1 week, ≤4 weeks"). More specifically, the lowest effect size was found for short-term durations (*ES* = 0.397, "one session, ≤1 week"), followed by long-term durations (*ES* = 0.895, ">4 weeks, ≤one semester") and intermediate-term (*ES* = 0.948, ">1 week, ≤4 weeks"). Although there was no significant difference between the short-term and intermediate-term, effect size of the long-term (*ES* = 0.895) was found to be significantly higher than that of short-term (*ES* = 0.397), indicating that long-term durations were favored, which is in alignment with the existing studies (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016). The lowest effect for short-term suggests that the effectiveness of MALL for L2 reading comprehension might be limited, as learners need more time to be acquainted with using MALL applications and with the learning scenarios (Sung et al., 2015). Furthermore, the moderating effect of intermediate-term durations was slightly larger, though nonsignificant, than that of long-term durations suggests learners normally experienced novelty effect at the earlier stage of study (within the first 4 weeks) due to the curiosity and freshness of the MALL applications, and their sustained attention and the novelty effect would wear off for long-term of investigation (Chen et al., 2020). It should be cautioned when interpreting the statistical results and comparing the difference between the short-term (*k* = 3) and intermediate-term durations (*k* = 4) because of small numbers of studies involved, warranting further research in this regard.

Proficiency levels. There were two proficiency levels: low vs. intermediate. Moderator analysis of proficiency level showed intermediate proficiency learners achieved a large effect size of 0.998, and low proficiency learners obtained a moderate effect size of 0.659. This is consistent with Droop and Verhoeven (2003), which maintained that L2 reading involves a complex process of word decoding skills, morphosyntactic knowledge and intercultural awareness. It is thus understandable to speculate that intermediate proficiency learners would outperform low proficiency learners when using MALL for L2 reading comprehension, since "limited language proficiency has also been found to impede the L2 reading comprehension" (Droop & Verhoeven, 2003, p. 81).

Educational levels. Educational levels were roughly divided into primary, secondary and tertiary education. The largest effect size was found for tertiary education (*ES* = 1.003), followed by secondary education (*ES* = 0.740) and primary education (*ES* = 0.519), indicating the effect size increased from primary and secondary education to tertiary education, echoing Chen and colleagues (2020). One possible reason why learners of tertiary education received the largest effect size and learners of primary education had the smallest effect size was due to age effect. In other words, compared with young children and adolescents, learners of tertiary education are mature adults whose word decoding skills, morphosyntactic knowledge and intercultural awareness would be more mature to facilitate their L2 reading comprehension (Droop & Verhoeven, 2003).

Screen sizes. Screen sizes were roughly categorized into small screens (cellphones and PDAs) and larger screens (e.g., tablet PCs). It was found that effect size of larger screens was larger than that of small screens, which is well-attested in a number of studies (e.g., Chen & Lin, 2016; Gutiérrez-Colón et al., 2020; Wang & Higgins, 2005). For instance, Wang and Higgins (2005) found that the small screens on mobile devices would limit the amount and type of information that can be displayed. In the same vein, Gutiérrez-Colón and colleagues (2020) also noted that small screens may restrict how learners perceive the texts to be read, information transmission and attitude towards reading, so they could have limited access to reading a text in depth with high cognitive load, and, therefore, their reading performance on mobile devices may be negatively affected.

Software types. According to Chen et al. (2020), software types were classified into educational purposes and general purposes. The effect of using general-purpose applications is found to be slightly larger than that of educational-purpose applications. This result is contrary to Chen et al. (2020) findings, which indicated educational-purpose applications are better tailored to learners' needs. A plausible explanation might be that, applications for general purposes used in current study are instant messaging tools, e.g., WeChat, WhatsApp and telegram, that featured in peer interactions and collaborations (Li et al., 2019; Li et al., 2021). Reading supported by the interactive and collaborative MALL applications would sustain EFL learners' attention and motivate their reading interest and engagement (Chen & Lin, 2016).

