



The effects of using mobile devices on language learning: a meta-analysis

Zhenzhen Chen^{1,2} · Weichao Chen³ · Jiyou Jia² · Huili An¹

Published online: 27 June 2020

© Association for Educational Communications and Technology 2020

Abstract

Despite the rapid development of the field of Mobile-assisted Language Learning (MALL), research synthesis and systematic meta-analyses on MALL are still lacking. It remains unclear how effective mobile devices are for language learning under different conditions. Review studies on the overall effectiveness of the latest smart mobile devices are still scant. In order to evaluate the learning outcomes of MALL and the impact of moderator variables, we systematically searched journal articles, conference proceedings, and doctoral dissertations published during 2008–2018 and performed a meta-analysis based on a synthesis of 84 effect sizes from 80 experimental and quasi-experimental studies. A medium-to-high effect size of 0.722 was found for the overall effectiveness of using mobile devices for language learning. The findings indicate that the use of mobile devices for language learning is more effective than conventional methods. The effects of nine moderator variables were analyzed. The target language skill, target language and first/second language were found to be significant moderators. Implications for language teaching and research are discussed.

Keywords Mobile learning · MALL · Language learning · Meta-analysis · Literature review

✉ Zhenzhen Chen
czz@bupt.edu.cn

Weichao Chen
Weichao.Chen@virginia.edu

Jiyou Jia
jjy@pku.edu.cn

Huili An
Anhuili@bupt.edu.cn

¹ School of Humanities, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Beijing 100876, China

² Graduate School of Education, Peking University, No. 5 Summer Palace Road, Beijing 100871, China

³ Office of Medical Education, School of Medicine, University of Virginia, P.O. Box 800382, Charlottesville, MA 22908, USA

Introduction

With the rapid evolution of mobile technologies and related innovations, there has been a steady increase in the application of mobile devices to facilitate language learning, thus fostering the development of the field of Mobile-assisted Language Learning (MALL) (Burston 2014a; Hwang and Fu 2019). Mobile devices have enormous potential to enhance language learning, such as improving the interactivity and mobility of the learning experience and engaging learners in situated learning, augmented reality and game-based learning (Godwin-Jones 2016, 2017; Naismith et al. 2004; Reinders and Pegrum 2017). Recognizing the potential of mobile learning, an increasing number of researchers and educators have focused on the research and implementation of MALL (Burston 2014a, 2015; Duman et al. 2015; Hwang and Fu 2019; Kukulska-Hulme and Shield 2008), with the earliest MALL publications reportedly dated from 1994 (Burston 2014a). Yet, there remains a dearth of research that systematically evaluates the learning outcomes of MALL and the impact of moderator variables. This meta-analysis, which includes studies published most recently, seeks to address this discrepancy.

Use of mobile devices in language learning

Mobile-assisted Language Learning (MALL) is considered as an extension of both mobile learning and computer-assisted language learning (Stockwell and Hubbard 2013). A popular definition of mobile learning is “learning at anytime and anywhere” (Hwang and Tsai 2011, p. 66). However, a consensus on what exactly is mobile learning has not been reached. While different definitions exist, current perspectives on mobile learning generally can be categorized into the technology-centered and learner-centered points of view (Burston 2014b; Traxler 2005; Winters 2006). A technology-centered perspective emphasizes the mobility of mobile devices and the flexible access to instructional materials afforded by mobile learning. For example, Quinn (2000) described mobile learning as “elearning through mobile computational devices.” This view, however, has been challenged by a learner-centered perspective, which emphasizes the mobility of learners rather than the mobility of devices. Increasingly, researchers and educators have begun to recognize the potential of mobile learning to facilitate interaction, collaboration and conversation among learners and ultimately to foster knowledge construction (Kukulska-Hulme 2009; Sharples et al. 2005, 2009). According to our review of the most recent MALL literature, both technology- and learner-centered perspectives still dominate the field. This review treats mobile learning as learning that is mediated by mobile devices and can potentially transcend the restriction of time and space. The use of this working definition is consistent with the practice adopted in previous MALL reviews (e.g. Duman et al. 2015; Kukulska-Hulme and Shield 2008; Viberg and Grönlund 2012).

Within the specific context of MALL, various mobile devices have been adopted to facilitate language learning over the past decades (Burston 2014a). A number of early projects used mobile phones to deliver informal learning materials, such as vocabulary lessons (Levy and Kennedy 2005; Thornton and Houser 2005; Lu 2008). Other devices such as personal digital assistants (PDAs), iPods, MP3 players, tablet Personal Computers (PCs) and laptops have also been used to facilitate the learning of a wide range of language skills (Chen and Chung 2008; Ducate and Lomicka 2013; Li et al. 2013; Song and Fox 2008;

Zurita and Nussbaum 2004). With the recent development of mobile technologies, smartphones are increasingly replacing feature phones as the most commonly adopted mobile devices. Smart mobile devices, which include smartphones, tablets, and wearable devices, such as smartwatches and smart glasses, can be operated by users' touch, voice and gestures and allow learning based on the multiple functions of the devices, such as QR codes, augmented reality, and place-sensitive functionality (Reinders and Pegrum 2017). Silverio-Fernández et al. (2018) further summarized the features that are most frequently associated with smart devices based on a content analysis of relevant publications. According to them, key features of smart devices include autonomy, such as being able to autonomously perform tasks on the background without direct user commands; connectivity, which refers to connecting to a network or sharing information with devices on a network; content-awareness, referring to gathering information from the environment through sensors; and user-interaction facilitated by activities that solicit or give information to the user. How to leverage these advanced functionalities of smart devices for language learning purposes is a topic receiving continuous attention.

However, despite the fact that increasingly diverse types of mobile devices have been used in pedagogical settings, it is still not clear which type of mobile devices is more effective for language learning. Review studies on the overall effectiveness of the latest smart mobile devices are still scant.

Previous review studies on mobile-assisted language learning (MALL)

There has been a wealth of literature documenting the application of mobile devices in language learning. A number of review studies have been conducted to analyze the development status and research trends in the field of MALL over different time periods. Their findings are summarized below in chronological order.

An early review of MALL literature was conducted by Kukulska-Hulme and Shield (2008). Most of the literature they reviewed appeared to be published between 2002 and 2007, although the authors did not specify the publishing years of the articles reviewed. While not a comprehensive review, the study was able to identify the major trends of MALL research in this early period. It was reported in this review that mobile technologies were mainly used for content delivery rather than fostering communication and collaboration. Oftentimes, mobile devices were used to text message students with learning materials and present them with links to language learning websites. Besides, instead of engaging students in anytime anywhere learning, which is an important advantage of mobile learning, many mobile learning projects delivered their content at set times. Based on their findings, Kukulska-Hulme and Shield called for a shift from adopting teacher-centered to learner-centered pedagogical approaches, and they also argued for the need to conduct more studies that employ mobile devices to support collaborative learning activities.

Viberg and Grönlund (2012) reviewed the literature on mobile-assisted second language learning published during 2007–2012. Although the experimental design was found to be the most commonly used research method, most experiments were implemented with a small number of participants and over a brief duration. This called into question the reliability and scalability of the investigation findings. Furthermore, the effectiveness of MALL was often analyzed by surveying learners' attitudes rather than investigating their learning outcomes. In terms of target language skills, students' vocabulary learning and

their development of speaking and listening skills have received the greatest attention from MALL researchers while studies on grammar, pronunciation and writing were lacking.

