Language Learning & Technology
ISSN 1094-3501
2024, Volume 28, Issue 1
pp. 1–26

ARTICLE

Effects of mobile-assisted language learning on foreign language learners' speaking skill development

Rui Li, Hunan University

Abstract

Despite the growing body of research regarding the effectiveness of MALL (mobile-assisted language learning) technologies on foreign language (FL) learners' speaking skill development, a comprehensively quantitative meta-analysis regarding the effect sizes of these studies is still lacking. To solve the problem, this study reported results based on a meta-analysis of 20 effect sizes among 932 participants from 18 experimental and quasi-experimental studies. The results showed that the overall effect size was significantly large, suggesting the use of MALL for FL learners' speaking skill development is more effective than traditional methods. Furthermore, learner-related, instruction-related and methodology-related moderator analysis results indicated that instructional approaches and intervention durations were significant moderators, while proficiency levels, educational levels, language types, intervention settings, software types, measured outcome types and duration intensity did not find a significant moderating effect. The results of the study provide some pedagogical implications into the use of MALL technologies for FL learners' speaking skill development.

Keywords: *Evidence-based Applied Linguistics (EBAL), Foreign Language (FL), Meta-analysis, Mobile-assisted Language Learning (MALL), Speaking Skill*

Language(s) Learned in This Study: *English, Chinese, French*

APA Citation: Li, R. (2024). Effects of mobile-assisted language learning on foreign language learners' speaking skill development. *Language Learning & Technology*, *28*(1), 1–26. https://hdl.handle.net/10125/73553

Introduction

In recent years, the emergence of mobile-assisted language learning (MALL) technologies has garnered a great deal of attention in the field of computer assisted language learning (CALL) and educational technology. MALL technologies can provide foreign/second language (FL/L2) learners with portable, connective, context-sensitive, location-aware, multifunctional and ubiquitous learning environments (Sung et al., 2015), which enables them to seamlessly practice their speaking skill (Andujar & Salaberri-Ramiro, 2021). To measure MALL effectiveness, researchers often adopted experimental research methods using mobile technologies to develop FL learners' speaking skill, such as WhatsApp (Rezaee et al., 2019), WeChat (Mu, 2017) and mobile speaking systems (Hwang et al., 2014). In these studies, the experimental group was exposed to MALL technologies while the control group used traditional methods, including traditional pencil and paper instruction (e.g., Hwang, Shih, Ma, Shadiev, & Chen, 2016; Ko & Lim, 2022), traditional classroom instruction (e.g., Lan & Lin, 2016; Liakin, Cardoso, & Liakina, 2017), oral assignments at home (Sun et al., 2017) and non-game-based learning (e.g., Grimshaw & Cardoso, 2018; Liu & Chu, 2010), among others. Research into the use of MALL for FL speaking skill development is crucial, because in foreign language learning speaking is one of the most challenging skills for learners (Li, 2022a, 2022d). In addition, the mobility and ubiquity features of MALL technologies outperform the traditional foreign classroom that "is unable to offer enough opportunities for speaking practice" (Ahn & Lee, 2016, p.779).

Researchers have paid much attention to investigating the effects of MALL for FL learners' speaking skill development on learning outcomes, and numerous studies have obtained positive effects, such as increasing motivation, critical thinking, decreasing anxiety (Chien, Hwang, & Jong, 2020), promoting self-efficacy (Ko & Lim, 2022), enhancing engagement (Andujar & Salaberri-Ramiro, 2021), and improving the FL speaking performance (Ko & Lim, 2022), however synthesized empirical evidence of the overall effects is still in paucity.

Furthermore, in their recent critical review, Hsu and Liu (2021) reported that learner-related (proficiency levels, educational levels and language types), instruction-related (instructional approaches and intervention settings) and methodology-related (software types, intervention durations and measured outcome types) factors might moderate the effects of MALL for FL learners' speaking skill development on learning outcomes. These results afforded new insights into the general research trends addressing MALL for FL learners' speaking skill development. These factors' moderating effects (henceforth also referred to as 'moderators') remain to be identified, since Hsu and Liu neither directly calculated the aggregated effects of MALL for FL learners' speaking skill development, nor dealt with the variability of the aggregated effects with moderator analyses. This suggests an urgent need to meta-analyze the effects of MALL for FL learners' speaking skill development and examine whether the calculated effects were moderated by of the learner-related, instruction-related and research methodology-related factors or moderators. The need to fill this gap prompted the researchers to provide a novel evidence-based applied linguistic perspective on MALL for FL learners' speaking skill development, using a meta-analysis of rigorously peer-reviewed empirical research.