Instructional approaches. A scrutiny of the beneficial effects indicated that all the five instructional approaches are effective, and drill and practice (*g* = 1.079) and multimedia learning (*g* = 0.869) obtained high effects, while

game-based learning (g = 0.786), personalized learning (g = 0.785) and collaborative learning (g = 0.601) had moderate effects. This might be attributed to the overwhelming influence of the traditional "behaviorist, teacher-centred, transmission model of instruction" (Burston, 2014, p. 344) that normally used MALL for multimedia glosses (Chen et al., 2011; Yanagisawa et al., 2020). In other words, MALL technologies were often used informally as a multimedia complement for formal classroom instruction, highlighting the drill and practice features to develop FL learners' reading ability (García Botero et al., 2019). Aside from drill and practice and multimedia learning, other effective approaches should receive equal attention, since game-based learning, personalized learning and collaborative learning also obtained significantly moderate beneficial effects.

Measured outcome types. Measured outcome types consist of standardized tests and researcher-designed tests. Although effect size of researcher-designed tests (*ES* = 0.868) was slightly higher than that of standardized tests (*ES* = 0.734), no significant difference existed between the two measured outcome types, corroborating previous findings that compared the moderating effect between standardized tests and researcher-designed tests in language learning anxiety (Li, 2021a) and CALL for writing quality (Xu et al., 2019). This result suggests that both standardized tests and researcher-designed tests could warrant a good reliability to measure EFL learners' reading performance.

6. Implications

Some practical implications for teachers, providers, designers and researchers are inferred from the major findings that follow.

6.1. Implications for teachers/providers

For teachers or provides, reading materials should be adjusted to EFL learners' current proficiency levels. Teachers or providers should take learners' proficiency levels into consideration, before distributing reading tasks or assignments to learners. As low proficiency learners often accompanied with small vocabulary size and limited morphosyntactic knowledge (Droop & Verhoeven, 2003), a simplified version with high frequency words and easy-to-understand grammatical structures would be preferred. Besides, reading materials should be tailored to learners' cognitive development. Our findings showed that adult learners performed better than the children and adolescent learners when using MALL for L2 reading comprehension, suggesting that teachers and/or MALL providers should also consider learners' cognitive development when preparing the reading materials. For young readers, multimedia reading texts that integrated audio, pictorial and textual materials altogether would be better than unimodal text-only materials (Li, 2021b; Mayer, 2009). Importantly, given that all five instructional approaches were found to be significantly effective but only differed in the magnitude of effect sizes, MALL applications should be integrated into curriculums with tailored approaches depending on the educational need and purpose (Li, 2022).

6.2. Implications for designers

MALL application designers should consider learners' personalized needs and develop the easy-to-use and user-friendly interface of MALL applications to "automatically adjust text display type in different reading contexts to promote reading comprehension, sustained attention, or reduce cognitive load based on reading contexts" (Chen & Lin, 2016, p. 568; Gutiérrez-Colón et al., 2020). Moreover, since the general-purpose applications featured in interactions and collaborations have a larger effect size, designers should also consider the integration of interactive and collaborative features into learning materials for the application designs (Li, 2022; Li et al., 2021). As the beneficial effects of larger screens would be higher than the smaller ones, designers should also consider using "appropriate text display type for mobile reading in different contexts, adjusting it to the reading context to improve reading comprehension, attention, or cognitive load" (Gutiérrez-Colón et al., 2020, p. 7).

6.3. Implications for researchers

Given that MALL devices are "ideal tools for creating an interactive, collaborative and ubiquitous environment for language learning" (Chen, 2013, p. 20), researchers should try to establish informal outside-of-the-classroom learning settings for EFL learners, so that they can plan, monitor, manage and autonomously self-regulate their learning process. Furthermore, researchers should also consider adopting longer intervention durations to

improve the reliability and ecological validity of research design. With short-term durations (e.g., one session or less than 1 week), it would be rather difficult to ensure whether the effects "are produced by the features of MALL devices rather than by the experience of technology novelty" (Sung et al., 2016, p. 265-266). Moreover, as far as the short-term durations are concerned, the integrative effect between MALL devices and the curriculum would also not be satisfactorily achieved.

7. Conclusion

The results showed that the overall effect size was significantly large, suggesting the use of MALL for L2 reading comprehension is more effective than traditional methods. For moderator analysis results, the intervention settings and intervention durations were found to be significant moderators, while others did not find a significant moderating effect.