In the extensive review conducted by Burston (2014a), 345 MALL implementation studies published between 1994 and 2012 were reviewed. Burston found the existing studies highly imbalanced, characterized by a focus on the second language English learners, a dominance of research on college students and an emphasis on vocabulary learning. Burston's review findings echoed the concerns raised by Viberg and Grönlund (2012). Both noted short intervention duration, small sample sizes, and the implementation of rote learning that followed the behaviorist paradigm as the major issues in MALL research. Based on these findings, Burston (2015) further argued for a need to shift from adopting the teacher-centered, behaviorist and content-delivering instructional approach to embracing a student-centered, constructive and collaborative learning approach in MALL.

Duman et al. (2015) analyzed the research trends in MALL by reviewing 69 studies published during 2000–2012 and noted a steady increase in the number of MALL publications between 2008 and 2012. The majority of their findings were in agreement with those of the previous reviews (Viberg and Grönlund 2012; Burston 2014a). For instance, the experimental or quantitative research design remained the dominant inquiry approach. Vocabulary learning continued to be the most studied subject area, and other skills, such as writing skills and grammar, remained neglected. Regarding the types of mobile devices employed in the MALL studies, cell phones were most commonly used, followed by PDAs.

More recently, Hwang and Fu (2019) reviewed the 2007–2016 MALL publications and identified some new research trends. For example, the topic of MALL's impact on students' integrated/whole language skills began to get more attention, with its total number of publications ranked next only to that of vocabulary studies. It was also noted that an increasing number of studies began to adopt more rigorous designs that involved longer treatment duration and mixed research methods.

Systematic reviews mentioned above provide important insights and guidance for researchers and practitioners by analyzing the status quo and overall research trends related to MALL in different time periods. However, most of the MALL review studies are narrative reviews that have suffered from two limitations. First, narrative reviews are based on the author's subjective synthesis and inference, and the results could be influenced by the author's preferences and biases. Second, it is almost impossible to accurately synthesize a large number of studies in a narrative review when the body of research has been growing dramatically in a given field (DeCoster 2004).

Systematic meta-analysis synthesizes a large number of studies using statistical methods to compute effect sizes and could provide a more objective summary of all the included studies. To the best of our knowledge, only a few systematic meta-analyses of MALL have been conducted. Sung et al. (2015) presented a meta-analysis of 44 MALL-related journal articles and doctoral dissertations published between 1993 and 2003, and a medium effect size of 0.55 was found for the use of mobile devices in language learning. Another meta-analysis conducted by Cho et al. (2018) synthesized findings from 20 studies that were published between 2005 and 2017, reporting a similar overall effect size of 0.51. However, among these very few MALL meta-analyses, conflicting outcomes have been reported. For instance, Sung et al. reached dissimilar conclusions from Cho et al. regarding the impact of target language skills and study contexts as moderator variables. It is therefore necessary to conduct further investigation on the impact of potential moderator variables to solve the discrepancy. In addition, the rapid increase in the number of MALL publications and the great progress in mobile technologies in recent years also call forth the need to analyze the overall effectiveness of recent mobile technologies on language learning.

Purpose of this study and research questions

The primary purpose of the current study is to examine the effectiveness of using mobile devices in language learning. Specifically, two research questions were addressed:

- (1) How effective are mobile devices for language learning?
- (2) What moderator variables contribute to the differences in the effect sizes of mobile devices on language learning?

Methodology

Search strategy and screening method

To ensure a comprehensive solicitation of relevant literature, we conducted both electronic and manual searches of journal articles, conference proceedings, and doctoral dissertations published during 2008–2018. We chose 2008 as the starting point because previous reviews noted that the number of MALL publications began to increase at a fast pace from 2008 (Duman et al. 2015). Since iPhone was launched around 2008, studies published after 2008 are also more likely to reflect the influence of smart mobile devices.

Following the approach of Sung et al. (2015) in their meta-analysis of MALL, we conducted electronic literature search in these two databases: Education Resources Information Center (ERIC) and the Social Sciences Citation Index database of the Institute of Science Index (SSCI). The SSCI core database was searched for peer-reviewed articles, and ERIC for peer-reviewed articles, conference proceedings and doctoral dissertations. In addition, to ensure that no articles were missed, a manual search for full-text articles was conducted in eight major journals of educational technology and technology-enhanced language learning: *Computers & Education*, *Interactive Learning Environments*, *Educational Technology Research & Development*, *Journal of Computer Assisted Learning*, *Computer Assisted Language Learning*, *Language Learning & Technology*, *ReCALL* and *System*.

Informed by the previous reviews of MALL (Duman et al. 2015; Sung et al. 2015), we applied three sets of search terms to retrieve the relevant literature:

- (1) Mobile-device-related keywords, including mobile, portable, wireless, ubiquitous, handheld, mobile/smart phone, PDA, tablet PC, pad, laptop, iPad, e-book, pocket e-dictionary and classroom response system; and
- (2) Learning-related keywords, including teaching, learning, training, instruction and lecture; and
- (3) Language-learning-related keywords, including language, ESL, listening, speaking, reading, writing, grammar, pronunciation, translation and vocabulary.

These three sets of keywords were combined when searching the aforementioned electronic databases. The search in SSCI database returned 1042 journal articles, and the ERIC search returned 716 journal articles, 56 doctoral dissertations and 14 conference proceedings.

In the subsequent screening process, the following criteria were applied to determine if a study was eligible for inclusion in this meta-analysis:

- (1) The study adopts an experimental or a quasi-experimental design that includes a control group. Qualitative studies or pre-experimental studies of single group designs were excluded.
- (2) The use of mobile devices is the examined variable in the intervention. Experiments that only compare different learning approaches or strategies (e.g. Zheng et al. 2016) were excluded.
- (3) The study reports experimental results of learning achievement measured by test scores. Studies only reporting affective variables, such as motivation, attitudes and perceptions, were excluded.
- (4) The study has sufficient information to calculate effect sizes, such as mean, SD, sample sizes, t value, or F value.

The entire screening process consisted of three phases:

In Phase 1, two experienced researchers separately read through all the retrieved abstracts to determine whether the studies met the inclusion criteria specified above. The initial screening yielded 204 articles and 19 doctoral dissertations, including undecided articles. Because abstracts were not available for the 14 conference proceedings in the database, they were reviewed in Phase 2.

In Phase 2, full-text papers of all the above-mentioned articles identified in Phase 1 and conference proceedings were downloaded for further screening.

In Phase 3, a manual search of the selected journals mentioned earlier was conducted, and eight articles were added.

Two researchers discussed and resolved any discrepancies in the screening process. When an agreement could not be reached, a third researcher was involved to address the disagreement. Eventually, 80 publications were included in the meta-analysis, including 76 journal articles, three doctoral dissertations and one conference publication.

Coding scheme

In order to explore the impact of different moderator variables on the outcomes of MALL, all the eligible studies were coded. Two researchers conducted the coding. The coding scheme developed by Sung et al. (2015) was adapted for the current study. Adjustment was made to some of the sub-categories to fit the purpose of the current study. The coding scheme was also reviewed by a third experienced researcher and was subsequently adjusted based on the third researcher's suggestions. The two coders discussed and finalized the coding scheme together. The resulting coding scheme consists of the following nine categories:

- (1) Educational level. The sub-categories are elementary school, secondary school, post-secondary, and mixed.
- (2) Device type. The sub-categories are smart handheld mobile devices, such as smartphones, smartwatches, iPads, iPod Touches, and Android tablets; non-smart handheld mobile devices, such as mp3s, traditional feature phones, and traditional classroom response systems; other mobile (but not handheld) devices, such as laptops; and mixed.
- (3) Application type, including general-purpose and educational-purpose applications. General-purpose applications refer to programs that are not specifically developed for educational uses, such as LINE (Chen Hsieh et al. 2017) and WhatsApp (Andujar 2016), whereas educational-purpose applications refer to programs that are specifically

designed for learning purposes, which include both commercial applications, such as Kahoot! (Hung 2017), and applications designed by researchers (Liu and Chu 2010; Hwang et al. 2014; Wu 2015).