Literature Review

Related MALL Studies for FL Learners' Speaking Skill Development

MALL for FL learners' speaking skill development refers to the use of MALL technologies for the development of speaking-related skills, including foreign language pronunciation, monologues and dialogic communication (Hsu & Liu, 2021, see Appendix A for details). Researchers have found that MALL for FL learners' speaking skill development can be used to propel self-regulated learning with the support of automatic speech recognition (ASR) technology (Ahn & Lee, 2016), enhancing the interactivity and personalization of collaborative learning (Lan & Lin, 2016), and improving foreign language learning affects (Andujar & Salaberri-Ramiro, 2021; Ko & Lim, 2022). For instance, Chang and Lan (2021) adopted a quasi-experimental design to investigate the effect of a flipped foreign language classroom incorporating LINE (a chat mobile APP) on learner-related factors, namely, speaking performance and perceptions. Results indicated that learners who used the outside-of-the-classroom LINE had higher speaking performance, a positive impact on how learners perceived foreign language anxiety and increased engagement in speaking activity. Similarly, Ahn and Lee (2016) examined learners' use experience of *Speaking English 60 Junior* (a mobile APP) which featured ASR technology. They found that learners held positive attitudes and interest towards the use of the application as a platform for practicing to speak. In another quasi-experiment, Ko and Lim (2021) designed a mobile APP named WikiTalki that was able to offer personalized peer feedback to learners' speaking performance. A positive outcome regarding the use of WikiTalki was achieved in that it enhanced learner engagement and collaborative learning as it was endowed with instruction-related functionalities, e.g., audio recording and peer sharing.

However, other studies found that MALL for FL learners' speaking skill development produced limited pedagogical benefits in terms of speaking outcomes. For instance, Lan and Lin (2016) examined the effects of a mobile seamless learning platform (MOSE) on learners' Chinese-as-a-second-language (CSL) speaking performance with a quasi-experimental design. The results indicate that there exists no significant difference in learners' speaking performance between traditional classroom learning and mobile learning.

Likewise, in a critical review of 28 empirical studies published between 2010 to 2019, Hsu and Liu (2021) also posited that if learners lack self-direction or sustained motivation in MALL use, their speaking outcomes might not be improved. Furthermore, they discovered that the improved speaking outcomes might be influenced or moderated by a series of learner-related (proficiency levels, educational levels and language types), instruction-related (instructional approaches and intervention settings) and methodology-related (software types, intervention durations and measured outcome types) factors. These mixed results indicate that the effects of MALL for FL learners' speaking skill development on learning outcomes remain open for further investigations, therefore justifying the need to scrutinize which learner-related, instruction-related and methodology-related factors may underline such variability in the effectiveness of MALL for FL learners' speaking skill development.

Related Reviews and Meta-analyses of MALL Studies

In the past decades, several qualitative critical reviews and quantitative meta-analyses on MALL have been conducted (see Appendix B). Although these studies may provide valuable insights into domain-general language skills, a closer scrutiny of these studies suggests that domain-specific skills remain poorly researched (e.g., Hsu & Liu, 2021 for MALL-related speaking skill development; Mahdi, 2017 for MALL-related vocabulary learning) and warrant further attention.

On the other hand, while numerous empirical studies (e.g., Andujar & Salaberri-Ramiro, 2021; Grimshaw & Cardoso, 2018; Ko & Lim, 2022) on MALL for FL learners' speaking skill development have been done, it is still unclear what the research trends should be in an attempt to gain a holistic understanding of the current state. To the best of our knowledge, only Hsu and Liu (2021) adopted a qualitative review, reporting that MALL enables educators to create learner-centered self-regulated learning environments and facilitate a collaborative-based instruction through self-evaluation, self-instruction and self-correction. Given the rapid progress in MALL technologies that have progressively led to accumulating sufficient multiple data sources over the past two decades, it is therefore necessary to further examine the effectiveness of MALL for FL learners' speaking skill development with sound evidence from an evidence-based applied linguistic (EBAL) perspective (Li, 2023a, 2023b).

The Present Study

What emerges from the above review is that despite the growing diversity of studies, meta-analysis research addressing MALL for FL learners' speaking skill development is still in its infancy. In this regard, one crucial domain of study that remains underexplored concerns the quantitative meta-analysis of converging empirical evidence. Subsequently, two objectives are derived from this purpose. First, motivated by existing meta-analyses regarding the inclusion criteria (e.g., Chen et al., 2020; Li, 2022c; Wang et al., 2020; Xu et al., 2019), a meta-analysis was conducted on the aggregated overall effect sizes of MALL for FL learners' speaking skill development. Second, drawing on several existing studies (Chen et al., 2020; Hsu & Liu, 2021; Sung et al., 2015, 2016), the moderating effects were assessed in relation to learner-related (proficiency levels, educational levels and language types), instruction-related (instructional approaches and intervention settings) and methodology-related (software types, intervention durations, durations and measured outcome types) moderators. Consequently, two research questions were developed.

RQ1: What is the overall effect size of MALL vs. non-MALL for FL learners' speaking skill development?

RQ2: How do learner-related, instruction-related and methodology-related moderators affect the aggregated effect sizes?