There are some limitations to be addressed though. On the one hand, due to the strict inclusion/exclusion criteria, some of the highly related empirical studies that did not report sufficient statistical information for effect size calculation, unpublished studies and publications written in other languages were not included, resulting in only 20 independent studies that met the inclusion criteria, which might affect the comprehensiveness of the meta-analysis results. Future study should include more eligible empirical studies with more keyword combinations and wider time range. On the other hand, the moderators included in this study were based on several existing meta-analyses of MALL for language learning (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), other potential moderators were not considered in the literature. Thus, researchers should include more potential moderators with sufficient information for calculation in the future.

Acknowledgement

This research was supported by the Social Science Foundation of Hunan Province (grant number 20ZDB005). The author would like to thank the anonymous reviewers and Prof. Yu-Fen Yang for their constructive and insightful comments.

References

Articles marked with an asterisk (*) were analyzed in the meta-analysis.

Alemi, M., & Lari, Z. (2012). SMS vocabulary learning: A Tool to promote reading comprehension in L2. *International Journal of Linguistics*, *4*(4), 275–287. https://doi.org/10.5296/ijl.v4i4.2318

*Ataee, E., Fatemi, M. A., & Ashraf, H. (2015). The Effect of short message service on Iranian lower intermediate EFL learners' reading comprehension through skimming and guessing strategies awareness. *Theory and Practice in Language Studies*, *5*(7), 1428–1437. https://doi.org/10.17507/tpls.0507.15

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, *10*(2), 141–163. https://doi.org/10.1177/004912418101000205

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis version 2.0* (Computer software). Biostat.

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

Burston, J. (2014). MALL: The Pedagogical challenges. *Computer Assisted Language Learning*, *27*(4), 344–357. https://doi.org/10.1080/09588221.2014.914539

Burston, J. (2015). Twenty years of MALL project implementation: A Meta-analysis of learning outcomes. *ReCALL*, *27*(1), 4–20. https://doi.org/10.1017/S0958344014000159

Chen, Z., Chen, W., Jia, J., & An, H. (2020). The Effects of using mobile devices on language learning: A Meta-analysis. *Educational Technology Research and Development*, *68*, 1769–1789. https://doi.org/10.1007/s11423-020-09801-5

*Chen, C., & Hsu, S. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society*, *11*(3), 153–180.

Chen, C. M., & Lin, Y. J. (2016). Effects of different text display types on reading comprehension, sustained attention and cognitive load in mobile reading contexts. *Interactive Learning Environments*, *24*(3), 553–571. https://doi.org/10.1080/10494820.2014.891526

*Chen, N., Teng, D. C., Lee, C., & Kinshuk. (2011). Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Computers & Education*, *57*, 1705–1715. https://doi.org/10.1016/j.compedu.2011.03.013

Chen, X. B. (2013). Tablets for informal language learning: Student usage and attitudes. *Language Learning & Technology*, *17*(1), 20–36. https://scholarspace.manoa.hawaii.edu/bitstream/10125/24503/1/Chen%20X.pdf

Cho, K., Lee, S., Joo, M., & Becker, B. (2018). The Effects of using mobile devices on student achievement in language learning: A Meta-analysis. *Education Sciences*, *8*(3), 1–16. https://doi.org/10.3390/educsci8030105

Davis, J. N., & Lyman-Hager, M. A. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, *30*(1), 58–72. https://doi.org/10.1111/j.1944-9720.1997.tb01317.x

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*, 78–103. https://doi.org/10.1598/RRQ.38.1.4

Duman, G., Orhon, G., & Gedik, N. (2014). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, *27*(2), 197–216. https://doi.org/10.1017/S0958344014000287

García Botero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language Learning*, *32*(1-2), 71–97. https://doi.org/10.1080/09588221.2018.1485707

*Gheytasi, M., Azizifar, A., & Gowhary, H. (2015). The Effect of smartphone on the reading comprehension proficiency of Iranian EFL learners. *Procedia - Social and Behavioral Sciences*, *199*, 225–230. https://doi.org/10.1016/j.sbspro.2015.07.510

Grami, S., & Hashemian, M. (2017). Effect of paper and e-dictionaries on Iranian EFL learners' reading comprehension. *International Journal of Applied Linguistics & English Literature*, *6*(5), 199–205. https://doi.org/10.7575/aiac.ijalel.v.6n.5p.199