- (4) Instructional approach, including self-directed learning (mobile learning at students' own pace, e.g., Hwang and Chen 2013; Wu 2015); flipped learning (students conducted mobile learning before class and subsequently used class time to apply their newly acquired knowledge, e.g., Wang 2016; Wu et al. 2017); collaborative learning (students conducted mobile learning in pairs or groups to discuss a concept or find a solution to a problem, e.g., Lai 2016; Lan and Lin 2016); situated learning (contextualized language learning in personalized real-life contexts, e.g., Sandberg et al. 2011; Wu et al. 2011); game-based learning (incorporation of games or game design elements, such as scoring, competition and rules, to enhance learning, e.g., Grimshaw and Cardoso 2018; Hung, 2017); teacher-led instruction (the teacher guided the students throughout the learning process, e.g., Lin 2017); assessment (students used mobile devices for formative assessment or quizzes, e.g., Agbatogun 2014) and mixed approach.
- (5) Learning context, including classroom, outdoor, and unrestricted learning contexts. Classroom learning contexts refer to situations where mobile learning activities took place within the confines of formal classrooms. Studies of mobile language learning in unstructured formats without the constraints of time and space were labeled as unrestricted learning contexts. Outdoor contexts refer to settings where students finished instruction or homework assignments on a mobile field trip in authentic environments, such as the zoo or the campus site.
- (6) Intervention duration, including five categories: one session, and learning spreading over up to 4 weeks, up to 10 weeks, up to 20 weeks and more than 20 weeks.
- (7) Target language skills, including speaking, listening, reading, writing, vocabulary, pronunciation, grammar, integrated/whole language skills and early literacy.
- (8) Target language, including English, Chinese, Spanish, French, Norwegian, Turkish and mixed.
- (9) L1/L2: first language (L1), second language (L2) and mixed.

During the coding process, each researcher first separately coded 10% of all the eligible studies. The two coders subsequently discussed and resolved any discrepancies in the initial coding process. After this initial training, the two researchers separately coded the rest of the included studies. All the differences in coding were resolved through discussion.

Calculation of effect size

The Comprehensive Meta Analysis 3.0 (<https://www.meta-analysis.com/>) software was used to compute the effect sizes. Effect size was calculated as the difference in means between the experimental group and the control group divided by the pooled standard deviation. Hedge's *g* was used when reporting effect sizes.

Eighty-five independent studies were extracted from the 80 articles that remained after the screening process. According to Borenstein et al. (2009), if multiple outcomes within a study are based on the same participants, they should not be extracted as separate studies, because it will lead to an inaccurate estimate of the summary effect. Following the guideline by Borenstein et al., the different outcomes from the same participants were combined and included as a single study when calculating the overall effect size.

Publication bias

Research has established that studies reporting significant results are more likely to be published (Dickersin et al. 1987; Easterbrook et al. 1991). Since published studies are more likely to be included in meta-analysis, there exists a risk of publication bias in review studies. Multiple methods are available to examine the presence of publication bias. The first approach involves using a funnel plot. The funnel plot displays the relationship between the standard errors of the included studies and their effect sizes in the shape of a funnel (Borenstein et al. 2009), which provides a subjective impression of the presence of publication bias based on the spread of studies. The second approach involves using the measure of Classic Fail-safe N , which was proposed by Rosenthal (1979) to represent how many missing studies need to be incorporated into the meta-analysis to bring the p -value to a non-significant level. The third approach uses Orwin's Fail-safe N proposed by Orwin (1983). Compared with Rosenthal's method, which focuses on statistical significance and assumes that the mean effect size in the missing studies is zero, Orwin's approach allows the researchers to compute how many missing studies are needed to bring the summary effect to a level below a specified value other than zero (Borenstein et al. 2009). Considering that the funnel plot can only give a rough estimation, this study adopts Rosenthal's Classic Fail-safe N and Orwin's Fail-safe N to assess publication bias.

Results and discussion

Overall effect size

One of the eligible studies, a study by Lin and Hwang (2018), yielded an unusually large effect size ($g = 10.364$), which was much larger than the overall effect size of the 85 studies ($g = 0.765$). Based on the suggestions from Lipsey and Wilson (2001), this study was excluded from the current analysis. A random-effect model was used to synthesize the effect sizes of the remaining 84 studies. These studies compared the effect of mobile learning with control groups adopting one of the following approaches: face-to-face learning (e.g. Liakin et al. 2017), traditional paper-based learning (e.g. Lu 2008; Saran et al. 2012), computer-based learning (e.g. de la Fuente 2014) and other unspecified traditional learning approaches.

As summarized in Table 1, the overall effect size of using mobile devices on language learning is 0.722 ($p < .001$), with a 95% confidence interval of 0.611–0.833. According to Cohen (1988), the effect sizes of ≥ 0.50 and ≥ 0.80 are considered medium and large respectively. Therefore, the results suggest using mobile devices for language learning is significantly more effective than learning with other conventional approaches, with a medium-to-high effect size. The average score of the students learning with a mobile device is 0.722

Table 1 Overall effect size of using mobile devices on language learning

	k	g	z	95% CI	Q-value	I^2
All studies	84	0.722	12.731***	[0.611–0.833]	366.396***	77.347

CI confidence interval

*** $p < .001$

standard deviation above that of learners who were not using one. An effect size of 0.7 also indicates that 76% percent of learners in the control group would underperform the average mobile learners in the experimental group (Coe 2002).

As listed in Table 1, the result of the Q statistic ($Q=366.396$, $p<.001$) provides evidence that heterogeneity in effect sizes existed among the 84 included studies and that the observed variance was attributed to sources other than sampling errors. The I^2 statistic was computed to further analyze the extent of inconsistency across these findings. The resulting I^2 index of 77.347% is regarded as high according to the benchmark of 25%, 50% and 75% suggested by Higgins et al. (2003). This further justifies a subgroup analysis to find out potentially important moderator variables contributing to the variation of effect sizes (Borenstein et al. 2009).

Effect sizes of moderator variables

We examined the impact of nine groups of potential moderators in the subgroup analysis. The result of our subgroup analysis is summarized in Table 2 and discussed below.

Educational level

We examined the effect of implementing MALL across different educational levels. According to Table 2, over half ($k=44$, 52%) of the studies were implemented in post-secondary educational settings, and these studies yielded a large effect ($g=0.843$, $p<.001$). A medium effect size was found for kindergarten/preschool children ($g=0.493$, $p<.001$), elementary school students ($g=0.603$, $p<.001$) and secondary school students ($g=0.653$, $p<.001$) engaging in MALL. The effect of using mobile devices on language learning was not statistically significant for mixed learner populations ($g=0.496$, $p>.05$). According to the statistic of Q_B ($Q_B=8.047$, $p>.05$), no statistically significant difference existed between the effect sizes of these learner populations.