Method

Literature Search

Exhaustive literature search procedures were adopted to examine the effectiveness of MALL for FL learners' speaking skill development. First, the potential keywords and/or keyword combinations were consulted based on Hsu and Liu (2021) to promote a comprehensive retrieval. It is noteworthy here that, while Hsu and Liu's review focuses on a narrative account of information that is already currently available, this study adopts a mathematical synthesis of the results. As such, only (quasi-)experiments that reported control and treatment/experimental groups were included in data collection. Related primary studies were searched from several electronic online databases (e.g., Web of Science, ScienceDirect, Springer, ProQuest, Scopus, Wiley and ERIC) and search engines (Google Scholar and Baidu Scholar) by using a combination of the following MALL-related and speaking-related keywords integrated with Boolean operators, including the 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 seamless learning OR ubiquitous learning OR u-learning OR mobile phone OR cell phone OR smartphone OR iPad OR iPhone OR tablets OR personal digital assistants OR PDAs) AND (speaking OR spoken OR pronunciation OR fluency OR speaking skill OR speaking ability OR speaking performance OR verbal skill OR oral OR communication) AND (control OR experiment OR experimental OR treatment). Second, a second round of backward and forward citation search that checked literature review and references sections of the selected literature was conducted to further avoid missing out a significant portion of the relevant literature, after the first round used a snowballing technique (Biernacki & Waldorf, 1981). Third, this paper drew on existing studies (Li, 2022a, 2022b, 2022c; Sung et al., 2015; Sung, Lee, Yang, & Chang, 2019), manually searched each of the following CALL journals (e.g., *Language Learning & Technology*, *Computer Assisted Language Learning*, *ReCALL*, *System* and *CALICO Journal*) and educational journals (*Computers & Education*, *Internet and Higher Education*, *Computers in Human Behavior*, *British Journal of Educational Technology*, *Interactive Learning Environments*, *Educational Technology & Society*, *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*, and *The Asia-Pacific Education Researcher*, among others) to avoid excluding key literature.

Inclusion and Exclusion Criteria

The initial literature retrieval yielded 63 primary studies related to MALL for FL learners' speaking skill development. To ensure the eligibility of the included studies, the following inclusion and/or exclusion criteria were proposed:

Time range: Studies published during 2000 to 2021 (until July, 2021) were considered, because MALL technologies remained few in number before 2000 (Duman et al., 2015).

Literature type: Studies should be peer-reviewed publications written in English (Peng et al., 2021), since Burston and Giannakou (2022, p.157) who comprehensively meta-analyzed MALL reported "all but two were written in English, one in Korean, and one in Turkish". Non-peer-reviewed publications or those not written in English were excluded. Furthermore, other review publications (e.g., review articles, book reviews, and editorial materials) were excluded as well.

MALL technologies: Studies should adopt a form of MALL devices/tools (e.g., mobile phones, PDAs and tablets) on foreign or second language speaking skill development. Those studies that failed to use educational technologies, or used CALL technologies, or used MALL technologies for first language speaking skill development were excluded (Chen et al., 2020; Sung et al., 2015, 2016).

Effect size calculation: Studies should report sufficient data for the effect size calculation. In doing so, only the experimental or quasi-experimental studies that examined the effectiveness of a mobile-assisted device on the performance of foreign or second language speaking skill (viz. standardized or researcher-designed speaking test scores) were included. More specifically, the independent variables 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 includes a measure of the researcher-designed or standardized pre- and post-tests between different modes. Publications that investigated only L2 learners' motivation, attitudes or perceptions, pedagogical or theoretical recommendations were excluded. As a result, 18 eligible publications were finalized for meta-analysis.

Study Quality Analysis Instrument

As Valentine (2019, p.130) puts it, study quality that "can affect study results, which can in turn affect the conclusions drawn" is one of the most important issues to consider. It is thus necessary to "develop an explicit, transparent, and reproducible instrument of assessing study quality" (Valentine, 2019, p.130). Informed by Valentine (2019), we adopted the commonly used study quality instrument, viz. the Medical Education Research Study Quality Instrument (MERSQI) (Appendix C, Reed et al., 2007), to evaluate the methodological quality of the selected studies used for this meta-analysis. Although the MERSQI was originally developed to evaluate the study quality of medical education research, it has been increasingly proven to be discipline neutral and commonly applied in educational research (e.g., Jensen & Konradsen, 2018; Wu et al., 2020). The MERSQI contains six domains, including study design, sampling, type of data, validity of evidence for evaluation instrument scores, data analysis and outcome. Each domain has a maximum score of 3, making a maximum total score of 18 and a potential range 5 to 18. It was reported that the average total score of 11 or so could be taken as the benchmark of satisfied study quality. For instance, Cook and Reed (2015) achieved an average score of 11.3, with a range of 8.9 to 15.1. Wu and colleagues (2020) found that the mean MERSQI score was 12.6, with a range of 10.4 to 14.5. Likewise, Jesen and Konradsen (2018), in their meta-analysis, obtained an average total score of 10.9, with a range of 6 to 14.5. Our meta-analysis showed that the average total score is 13.833 (SD = 0.686), with a range of 12.5 to 15, suggesting that the selected primary studies met the criteria for the coming meta-analysis.

Moderators' Coding Schemes

According to Wilson (2019, p.154), a well-designed code 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". Detailed information for the proposed code scheme of moderators was presented in Appendix D.