Gutiérrez-Colón, M., Frumuselu, A. D., & Curell, H. (2020). Mobile-assisted Language learning to enhance L2 reading comprehension: A Selection of implementation studies between 2012–2017. *Interactive Learning Environments*, forthcoming. https://doi.org/10.1080/10494820.2020.1813179

Hazaea, A. N., & Alzubi, A. A. (2016). The Effectiveness of using mobile on EFL learners' reading practices in Najran University. *English Language Teaching*, *9*(5), 8–21. https://doi.org/10.5539/elt.v9n5p8

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

Hwang, G., & Fu, Q. (2019). Trends in the research design and application of mobile language learning: A Review of 2007-2016 publications in selected SSCI journals. *Interactive Learning Environments*, *27*(4), 567–581. https://doi.org/10.1016/j.compedu.2018.01.004

Keezhatta, M. S., & Omar, A. (2019). Enhancing reading skills for Saudi secondary school students through mobile assisted language learning (MALL): An Experimental study. *International Journal of English Linguistics*, *9*(1), 437–447. https://doi.org/10.5539/ijel.v9n1p437

*Khubyari, L., & Narafshan, M. H. (2016). A Study on the impact of MALL (mobile assisted language learning) on EFL learners' reading comprehension. *International Journal of English Language Teaching*, *4*(2), 58–69.

Klimova, B., & Zamborova, K. (2020). Use of mobile applications in developing reading comprehension in second language acquisition: A review study. *Education Sciences*, *10*(12), 1–11. https://doi.org/10.3390/educsci10120391

*Lan, Y., Sung, Y., & Chang, K. (2013). From particular to popular: Facilitating EFL mobile-supported cooperative reading. *Language Learning & Technology*, *17*(3), 23–38.

*Lan, Y., Sung, Y., & Chang, K. (2009). Let us read together: Development and evaluation of a computer-assisted reciprocal early English reading system. *Computers & Education*, *53*(4), 1188–1198. https://doi.org/10.1016/j.compedu.2009.06.002

Li, R. (2021a). Foreign language reading anxiety and its correlates: A Meta-analysis. *Reading and Writing*, forthcoming. https://doi.org/10.1007/s11145-021-10213-x

Li, R. (2021b). Does game-based vocabulary learning APP influence Chinese EFL learners' vocabulary achievement, motivation and self-confidence? *Sage Open*, *11*(1), 1–12. https://doi.org/10.1177/21582440211003092

Li, R. (2022). Research trends of blended language learning: A Bibliometric synthesis of SSCI-indexed journal articles during 2000–2019. *ReCALL*, forthcoming. https://doi.org/10.1017/S0958344021000343

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2021). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning*, *34*(4), 483–505. https://doi.org/10.1080/09588221.2019.1619585

Li, R., Meng, Z., Tian, M., Zhang, Z., Ni, C., & Xiao, W. (2019). Examining EFL learners' individual antecedents on the adoption of automated writing evaluation in China. *Computer Assisted Language Learning*, *32*(7), 784–804. https://doi.org/10.1080/09588221.2018.1540433

Li, Y., & Hafner, C. A. (2022). Mobile-assisted vocabulary learning: Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. *ReCALL*, *34*(1), 66–80. https://doi.org/10.1017/S0958344021000161

*Lin, C. (2014). Learning English reading in a mobile-assisted extensive reading program. *Computers & Education*, *78*, 48–59. https://doi.org/10.1016/j.compedu.2014.05.004

*Lin, C. (2017). Learning English with electronic textbooks on tablet PCs. *Interactive Learning Environments*, *25*(8), 1035–1047. https://doi.org/10.1080/10494820.2016.1242505

Lin, C., Lin, V., Liu, G., Kou, X., Kulikova, A., & Lin, W. (2020). Mobile-assisted reading development: A Review from the Activity Theory perspective. *Computer Assisted Language Learning*, *33*(8), 833–864. https://doi.org/10.1080/09588221.2019.1594919

Lin, 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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