Our findings indicate the effect of using mobile devices in language learning increased from kindergarten/preschool to universities. One possible reason why MALL was less effective for younger students might be related to the learning contexts. Among the 22 eligible studies that involved preschool/kindergarten children and elementary school students, the majority ($k=18$, 82%) of them were conducted in classroom settings. By contrast, among the 44 studies involving post-secondary students, only 29.5% ($k=13$) were conducted in formal settings. According to Sung et al. (2016), mobile learning tends to produce larger effect sizes in informal settings than in formal classrooms. Therefore, the differences in the learning contexts might have contributed to the difference in learning outcomes between student populations.

Intervention duration

The intervention duration moderator includes five major categories: one session ($k=5$, 6%), up to 4 weeks ($k=23$, 27%), up to 10 weeks ($k=32$, 38%), up to 20 weeks ($k=18$, 21%), and more than 20 weeks ($k=4$, 5%). According to Table 2, studies lasting for one session achieved a medium-to-large effect ($g=0.731$, $p<.05$). Among the studies that lasted for longer than one session, the effect of MALL decreased with longer duration: Interventions that were completed within 4 weeks yielded the largest effect ($g=0.796$,

Table 2 Effect sizes of moderator variables

Category	<i>k</i>	<i>g</i>	<i>z</i>	95% CI	Q_B	df
<i>Educational level</i>					8.047	4
1. Kindergarten/preschool	8	0.493	4.530***	[0.280–0.706]		
2. Elementary	14	0.603	4.292***	[0.328–0.897]		
3. Secondary	15	0.653	4.546***	[0.372–0.935]		
4. POST-secondary	44	0.843	10.744***	[0.689–0.997]		
5. Mixed	3	0.496	1.872	[– 0.023–1.016]		
<i>Intervention duration</i>					1.373	5
1. One session	5	0.731	2.461**	[0.149–1.314]		
2. Less than 4 weeks	23	0.796	7.128***	[0.577–1.015]		
3. Less than 10 weeks	32	0.746	6.862***	[0.533–0.959]		
4. Less than 20 weeks	18	0.662	6.172***	[0.452–0.872]		
5. More than 20 weeks	4	0.575	2.49**	[0.122–1.027]		
6. Not specified	2	0.558	1.356	[– 0.248–1.365]		
<i>Device type</i>					6.650	3
1. Non-smart handhelds	17	0.787	5.739***	[0.518–1.056]		
2. Smart handhelds	56	0.763	9.957***	[0.613–0.914]		
3. Other mobile devices	8	0.489	4.530***	[0.277–0.700]		
4. Mixed	3	0.467	2.922**	[0.154–0.780]		
<i>Application type</i>					0.643	1
1. Educational-purpose	51	0.759	9.934***	[0.609–0.909]		
2. General-purpose	33	0.667	7.721***	[0.498–0.836]		
<i>Instructional approach</i>					9.802	7
1. Self-directed learning	43	0.768	9.759***	[0.614–0.922]		
2. Collaborative learning	7	0.802	2.412**	[0.150–1.454]		
3. Situated learning	11	0.795	4.284***	[0.431–1.159]		
4. Flipped learning	3	0.769	1.938	[– 0.009–1.547]		
5. Game learning	9	0.570	4.199***	[0.304–0.837]		
6. Teacher-led	5	0.373	2.284**	[0.053–0.694]		
7. Assessment	3	1.168	3.51***	[0.516–1.820]		
8. Others	3	0.375	1.584	[– 0.089–0.838]		
<i>Learning context</i>					3.58	2
1. Classroom	41	0.612	7.712***	[0.457–0.768]		
2. Unrestricted	38	0.804	11.614***	[0.668–0.939]		
3. Outdoor	5	0.959	2.146**	[0.083–1.835]		
<i>Target language skill</i>					18.903**	7
1. Listening	4	1.08	4.043***	[0.556–1.603]		
2. Reading	14	0.375	3.469**	[0.163–0.587]		
3. Speaking	5	1.056	3.712***	[0.499–1.614]		
4. Vocabulary	23	0.772	7.905***	[0.581–0.964]		
5. Writing	4	1.041	3.545***	[0.466–1.617]		
6. Integrated/whole skill	23	0.824	6.312***	[0.568–1.079]		
7. Early literacy	8	0.493	4.53***	[0.280–0.706]		
8. Others	3	0.487	2.384**	[0.086–0.887]		
<i>Target language</i>					25.166***	3

Table 2 (continued)

Category	<i>k</i>	<i>g</i>	<i>z</i>	95% CI	Q_B	df
1. English	72	0.785	12.853***	[0.665–0.904]		
2. Chinese	4	0.142	1.033	[– 0.128–0.413]		
3. Spanish	3	0.725	1.685	[– 0.118–1.569]		
4. Others	5	0.300	2.259**	[– 0.040–0.559]		
<i>L1/L2</i>					49.994***	2
1. L1	16	0.442	5.954***	[0.297–0.588]		
2. L2	66	0.825	11.993***	[0.690–0.960]		
3. Mixed	2	0.024	0.258	[– 0.156–0.203]		

CI confidence interval

*** $p < .001$; ** $p < .05$

$p < .001$). This was followed by those finished within 10 weeks ($g = 0.746$, $p < .001$) and 20 weeks ($g = 0.662$, $p < .001$), respectively. The effect of long-term studies lasting for more than 20 weeks was the smallest ($g = 0.575$, $p < .05$). According to the statistic of Q_B ($Q_B = 1.373$, $p > .05$), the effect sizes did not differ significantly between different intervention durations.

A decline in effect size was observed in our analysis for those interventions that lasted for more than 4 weeks. Although short implementation duration is considered a major flaw in MALL studies (Burston 2014a; Viberg and Grönlund 2012), our findings suggest that shorter-term interventions yielded larger effect sizes than longer-term ones. There are two possible explanations for the results. First, students typically experience novelty effect at the beginning of the study due to the freshness and curiosity of the new technology, but in longer-term investigations the novelty effect tends to wear off (Liakin et al. 2017), which might bring down the learning effect. Second, researchers are more likely to invest the best possible resources and energy into studies within a shorter time frame (Sung et al. 2016). For long-term studies, it is difficult to maintain the same level of input.

Device type

Device types include four categories: smart ($k = 56$, 67%) and non-smart handheld mobile devices ($k = 17$, 20%), other mobile (but not handheld) devices ($k = 8$, 10%), and mixed ($k = 3$, 3%). Table 2 shows using different types of devices in MALL resulted in medium-to-large effects. The effect size of using non-smart handheld devices was 0.787 ($p < .001$), slightly larger than that of using smart handheld devices ($g = 0.763$, $p < .001$). Using other mobile (but not handheld) devices, such as laptops, yielded a moderate effect size of 0.489 ($p < .001$). A similar effect was reported for using mixed types of devices ($g = 0.467$, $p < .05$). According to Q_B statistic ($Q_B = 6.650$, $p > .05$), no statistically significant difference existed between the effect sizes of different types of devices.

Both smart and non-smart handheld devices outperformed other mobile devices that have larger screen sizes in facilitating language learning. This might be attributed to the smaller sizes of handheld devices, which further encourages learners to study anytime anywhere.