After the code scheme was developed, coding procedures were observed as follows: First, given the recommended practice for data dependencies (Li, 2023a, 2023b), 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 18 eligible primary studies that yielded 20 effect sizes as independent studies. For instance, Ataeifar et al. (2019) conducted a quasi-experimental design that compared two experimental groups and a control group. The effect sizes were coded separately, since foreign language learners of the experimental groups involved different participants with mobile Twitter and Voicethread, respectively, while those of the control group adopted the traditional approach with Series books. Further, two coders negotiated with each other in the study to ensure the consistent understanding of each item. Then, they independently coded the items, and the interrater reliability (percentage of agreement) was 96.3%. The discrepancies were resolved by consensus through discussions (e.g., Chen et al., 2020; Li, 2022b, 2022c).

Analysis of the Effect Sizes

Hedges' g was taken as the effect size index, as it could "provide a simple correct for the bias of small sample size" (Lipsey & Wilson, 2001, p.48). For most studies, Hedges' g was calculated based on means, sample sizes and standard deviations for the experimental and control group. When the means, sample sizes and standard deviations were not sufficiently reported, other statistical values, e.g., t-value, confidence interval, difference in means and sample sizes, were used to calculate effect sizes. For instance, we used difference in means, lower and upper limits, sample sizes for both the experimental and control group (Ataeifar et al., 2019; Mu, 2017), and means, sample sizes and t-value (Chang & Lan, 2021; Hwang et al., 2014), to compute the effect sizes. The interpretations of the magnitude of an effect size were based on Li (2023a): 0.200, 0.500, and 0.800 for small, moderate, and large effects, respectively.

Characteristics of Included Studies

Table 1 demonstrated the main characteristics of 18 primary studies included in the meta-analysis (see Appendix A for more). The 18 studies involved participants with seven different first languages, indicating that the research on MALL for FL learners' speaking skill development has been internationally widespread. While most of the studies used MALL for FL learners' speaking skill development to learn a foreign language, only two studies used it to learn a second language. Participants represented all educational levels: primary, secondary and tertiary levels. The majority of learners were college students of the tertiary educational level, but only a small number of learners were of the primary and secondary levels. The sample sizes ranged from 16 to 107 participants, and most were within the scope of less than 100 participants. The total sample size was 932 ($N_{\text{experimental}} = 497$, $N_{\text{control}} = 435$). For the consistence of intervention durations, the original data were transformed into units of week, ranging from 0 (representing 'one session') to 18 weeks. As far as journals are concerned, 27.78% were from CALL journals ($k_{\text{total}} = 5$: $k = 3$ from *Computer Assisted Language Learning*, $k = 1$ from *Language Learning & Technology* and $k = 1$ from *ReCALL*) and the remainder were from educational journals ($k_{\text{total}} = 13$, e.g., $k = 3$ from *Educational Technology & Society*, $k = 2$ from *Computers & Education*, among others), resonating with Li (2022a) who claimed the interdisciplinary nature of MALL in general and MALL for FL learners' speaking skill development in particular.

Table 1

Main Characteristics of 18 Included Studies

No.	Study	L1*	LT	TL	EL*	*N* (E/C)	D (weeks)
1	Ali et al. (2015)	Malaysian	FL	English	tertiary	40 (20/20)	0
2	Ataeifar et al. (2019)	Persian	FL	English	tertiary	60 (30/30)	12.857
3	Chang & Lan (2021)	Mandarin	FL	English	tertiary	60 (30/30)	18
4	Chew et al. (2018)	Mandarin	FL	English	tertiary	40 (20/20)	4
5	Chien et al. (2020)	Mandarin	FL	English	secondary	69 (33/36)	4
6	Fang et al. (2021)	Mandarin	FL	English	tertiary	66 (36/30)	3
7	Grimshaw & Cardoso (2018)	French	L2	English	tertiary	20 (10/10)	6
8	Hwang et al. (2014)	Mandarin	FL	English	primary	33 (33/NA)	0
9	Hwang et al. (2016)	Mandarin	FL	English	secondary	40 (20/20)	4
10	Ko & Lim (2022)	Korean	FL	English	secondary	107 (56/51)	4
11	Lan & Lin (2016)	Thai	FL	Chinese	tertiary	34 (18/16)	4
12	Liakin et al. (2017)	English	L2	French	tertiary	18 (9/9)	7
13	Lin & Hwang (2018)	Mandarin	FL	English	tertiary	49 (33/16)	18
14	Liu & Chu (2010)	Mandarin	FL	English	secondary	64 (32/32)	8
15	Mu (2017)	English	FL	Chinese	tertiary	64 (32/32)	16
16	Rezaee et al. (2019)	Iranian	FL	English	tertiary	80 (40/40)	4
17	Saran et al. (2009)	Turkey	FL	English	tertiary	16 (8/8)	4
18	Sun et al. (2017)	Mandarin	FL	English	primary	72 (37/35)	12

Notes. L1 = first language; LT = language types; FL = foreign language; L2 = second language; TL = target language; EL = educational levels; N = number of sample sizes; E/C = number of experimental group/number of control group; D = durations; *Since very few of these L1 are actually stated explicitly by the authors of these studies, the attribution of L1 is only a presumption based on authors' affiliation information when the L1 information is absent; *Educational levels of samples were annotated based on "Participants" section of these studies.