*Mays, B. R., Yeh, H., & Chen, N. (2020). The Effects of using audience response systems incorporating student-generated questions on EFL students' reading comprehension. *The Asia-Pacific Education Researcher*, *29*(6), 553–566. https://doi.org/10.1007/s40299-020-00506-0

Moon, A. L., Francom, G. M., & Wold, C. M. (2021). Learning from versus learning with technology: Supporting constructionist reading comprehension learning with iPad applications. *TechTrends*, *65*(1), 79–89. https://doi.org/10.1007/s11528-020-00532-1

Motallebzadeh, K., & Ganjali, R. (2011). SMS: Tool for L2 vocabulary retention and reading comprehension ability. *Journal of Language Teaching and Research*, *2*(5), 1111–1115. https://doi.org/10.4304/jltr.2.5.1111-1115

*Naderi, S., & Akrami, A. (2018). EFL learners' reading comprehension development through MALL: Telegram groups in focus. *International Journal of Instruction*, *11*(2), 339–350. https://doi.org/10.12973/iji.2018.11223a

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

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

*Priyanti, N. W. I., Santosa, M. H., & Dewi, K. S. (2019). Effect of quizizz towards the eleventh-grade English students' reading comprehension in mobile learning context. *Language and Education Journal Undiksha*, *2*(2), 71–80. https://doi.org/10.23887/leju.v2i2.20323

Reiber-Kuijpers, M., Kral, M., & Meijer, P. (2021). Digital reading in a second or foreign language: A Systematic literature review. *Computers & Education*, *163*, 1–26. https://doi.org/10.1016/j.compedu.2020.104115

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage.

Shadiev, R., Liu, T., & Hwang, W. Y. (2020). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology*, *51*(3), 709–720. https://doi.org/10.1111/bjet.12839

Singhal, M. (1998). A Comparison of L1 and L2 reading: Cultural differences and schema. *The Internet TESL Journal*, *4*(10). http://iteslj.org/Articles/Singhal-ReadingL1L2.html

*Sofiana, N., & Mubarok, H. (2020). The Impact of English game-based mobile application on students' reading achievement and learning motivation. *International Journal of Instruction*, *13*(3), 247–258. https://doi.org/10.29333/iji.2020.13317a

Sung, Y., Chang, K., & Yang, J. (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

Sung, Y., Chang, K., & Liu, T. (2016). The Effects of integrating mobile devices with teaching and learning on students' learning performance: A Meta-analysis and research synthesis. *Computers & Education*, *94*, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008

*Wang, Y. (2017). Integrating self-paced mobile learning into language instruction: Impact on reading comprehension and learner satisfaction. *Interactive Learning Environments*, *25*(3), 397–411. https://doi.org/10.1080/10494820.2015.1131170

Wang, S., & Higgins, M. (2005). Limitations of mobile phone learning. Paper presented at the *IEEE International Workshop on Wireless and Mobile Technologies in Education* (*WMTE'05*), Tokushima, Japan. https://doi.org/10.1109/WMTE.2005.43

Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A Focus on word frequency effects. *Journal of Experimental Child Psychology*, *173*, 318–337. https://doi.org/10.1016/j.jecp.2018.03.014

Wilson, D. B. (2019). Systematic coding for research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis (3rd edition)* (pp.153–172). Russell Sage Foundation.

*Wu, T., Sung, T., Huang, Y., & Yang, C. (2010). Location awareness mobile situated English reading learning system. *Journal of Internet Technology*, *11*(7), 923–933. https://doi.org/10.6138/JIT.2010.11.7.05

*Wu, T., Sung, T., Huang, Y., Yang, C., & Yang., J. (2011). Ubiquitous English learning system with dynamic personalized guidance of learning portfolio. *Educational Technology & Society*, *14*(4), 164–180. https://drive.google.com/file/d/1dpSCn0MNEUUSh9DhgdwhJm99JITw_DBb/view

Xu, Z., Banerjee, M., Ramirez, G., Zhu, G., & Wijekumar, K. (2019). The Effectiveness of educational technology applications on adult English language learners' writing quality: A Meta-analysis. *Computer Assisted Language Learning*, *32*(1-2), 132–162. https://doi.org/10.1080/09588221.2018.1501069

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*(2), 411–438. https://doi.org/10.1017/S0272263119000688