According to Table 2, using non-smart handheld devices led to the strongest effect on language learning. Many of these studies utilized the Short Messaging Service (SMS)/Multimedia Messaging Service (MMS) service of traditional mobile phones. For example,

both Lu (2008) and Alemi et al. (2012) examined the educational outcomes of SMS-based vocabulary lessons, while Saran et al. (2012) investigated the effectiveness of learning vocabulary via multimedia messages. Lin and Lin (2019) also concluded from their meta-analysis that implementing the SMS/MMS mode of vocabulary learning yields a very high effect, which might explain the strong positive outcomes of using non-smart handheld devices in this study.

The benefits of using smart handheld devices on language learning are also supported in our meta-review. Among the studies using smart handheld devices, 42.9% used smartphones. Smartphones have such affordances as connectivity, portability, touchscreens, an infinite number of applications available and the GPS function, and they have greatly enhanced the usability and functionality of mobile phones (Godwin-Jones 2017). The widespread ownership of smartphones has also facilitated the adoption of the Bring Your Own Device (BYOD) model in language learning (Burston 2013). For example, the effect of using students' personal smartphones as clickers in English as a Foreign Language (EFL) classrooms was investigated by both Hung (2017) and Chou et al. (2017). Both reported a strong effect on learning achievement, and students responded positively to the BYOD mode of learning. iPads were the second most frequently used (17.9%) smart handheld devices in this review. However, except for two studies, all iPad-based studies were conducted in classrooms, and the participants were all preschool/kindergarten children. More studies introducing iPads to other learner populations and different learning settings are needed. While most studies using smart devices focused on these commonly adopted ones, Shadiev et al. (2018) examined the outcomes of using smartwatches to engage learners in authentic learning environments. Additional investigation is warranted to examine language learning outcomes from using emerging smart devices.

Application type

The application type moderator includes two categories: general-purpose (39%) and educational-purpose applications (61%). Table 2 shows a medium-to-large effect size for both general-purpose ($g=0.667$, $p<.001$) and educational-purpose applications ($g=0.759$, $p<.001$). The effect of using educational applications is slightly larger, but no statistically significant difference existed between the effect sizes of these two types of applications ($Q_B=0.643$, $p>.05$).

Applications developed for educational purposes are better tailored to students' needs and pedagogical goals, which accounts for the larger effect size associated with studies using educational-purpose applications. Our findings also corroborate previous findings comparing the moderating effect between using educational and general-purpose software in general mobile learning (Sung et al. 2016) and computer-mediated language learning (Grgurovic et al. 2013).

Instructional approach

The moderator of instructional approach includes nine categories: self-directed ($k=43$, 51%), situated learning ($k=11$, 13%), game-based learning ($k=9$, 11%), collaborative learning ($k=7$, 8%), teacher-led ($k=5$, 6%), flipped learning ($k=3$, 4%), assessment ($k=3$, 4%), mixed ($k=1$, 1%) and unspecified ($k=2$, 2%). The categories of mixed and unspecified approaches were combined due to their small number of eligible studies, with no statistically significant effect found ($g=0.375$, $p>.05$). Over half of the studies

included in this review adopted a self-directed instructional approach, with a medium-to-high effect size reported ($g=0.768$, $p<.001$). A similar effect size was reported for the situated learning approach ($g=0.795$, $p<.001$). A large effect was reported for collaborative learning ($g=0.802$, $p<.05$), and a medium effect for game learning ($g=0.570$, $p<.001$). Adoption of the assessment approach achieved the largest effect ($g=1.168$, $p<.001$), whereas teacher-led instruction had a small effect on language learning ($g=0.373$, $p<.05$). No statistically significant effect was found for the flipped learning approach ($g=0.769$, $p>.05$). According to Q_B statistic ($Q_B=9.802$, $p>.05$), no statistically significant difference existed between the effect sizes of different instructional approaches.

Given that the use of mobile devices allows language learning to take place across time and space, it comes as no surprise that the majority of MALL studies adopted the self-directed learning approach. The positive effect of self-directed mobile learning is also supported by previous meta-reviews (Sung et al. 2015, 2016).

Learning is by nature a situated activity, and authentic activities play an essential role in the acquisition of language skills (Brown et al. 1989). This study finds evidence to support the strong benefit of adopting a mobile-assisted situated learning approach (Godwin-Jones 2017), where mobile devices are used to engage learners in meaningful real-world or place-based learning. For example, Chang (2018) engaged students in place-based learning of English vocabulary in an authentic environment while they were taking a field trip to a zoo. Taking advantage of the location-awareness function of mobile devices, Chang was able to present individualized learning materials to the students based on their specific locations in the zoo. In this way, the learning process is made both personalized and enjoyable, which might have contributed to the learning gains attained. In another study, Hwang et al. (2014) developed a learning system that facilitated elementary students to practice writing by engaging with their familiar contexts at school and reported positive learning outcomes.

Our study also finds support for the positive effect of mobile-assisted collaborative learning, the importance of which has been repeatedly emphasized in prior reviews and studies (Burston 2015; Herrington et al. 2009; Kukulska-hulme and Shield 2008). The mobile version of social media and Web 2.0 tools, such as WhatsApp (Lai 2016; Andújar-Vaca and Cruz-Martínez 2017) and LINE (Chen Hsieh et al. 2017), have been successfully used to promote language learners' interaction and collaboration.

Game-based learning is moderately effective for developing language skills. For instance, Grimshaw and Cardoso (2018) successfully used mobile games to enhance learners' speaking fluency, and Hung (2017) introduced gamified foreign language learning through a BYOD model.

Our analysis also provides evidence that mobile-assisted assessment can be an effective strategy to support language learning. For example, Agbatogun (2014) examined the effectiveness of using traditional clickers on the development of students' communicative competence. Asmali (2018) reported positive effect of providing formative assessment through a BYOD approach in English classes. Consistently, prior research demonstrates that testing is an effective instructional approach for information retention and retrieval (Binks 2018). Mobile systems also support providing feedback in a real-time fashion, which greatly enhances language learning.

Teacher-led instruction has a small-to-medium effect on language learning, and the effect size is the smallest among all the significant moderators. This suggests that mobile learning led by teachers might be less effective than other approaches that best utilize the features of mobile devices, such as informality, collaboration and location-awareness.

However, our findings indicate mobile-supported flipped learning did not outperform traditional flipped learning in the learning outcomes. One possible explanation has to do with the feature of flipped learning that emphasizes lesson previews. For example, Wang (2016) set strict requirements for both the treatment and the control groups to study before class meetings. The extra efforts that the students had invested before the class might account for the small difference between mobile-supported and non-mobile-supported flipped learning groups. However, due to the limited number of included studies ($k=3$), caution should be exercised when generalizing our findings to other mobile-supported flipped learning contexts.

Learning context

We examined the impact of three different mobile learning contexts: classroom ($k=41$, 49%), unrestricted ($k=38$, 45%) and outdoor ($k=5$, 6%). Large effect sizes were reported for language learning in both outdoor ($g=0.959$, $p<.05$) and unrestricted settings ($g=0.804$, $p<.001$), whereas implementing MALL in classrooms yielded a medium effect ($g=0.612$, $p<.001$). According to Q_B statistic, no statistically significant difference existed between the effect sizes of these learning contexts ($Q_B=3.58$, $p>.05$).