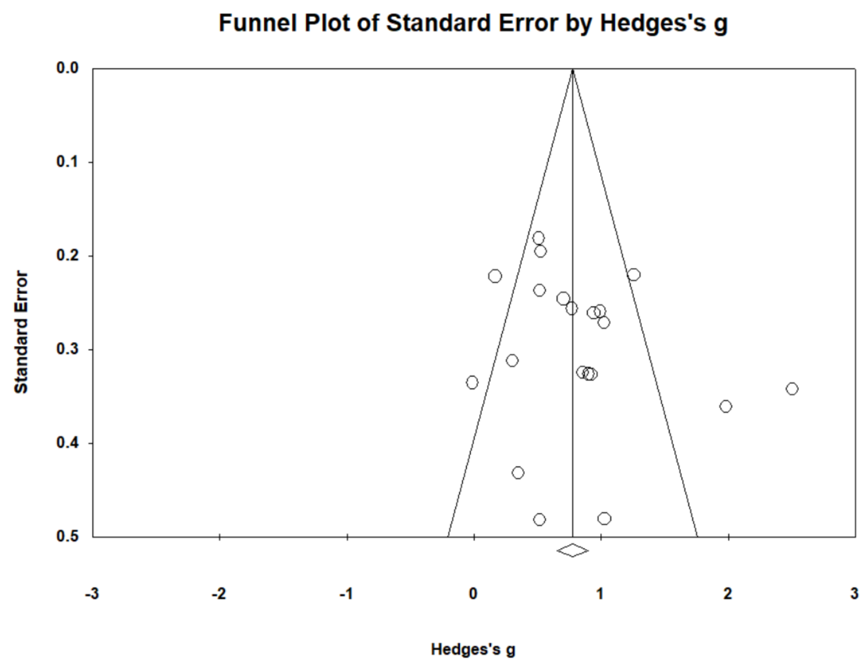
Results

Publication Bias Analysis Results

The publication bias refers to the phenomenon that researchers normally did not publish non-significant results, and unpublished studies might differ from the published studies (Borenstein et al., 2005, 2009). Based on the existing studies (Borenstein et al., 2009; Li, 2022c, 2022d), the inspection of publication bias was based on funnel plot (Figure 1) and a fail-safe *N* method. First, as the symmetrical distribution was apparent in Figure 1, the probability of having a publication bias is rare. In addition, by evaluating how many unpublished studies with non-significant results would change the meta-analytic results from significant to non-significant, 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 834, which was significantly higher than the respective observed number 20 ($Z = 12.801$, $p < 0.001$), that Rosenthal (1991) suggested for the file-drawer problem.

Figure 1

Funnel Plot of Standard Error by Hedges' g for the Selected Studies



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), which was summarized in Table 2. As shown in Table 2, the overall effect size computed from the 20 effect sizes was large, $g = 0.824$, 95% CI [0.594, 1.054] and significant, $Z(19) = 7.016$, $p < 0.001$, indicating that the use of MALL for FL learners' speaking skill development is more effective than traditional methods.

Table 2

Overall Effect Size Results of the Included Studies

		95% CI			Heterogeneity	
k	***g***	**LL**	**UL**	***Z*-value**	***Q*-value**	I^2
20	0.824	0.594	1.054	7.016***	65.728***	71.093

Notes. *k* = number of effect sizes; *g* = Hedges' g; LL = lower limit, UL = upper limit; 95 % CI = 95% confidence interval; ****p* < 0.001.

Moderator Analysis Results

The heterogeneity results in Table 2 also indicated that there were between-group differences among the effect sizes resulting from factors other than subject-level sampling error (Q-value = 65.728, $p < 0.001$). The I^2 for the overall model showed the high heterogeneity ($I^2 = 71.093$), indicating that one or more moderators could account for the heterogeneity (Borenstein et al., 2005, 2009). Hence, moderator analysis results regarding learner-related (proficiency levels, educational levels and language types), instruction-related (instructional approaches and intervention settings) and methodology-related (software types, intervention durations, durations and measured outcome types) moderators were reported in the remainder of this section.

Learner-Related Moderators

Table 3 reported the results of moderator analysis regarding learner-related moderators, including proficiency levels, educational levels and language types.

Proficiency levels: High proficiency learners achieved large effect size ($g = 1.749$, 95% CI [0.298, 3.200]), followed by low proficiency learners ($g = 0.745$, 95% CI [0.530, 0.959]) and intermediate proficiency learners ($g = 0.571$, 95% CI [0.119, 1.024]). However, the moderating effect of proficiency levels was non-significant, $Q = 2.385$, $p = 0.304$.

Educational levels: Learners of tertiary educational level achieved a large effect size ($g = 0.908$, 95% CI [0.562, 1.253]), followed by learners of secondary educational level ($g = 0.708$, 95% CI [0.469, 0.948]) and learners of primary educational level ($g = 0.513$, 95% CI [0.231, 0.795]). However, the moderating effect of education levels was non-significant, $Q = 3.057$, $p = 0.217$.

Language types: Moderator analysis results between foreign language ($g = 0.837$, 95% CI [0.593, 1.080]) and second language ($g = 0.656$, 95% CI [-0.006, 1.317]) were not significant, $Q = 0.253$, $p = 0.615$.

Table 3

Moderator Analysis Results of Learner-related Moderators

			95% CI			Heterogeneity		
Moderators	***k***	***g***	**LL**	**UL**	***Z*-value**	***Q*-value**	***df***	***p***
Proficiency levels	20	0.731	0.539	0.923	7.461***	2.385	2	0.304
1. low	16	0.745	0.530	0.959	6.808***			
2. intermediate	2	0.571	0.119	1.024	2.474*			
3. high	2	1.749	0.298	3.200	2.363*			
Educational levels	20	0.688	0.527	0.850	8.354***	3.057	2	0.217
1. primary education	2	0.513	0.231	0.795	3.564***			
2. secondary education	4	0.708	0.469	0.948	5.797***			
3. tertiary education	14	0.908	0.562	1.253	5.152***			
Language types	20	0.815	0.586	1.043	6.986***	0.253	1	0.615
1. FL	18	0.837	0.593	1.080	6.729***			
2. L2	2	0.656	-0.006	1.317	1.931			

Notes. k = number of effect sizes; g = Hedges' g; LL = lower limit, UL = upper limit; FL = foreign language, L2 = second language; 95 % CI = 95% confidence interval; * $p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Instruction-Related Moderators

Table 4 reported the results of moderator analysis regarding instruction-related moderators, including instructional approaches and intervention settings.

Instructional approaches: The marginally significant effect of instructional approaches was found, Q =

10.692, p = 0.058. Post-hoc comparison to locate the source of significant effect showed that drill and practice (g = 1.299, 95% CI [0.431, 2.166]), flipped classroom (g = 1.292, 95% CI [0.690, 1.894]), multimedia learning (g = 0.911, 95% CI [0.459, 1.363]) and game-based approaches (g = 0.807, 95% CI [0.447, 1.167]) had high effect sizes without a significant between-group difference, Q = 2.485, p = 0.478, which was significantly higher than peer feedback (g = 0.528, 95% CI [0.313, 0.743]), Q = 9.298, p = 0.054, and task-based approach (g = 0.434, 95% CI [0.019, 0.848]), Q = 10.864, p = 0.054. No significant difference between peer feedback and task-based approach was observed, Q = 0.161, p = 0.688.

Intervention settings: The outside classroom setting (g = 0.987, 95% CI [0.425, 1.549]) had a high moderating effect, while classroom setting (g = 0.698, 95% CI [0.546, 0.850]) had a moderate effect size. However, the moderating effect of intervention settings was non-significant, Q = 0.948, p = 0.330.

Table 4

Moderator Analysis Results of Instruction-related Moderators

			95% CI			Heterogeneity		
Moderators	***k***	***g***	**LL**	**UL**	***Z*-value**	***Q*-value**	***df***	***p***
Instructional approaches	20	0.677	0.527	0.828	8.831***	10.692†	5	0.058
1. drill and practice	4	1.299	0.431	2.166	2.935***			
2. flipped classroom	3	1.292	0.690	1.894	4.208***			
3. game-based	3	0.807	0.447	1.167	4.389***			
4. multimedia learning	2	0.911	0.459	1.363	3.950***			
5. peer feedback	4	0.528	0.313	0.743	4.824***			
6. task-based	4	0.434	0.019	0.848	2.052***			
Intervention settings	20	0.718	0.571	0.864	9.587***	0.948	1	0.330
1. classroom	12	0.698	0.546	0.850	9.000***			
2. outside classroom	8	0.987	0.425	1.549	3.443***			

Notes. k = number of effect sizes; g = Hedges' g; LL = lower limit, UL = upper limit; FL = foreign language, L2 = second language; 95 % CI = 95% confidence interval; †p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Methodology-Related Moderators

Table 5 and Figure 2 presented the results of moderator analysis regarding methodology-related moderators, including a continuous (duration intensity) and four categorical moderators (software types, intervention durations and measured outcome types).

Software types: Software for general purposes (g = 1.077, 95% CI [0.539, 0.851]) had a high effect size, while software for educational purposes (g = 0.655, 95% CI [0.491, 0.819]) had a moderate effect size. No significant between-group difference of software types was observed, Q = 2.411, p = 0.120.

Intervention durations: Significant moderating effect of intervention durations was found, Q = 7.058, p = 0.029. Post-hoc comparison indicated that long durations (g = 1.168, 95% CI [0.756, 1.580]) had a high moderating effect, which was significantly higher than the other two duration levels, $Q_{\text{short-long}}$ = 6.980, p = 0.008, and $Q_{\text{intermediate-long}}$ = 3.702, p = 0.054.

Duration intensity: Since the random-effect meta-regression analysis with 10 effect sizes and no covariate (viz. only one predictor) offers a good power to detect nonzero effect (Berkey et al., 1995), the current study with a total of 20 effect sizes that reported the intervention durations met the required criteria of meta-regression analysis. Figure 2 showed the rising regression line between duration intensity and effect sizes (in Hedges' g). The duration intensity was not found to be a significant predictor of the effect sizes (Q_{model} = 0.278, β = -0.001, p = 0.598). In other words, intensity of treatment durations per week tended not to moderate the effect sizes.