The stronger effect of learning with mobile devices in unrestricted and outdoor settings than in formal classrooms is aligned with findings from the previous meta-analysis of MALL studies (Sung et al. 2015). This finding also provides further evidence for the eco-dialogical perspective of learning. According to this view, language learning involves more than mastering a set of linguistic rules; instead, it represents an experience that emerges through the interaction between learners and the context (Zheng and Newgarden 2017; Zheng et al. 2018). Moreover, the eco-dialogical perspective emphasizes the differences between learning environments in their affordances, or the learning opportunities that different contexts can offer (Van Lier 2004). MALL that transcends time and space encourages learners to integrate their formal classroom experiences with real-world life experiences and provides them with plentiful authentic and structured learning tasks to practice language (Sung et al. 2015). Therefore, implementing MALL in outdoor and unrestricted settings tend to achieve larger effects than in classroom settings.

Target language skill

The target language skill variable includes nine categories: grammar ($k=1$, 1%), pronunciation ($k=2$, 2%), listening ($k=4$, 5%), writing ($k=4$, 5%), speaking ($k=5$, 6%), early literacy ($k=8$, 10%), reading ($k=14$, 17%), vocabulary ($k=23$, 27%) and integrated/whole language skill ($k=23$, 27%). Given that both categories of grammar and pronunciation included few studies, they were combined as “others” in the moderator analysis.

Nearly one third of the studies included in this review focused on the effect of MALL on integrated language skills, with a large effect size of 0.824 reported ($p<.001$). Among all the sub-skills studies, vocabulary learning received the greatest attention, accounting for about another one third of all the included studies. A medium-to-large effect was achieved through mobile-assisted vocabulary learning ($g=0.772$, $p<.001$). Using mobile devices to support students’ development of listening ($g=1.080$), speaking ($g=1.056$) and writing skills ($g=1.041$) yielded very high effects, all reaching statistical significance ($p<.001$). A medium effect was found for studies on early literacy ($g=0.493$, $p<.001$).

reading ($g=0.375$, $p<.05$) and other skills ($g=0.487$, $p<.05$). According to Q_B statistic ($Q_B=18.903$, $p<.05$), statistically significant difference existed between the effect sizes of different language skills.

Overall, our results demonstrate a positive effect in favor of using mobile devices to learn different language skills. Our findings regarding the effect of vocabulary learning ($g=0.772$) are also consistent with findings reported in previous research synthesis and meta-analysis of mobile-assisted vocabulary learning (Mahdi 2018; Lin and Lin 2019). For example, in a meta-analysis of 16 mobile-assisted vocabulary studies, Mahdi (2018) also reported a medium-to-high effect. The very large effect of mobile-assisted listening, speaking and writing and the positive effect of mobile-assisted reading further reveal the great potential of using mobile devices to facilitate learning a wide range of language skills. However, it should be noted that among the 14 studies on mobile reading, 12 studies used either laptops, iPad or tablet PCs and all the 12 studies were implemented in classroom settings. In other words, the effects of reading with smaller devices such as smartphones, and the effects of mobile-assisted reading in informal settings were still largely unexplored. Our study shows using mobile devices for early literacy has a medium effect size of 0.493 ($p<.001$), which means mobile devices are also effective for the development of literacy such as letter recognition and letter writing.

Target language

Target language moderator includes seven categories: English ($k=72$, 86%), Chinese ($k=4$, 5%), Spanish ($k=3$, 4%), French ($k=1$, 1%), Norwegian ($k=1$, 1%), Turkish ($k=1$, 1%) and mixed ($k=2$, 2%). As there were very few studies in some categories, the categories of French, Norwegian, Turkish and mixed languages were combined into “others” in the moderator analysis. As is shown in Table 2, research on English language ($k=72$) dominated the field of MALL, with a medium-to-large effect size reported ($g=0.785$, $p<.001$). However, no statistically significant effect was found for mobile learning of Chinese ($g=0.142$, $p>.05$) or Spanish ($g=0.725$, $p>.05$). The effect of mobile devices on learning other languages reached significance with a small effect size ($g=0.300$, $p<.05$). Q_B statistics ($Q_B=25.166$, $p<.001$) indicates the mean effect sizes differed significantly between different target languages.

Our findings are in line with the overall research trend identified in the previous reviews (Burston 2014a; Sung et al. 2015) regarding the *imbalanced* nature of MALL studies, i.e. English is the most-researched language. Our study also finds evidence for the effectiveness of using mobile devices to facilitate English learning (Kukulska-Hulme 2009). However, using mobile devices to learn Chinese or Spanish did not yield statistically significant effects. Given that we were only able to locate a very limited number of eligible studies on learning non-English languages, more research is needed to provide more solid evidence for the effectiveness of using mobile devices to learn languages other than English.

First/second language (L1/L2)

We compared the effect sizes for mobile learning of first ($k=16$, 19%), second ($k=66$, 79%) and mixed ($k=2$, 2%) languages. According to Table 2, using mobile devices to learn a second language produced a large effect ($g=0.825$, $p<.001$). A moderate effect was achieved through mobile-assisted learning of the first language ($g=0.442$, $p<.001$). Mobile devices were not effective in supporting the learning of mixed languages ($g=0.024$, $p>.05$). The value of

Table 3 Result of classic fail-safe *N* test

Item	Statistic
Z-value for observed studies	24.677
p-value for observed studies	.000
Alpha	0.05
Z for alpha	1.95996
Number of observed studies	84
Number of missing studies that would bring p-value > .05	3232

Table 4 Result of Orwin's fail-safe *N* test

Item	Statistic
Hedge's <i>g</i> in observed studies	0.6088
Criterion for "trivial" Hedge's <i>g</i>	0.01
Mean Hedge's <i>g</i> in missing studies	0.0000
Number of missing studies needed to bring Hedge's <i>g</i> under 0.01	5031

Q_B indicates ($Q_B=49.994$, $p<.001$) statistically significant difference existed between the effect sizes of these categories. Using mobile devices to learn a second language was more effective than using them to learn the first language.

Both learner motivation and opportunities to use the language are critical to the success of language learning (Rubin 1975). Features of mobile learning such as easy access to learning materials, informality and multimedia function enable language learners to spend more time on instructional tasks and to practice their language skills in an interesting and enjoyable manner, which might explain the positive effect of learning for both L1 and L2. Our study shows using mobile devices in learning a second language is more effective than learning the first language. This could be due to the difference between L2 and L1 learning (Cook 2010; Ellis 1994) and that instruction plays a more important role in L2 learning than in L1 acquisition (Ellis 1994; Cook 1973). Another explanation for the larger effect size of L2 learning may be that the studies on mobile-assisted learning of a first language all involved preschool or elementary school students as participants; the effect sizes for these groups of learners are smaller than for other learner populations, as discussed earlier in the section of "Educational Level."

Evaluation of publication bias

Classic Fail-safe *N* and Orwin's Fail-safe *N* were adopted to evaluate the presence of publication bias. The tolerance level suggested by Rosenthal (1979) is $5k+10$, i.e. 430 for the current study. According to the result of Classic Fail-safe *N* test, a total number of 3232 studies would be needed to bring the effect size to a non-significant level (Table 3). According to Orwin's Fail-safe *N* test, to bring the effect size to a trivial level of 0.01, 5,031 studies need

to be incorporated into the analysis (Table 4). Therefore, we could draw a conclusion that the impact of publication bias on the effect size was trivial.

Conclusion and implications

Conclusion

This study examined the overall effectiveness of using mobile devices on language learning based on a synthesis of 84 separate studies extracted from journal articles, doctoral dissertations and conference proceedings. We found a medium-to-high overall effect size for mobile devices on language learning achievement, which confirms the positive outcomes of using mobile devices in language learning.