Measured outcome types: It was found that effect sizes did not vary based on whether the standardized or researcher-designed tests were used, $Q = 0.824$, $p = 0.364$, indicating the robustness of outcome data regardless of difference in the measured outcome types.

Table 5

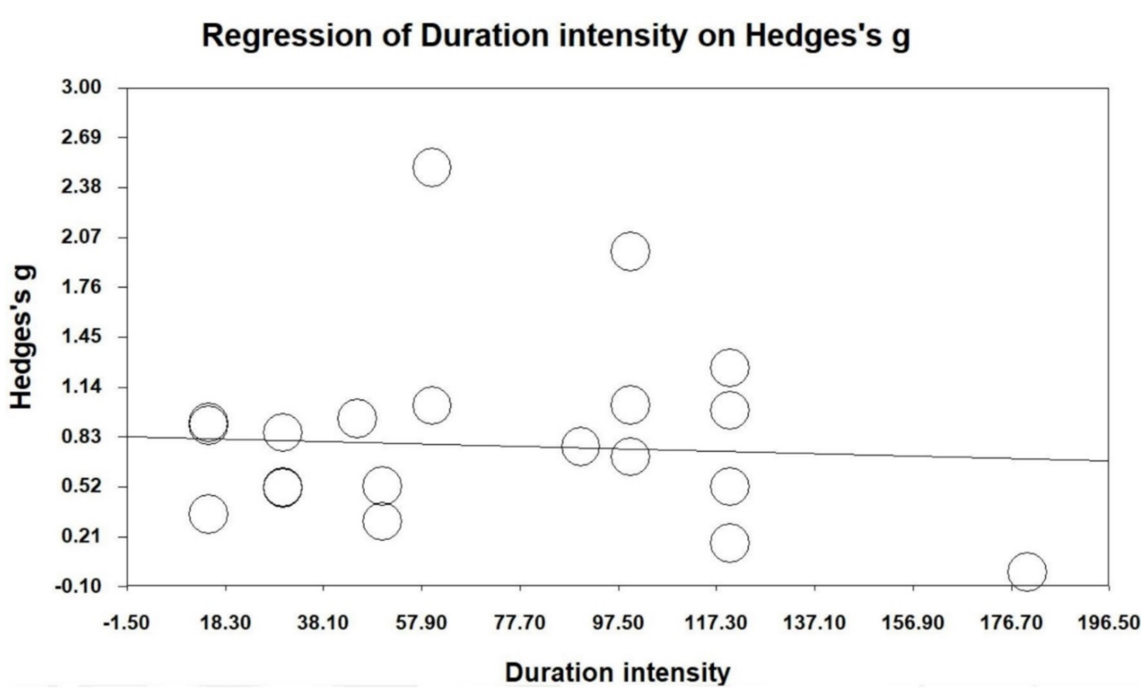
Moderator Analysis Results of Methodology-related Moderators

			95% CI			Heterogeneity		
Moderators	***k***	***g***	**LL**	**UL**	***Z*-value**	***Q*-value**	***df***	***p***
Software types	20	0.695	0.539	0.851	8.729***	2.411	1	0.120
1. general purposes	8	1.077	0.570	1.583	4.166***			
2. educational purposes	12	0.655	0.491	0.819	7.827***			
Intervention durations	20	0.683	0.515	0.850	7.997***	7.058*	2	0.029
1. one session, ≤1 week	3	0.678	0.393	0.964	4.654***			
2. >1 week, ≤4 weeks	8	0.523	0.284	0.762	4.295***			
3. >4 weeks, ≤one semester	9	1.168	0.756	1.580	5.560***			
Measured outcome types	20	0.787	0.566	1.008	6.988***	0.824	1	0.364
1.standardized	7	0.997	0.492	1.503	3.869***			
2.researcher-designed	13	0.737	0.492	0.983	5.890***			

Notes. k = number of effect sizes; g = Hedges' g; LL = lower limit, UL = upper limit; 95 % CI = 95% confidence interval; $^{***}p < 0.001$.

Figure 2

Meta-Regression of the Duration Intensity on Hedges' g



Discussion

The purpose of this study was to quantitatively aggregate the findings on the identified studies of MALL for FL learners' speaking skill development. Meanwhile, moderator analyses on the learner-related, instruction-related and methodology-related categories were carried out to justify the variability in the aggregated effect, including a continuous variable (duration intensity) and eight categorial variables (proficiency levels, educational levels, language types, instructional approaches, intervention settings, software types, intervention durations and measured outcome types). In what follows, we will discuss the findings in relation to the two research questions.