Through the analysis of potential moderator variables, we could draw the following conclusions:

- (1) Educational level, implementation duration, device type, instructional approach, learning context and application type were not statistically significant moderators explaining the differences in effect sizes of MALL.
- (2) Target language was a statistically significant moderator explaining the effect-size variation. In terms of learning outcomes, using mobile devices to learn English is more effective than learning many other languages.
- (3) L1/L2 was a statistically significant variable moderating the variation in effect sizes. Students benefit more from using mobile devices to learn a second language than from engaging in mobile learning of their first language.
- (4) Target language skill was a significant moderator explaining the different effects of MALL. More success has been achieved with adopting mobile devices to enhance the learning of speaking, listening, writing and vocabulary than to learn other subskills such as reading.

Implications for MALL research and application

Our findings have implications for future MALL-related studies and applications. First, our study confirms that language learning through mobile devices is more effective than the conventional instructional approach. Therefore, further investigations to explore the pedagogical potential of MALL should be encouraged. Secondly, although this study shows mobile learning is effective for learning a wide range of language skills under different conditions, research interest in different areas of MALL tends to be relatively unbalanced. For example, English language is the dominant target language, and self-directed learning is the major type of learning approaches in MALL studies. To benefit various kinds of learners in a wider range of contexts, more attention should be given to those areas where the potential of MALL has not been fully explored. Thirdly, our study shows MALL studies employing the situated and the collaborative features of mobile learning produce a high effect. These features of mobile technologies are increasingly transforming the way we live, work and learn, and it is necessary for future research to explore the mediating role that mobile devices play in shaping the relationship between people, technologies and learning contexts.

Limitation of the study

First, this review focused on synthesizing studies reporting cognitive learning outcomes. With increasingly more rigorous MALL studies being conducted that address other learning outcomes such as affective outcomes, future reviews can consider expanding their focus and investigating the impact of mobile learning on non-cognitive outcomes.

Secondly, the impact of potential moderator variables needs to be further explored in future meta-analyses. This review examined the moderating effect of nine potentials variables, and three moderators were found to significantly moderate the effect-size variation in MALL; whereas the previous review by Sung et al. (2015) reported some dissimilar results. Given that some of the sub-categories only had a limited number of eligible studies, further investigations that include more studies would be necessary to confirm some of the findings.

Thirdly, this study adopted a meta-analysis approach to synthesize the results of multiple studies. Since a meta-analysis can only analyze outcomes of studies adopting quantitative research designs, studies adopting other research designs could not be included. It is likely that these studies could provide different and deep insights. Future meta-analyses can supplement their findings with understanding achieved through a systematic review of both qualitative and quantitative research investigations.

Acknowledgements This research is part of the project entitled *Classroom Interaction in Smartphone-supported College English Classes* supported by the Key Projects of Ministry of Education, National Education Sciences Planning Projects of 2017, China (Project Number: DCA170308). We would also like to acknowledge the constructive suggestions provided by the reviewers and the valuable editing suggestions given by Mr. Mitchell Wagner from Indiana University.

References

- Agbatogun, A. (2014). Developing learners' second language communicative competence through active learning: Clickers or communicative approach? *Educational Technology & Society*, 17(2), 257–269.
- Alemi, M., Sarab, M. R. A., & Lari, Z. (2012). Successful learning of academic word list via MALL: Mobile assisted language learning. *International Education Studies*, 5(6), 99–109.
- Andujar, A. (2016). Benefits of mobile instant messaging to develop ESL writing. *System*, 62, 63–76.
- Andújar-Vaca, A., & Cruz-Martínez, M. S. (2017). Mobile instant messaging: Whatsapp and its potential to develop oral skills. *Comunicar*, 25(50), 43–52.
- Asmali, M. (2018). Integrating technology into ESP classes: Use of student response system in English for specific purposes instruction. *Teaching English with Technology*, 18(3), 86–104.
- Binks, S. (2018). Testing enhances learning: A review of the literature. *Journal of Professional Nursing*, 34(3), 205–210.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Burston, J. (2013). Language learning technology MALL: Future directions for BYOD applications. *The IALLT Journal for Language Learning Technologies*, 43(2), 89–96.
- Burston, J. (2014a). The reality of MALL: Still on the fringes. *CALICO Journal*, 31(1), 103–125.
- Burston, J. (2014b). MALL: The pedagogical challenges. *Computer Assisted Language Learning*, 27(4), 344–357.
- Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(1), 4–20.
- Chang, C. C. (2018). Outdoor ubiquitous learning or indoor CAL? Achievement and different cognitive loads of college students. *Behaviour and Information Technology*, 37(1), 38–49.

- Chen, C. M., & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, *51*(2), 624–645.
- Chen Hsieh, J. S., Wu, W. C. V., & Marek, M. W. (2017). Using the flipped classroom to enhance EFL learning. *Computer Assisted Language Learning*, *30*(1–2), 1–21.
- Cho, K., Lee, S., Joo, M.-H., & Becker, B. J. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences*, *8*(3), 105.
- Chou, P. N., Chang, C. C., & Lin, C. H. (2017). BYOD or not: A comparison of two assessment strategies for student learning. *Computers in Human Behavior*, *74*, 63–71.
- Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the British Educational Research Association Annual Conference, England 12–14 September 2002. Retrieved January 25, 2020, from <https://www.leeds.ac.uk/educol/documents/00002182.htm>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cook, V. (1973). The comparison of language development in native children and foreign adults. *IRAL*, *15*, 1–20.
- Cook, V. (2010). The relationship between first and second language learning revisited. In E. Macaro (Ed.), *The Continuum companion to second language acquisition* (pp. 137–157). London: Continuum.
- de la Fuente, M. J. (2014). Learners' attention to input during focus on form listening tasks: The role of mobile technology in the second language classroom. *Computer Assisted Language Learning*, *27*(3), 261–276.
- DeCoster, J. (2004). *Meta-analysis notes*. Retrieved October 15, 2019, from <http://www.stat-help.com>.
- Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., & Smith, H. (1987). Publication bias in clinical trials. *Controlled Clinical Trials*, *8*, 348–353.
- Ducate, L., & Lomicka, L. (2013). Going mobile: Language learning with an ipod touch in intermediate French and German classes. *Foreign Language Annals*, *46*(3), 445–468.
- Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, *27*(2), 197–216.
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*(8746), 867–872.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Language Learning & Technology*, *20*(3), 9–19.
- Godwin-Jones, R. (2017). Smartphones and language learning. *Language Learning & Technology*, *21*(2), 3–17.
- Grgurovic, M., Chapelle, C., & Shelley, M. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, *25*(2), 165–198.
- Grimshaw, J., & Cardoso, W. (2018). Activate space rats! Fluency development in a mobile game-assisted environment. *Language Learning & Technology*, *22*(3), 159–175.
- Herrington, A., Herrington, J., & Mantei, J. (2009). Design principles for mobile learning. In Herrington, J., Herrington, A., Mantei, J., Olney, I. & Ferry, B. (Eds.), *New technologies, new pedagogies: Mobile learning in higher education* (p. 138). Wollongong: Faculty of Education, University of Wollongong.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.
- Hung, H. T. (2017). Clickers in the flipped classroom: Bring your own device (BYOD) to promote student learning. *Interactive Learning Environments*, *25*(8), 983–995.
- Hwang, G.-J., & Fu, Q.-K. (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.
- Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *42*(4), 65–70.
- Hwang, W. Y., & Chen, H. S. L. (2013). Users' familiar situational contexts facilitate the practice of EFL in elementary schools with mobile devices. *Computer Assisted Language Learning*, *26*(2), 101–125.
- Hwang, W. Y., Chen, H. S. L., Shadiev, R., Huang, R. Y. M., & Chen, C. Y. (2014). Improving English as a foreign language writing in elementary schools using mobile devices in familiar situational contexts. *Computer Assisted Language Learning*, *27*(5), 359–378.
- Kukulka-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL*, *21*(2), 157–165.
- Kukulka-hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(3), 271–289.
- Lai, A. (2016). Mobile immersion: An experiment using mobile instant messenger to support second-language learning. *Interactive Learning Environments*, *24*(2), 277–290.

- Lan, Y. J., & Lin, Y. T. (2016). Mobile seamless technology enhanced CSL oral communication. *Educational Technology & Society*, 19(3), 335–350.
- Levy, M., & Kennedy, C. (2005). Learning Italian via mobile SMS. In A. Kukulska-Hulme & J. Traxler (Eds.), *Mobile learning: A handbook for educators and trainers* (pp. 76–83). London: Taylor & Francis.
- Li, L., Chen, G., & Yang, S. (2013). Construction of cognitive maps to improve e-book reading and navigation. *Computers & Education*, 60(1), 32–39.
- Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning*, 30(3–4), 348–365.
- Lin, C. C. (2017). Learning English with electronic textbooks on tablet PCs. *Interactive Learning Environments*, 25(8), 1035–1047.
- Lin, C. J., & Hwang, G. J. (2018). A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Educational Technology & Society*, 21(2), 205–219.
- 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.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55(2), 630–643.
- Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning*, 24(6), 515–525.
- Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning. *Journal of Educational Computing Research*, 56(1), 134–154.
- Naismith, L., Sharples, M., Vavoula, G. & Lonsdale, P. (2004). *Literature review in mobile technologies and learning*. Bristol: University of Birmingham. (FutureLab Series, Report, 11). Retrieved October 20, 2019, from www.futurelab.org.uk/research/lit_reviews.htm.
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157–159.
- Quinn, C. (2000). mLearning: mobile, wireless, in your pocket learning. *LineZine*. Retrieved October 20, 2019, from <https://www.linezine.com/2.1/features/cqmmwiyp.htm>.
- Reinders, H., & Pegrum, M. (2017). Supporting language learning on the move: An evaluative framework for mobile language learning resources. In B. Tomlinson (Ed.), *Second language acquisition research and materials development for language learning* (pp. 116–141). (Second Language Acquisition Research Series). London: Routledge.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Rubin, J. (1975). What the “good language learner” can teach us? *TESOL Quarterly*, 9(1), 41–51.
- Sandberg, J., Maris, M., & De Geus, K. (2011). Mobile English learning: An evidence-based study with fifth graders. *Computers and Education*, 57(1), 1334–1347.
- Saran, M., Seferoğlu, G., & Çağiltay, K. (2012). Mobile language learning: Contribution of multimedia messages via mobile phones in consolidating vocabulary. *The Asia-Pacific Education Researcher*, 21(1), 181–190.
- Shadiev, R., Hwang, W. Y., & Liu, T. Y. (2018). A study of the use of wearable devices for healthy and enjoyable English as a foreign language learning in authentic contexts. *Educational Technology & Society*, 21, 217–231.
- Sharples, M., Milrad, M., Arnedillo-Sánchez, I., Milrad, M., & Vavoula, G. (2009). Mobile learning: Small devices, big issues. In N. Balacheff, S. Ludvigsen, T. Jong, A. Lazonder, & S. Barnes (Eds.), *Technology enhanced learning: Principles and products* (pp. 233–249). Heidelberg: Springer.
- Sharples, M., Taylor, J., & Vavoula, G. (2005). Towards a theory of mobile learning. *Proceedings of mLearn*, 1, 1–9.
- Silverio-Fernández, M., Renukappa, S., & Suresh, S. (2018). What is a smart device?—A conceptualisation within the paradigm of the internet of things. *Visualization in Engineering*, 6(1), 3.
- Song, Y., & Fox, R. (2008). Using PDA for undergraduate student incidental vocabulary testing. *ReCALL*, 20(3), 290–314.
- Stockwell, G., & Hubbard, P. (2013). *Some emerging principles for mobile-assisted language learning*. Monterey, CA: The International Research Foundation for English Language Education. Retrieved November 1, 2019, from <https://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning>.
- 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.

- 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>.
- Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, 21, 217–228.
- Traxler, J. (2005). *Defining mobile learning*. Paper presented at IADIS International Conference Mobile Learning 2005 (pp. 261–266).
- Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Dordrecht: Springer.
- Viberg, O., & Grönlund, A. (2012). *Mobile assisted language learning: A literature review*. Paper presented at the 11th World Conference on Mobile and Contextual Learning, Helsinki, Finland. mLearn 2012. 1–8.
- Wang, Y. H. (2016). Could a mobile-assisted learning system support flipped classrooms for classical Chinese learning? *Journal of Computer Assisted Learning*, 32(5), 391–415.
- Winters, N. (2006). What is mobile learning? In M. Sharples (Ed.), *Big issues in mobile learning: Report of a workshop by the kaleidoscope network of excellence mobile learning initiative* (pp. 5–9). Nottingham: University of Nottingham.
- Wu, Q. (2015). Designing a smartphone app to teach English (L2) vocabulary. *Computers & Education*, 85, 170–179.
- Wu, T. T., Sung, T. W., Huang, Y. M., Yang, C. S., & Yang, J. T. (2011). Ubiquitous English learning system with dynamic personalized guidance of learning portfolio. *Educational Technology & Society*, 14(4), 164–180.
- Wu, W. C., Chen Hsieh, J., & Yang, J. C. (2017). Creating an online learning community in a flipped classroom to enhance EFL learners' oral proficiency. *Educational Technology & Society*, 20(2), 142–157.
- Zheng, D., Liu, Y., Lambert, A., Lu, A., Tomei, J., & Holden, D. (2018). An ecological community becoming: Language learning as first-order experiencing with place and mobile technologies. *Linguistics and Education*, 44, 45–57.
- Zheng, D., & Newgarden, K. (2017). Dialogicality, ecology, and learning in online game worlds. In S. L. Thorne & S. May (Eds.), *Language, education and technology* (pp. 345–359). Cham: Springer International Publishing.
- Zheng, L., Li, X., & Chen, F. (2016). Effects of a mobile self-regulated learning approach on students' learning achievements and self-regulated learning skills. *Innovations in Education and Teaching International*, 55(6), 1–9.
- Zurita, G., & Nussbaum, M. (2004). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computer & Education*, 42, 289–314.

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

Zhenzhen Chen Associate Professor, Deputy Dean of School of Humanities, Beijing University of Posts and Telecommunications. She is currently a doctoral student at Graduate School of Education, Peking University. Her research interest includes technology enhanced language learning and mobile assisted language learning.

Weichao Chen PhD, is an instructional designer at the University of Virginia School of Medicine. Her research interests include educational design research, faculty development, medical education, educational technology, and language teaching. Jiyou Jia, PhD, and Weichao Chen, PhD, share senior authorship on this paper.

Jiyou Jia Professor and Head of Department of Educational Technology, Graduate School of Education, Peking University, China. His research interest includes AI in education and CALL.

Huili An is a graduate student at School of Humanities, Beijing University of Posts and Telecommunications.