Discussions of Overall Effects

RQ1 was associated with the overall effect size of MALL vs. non-MALL for FL learners' speaking skill development. A total of 20 effect sizes generated an aggregated effect size of 0.824. The meta-analysis result indicated a positive and significant effect for the use of MALL for FL learners' speaking skill development. The effect size of this study, however, is higher than those meta-analyses on MALL for domain-general language learning (Chen et al., 2020; Sung et al., 2015, 2016), but comparable with the results reported in a meta-analysis conducted by Lin and Lin (2019). 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; Sung et al., 2015, 2016) that dealt with MALL in general, this study meta-analyzed the use of domain-specific MALL for FL learners' speaking skill development. This explanation also lends support in another domain-specific meta-analysis (Lin & Lin, 2019), which also obtained a large effect size of 1.005 regarding mobile-assisted FL vocabulary learning.

Discussions of Moderating Effects

RQ2 considered the moderating effects of three categories of moderators. The homogeneity test for the effectiveness of MALL for FL learners' speaking skill development indicated a significant variability in the aggregated effect sizes, which justified a search for a series of moderator analyses that could help explain the variations in speaking outcomes between the MALL-related approach and the traditional approach. Eight categorial moderators and one continuous moderator were analyzed from the learner-related, instruction-related and methodology-related categories. Moderator analysis results indicated that instructional approaches and intervention durations were significant moderators, while proficiency levels, educational levels, language types, intervention settings, software types, measured outcome types and duration intensity did not find a significant moderating effect.

A scrutiny of the significant moderating effect of instructional approaches moderator indicated that such approaches as drill and practice ($g = 1.299$), flipped classroom ($g = 1.292$), multimedia learning ($g = 0.911$) and game-based approaches ($g = 0.807$) obtained high moderating effects, while peer feedback ($g = 0.528$) and task-based approach ($g = 0.434$) did not. The counterintuitive result is, however, inconsistent with many researchers who maintained that MALL technologies are often featured in task-based learning activities (e.g., Lan & Lin, 2016; Rezaee et al., 2019; Sun et al., 2017), and in providing comments to peers and receiving feedback from peers (Chien et al., 2020). This might be attributed to the overwhelming influence of the traditional "behaviorist, teacher-centered, transmission model of instruction" (Burston, 2014, p.344). In other words, the results hint that MALL technologies were often used outside of the classroom as a complement for formal classroom instruction instead, and the drill and practice features of MALL technologies were emphasized when developing FL learners' speaking ability (García Botero et al., 2019). A further implication is that, if learners fail to receive teachers' timely monitoring and feedbacks, their sustained motivation and interest in MALL might be undermined accordingly (García Botero et al., 2019; Li, 2021a; Li et al., 2019). Aside from the drill and practice approach, other approaches such as flipped classroom, multimedia learning and game-based learning should receive equal attention. For instance, using the multimedia learning approach, learners' verbal and nonverbal channels could be interconnected to reduce cognitive load and increase learning outcomes in turn (Li, 2021b).

Moderator analysis on the intervention durations moderator showed a result consistent with researchers who advocated that "long-term teaching interventions are important for obtaining reliable results" (Sung et al., 2016, p.264). Another plausible explanation for why long-term treatments have better moderating effects lies in the fact that FL learners' speaking skill, as the most challenging productive skill, cannot be substantially improved for a short period of time (Ahn & Lee, 2016; Hsu & Liu, 2021). In other words, FL learners' speaking skill could be improved when MALL for FL learners is applied for longer durations for speaking skill development, since it has been found to increase the frequency, interactivity and flow experiences of speaking practice with the provision of authentic, individualized, collaborative and learner-centered social contexts (Hsu & Liu, 2021; Li et al, 2021).

Regarding the three learner-related moderators, we neither find a significant moderating effect of proficiency levels moderator, nor of educational levels and language types moderators on the overall effect size, suggesting the effectiveness of MALL for FL learners' speaking skill development was stable regardless of the differences in learner-related factors, consistent with the findings of several researchers (Chen et al., 2020; Sung et al., 2015, 2016). For instance, meta-analyses evaluating the effectiveness of MALL applications conducted by Chen et al. (2020) and Sung et al. (2016) indicated that proficiency levels, educational levels and language types were not significant moderators for the effectiveness of MALL. Another possible explanation is that, the primary studies selected in this study are the experiments or quasi-experiments with strict manipulations, which might not be easily influenced by the learner-related confounds (e.g., proficiency levels, educational levels and language types, Li, 2022a).

Concerning the instructional settings moderator, it is not surprising that the effect size of outside classroom setting (g = 0.987) was higher than classroom setting (g = 0.698), which is consistent with similar studies conducted by Chen et al. (2020) and Sung et al. (2015) who posited learning with MALL applications in outside classroom setting achieved higher effect than in formal classrooms. This is because MALL applications can transcend temporal and spatial barriers to exert a maximal learning outcome by connecting the formal classroom and informal outside-of-the-classroom learning activities (Sung et al., 2015). The non-significant moderating effect of intervention settings, however, may warrant further research in this regard. Our result also indicated software for general purposes (g = 1.077) was higher than that for educational purposes (g = 0.655), which is contrary to Chen et al. (2020) findings that educational-purpose applications are better tailored to learners' needs. It should be noted here that applications for general purposes used in current study are instant messaging tools, e.g., mobile SNS, WhatsApp, LINE and telegram, that featured in peer interactions and collaborations (Li et al., 2021; Li, 2023c). Practicing FL speaking supported by the interactive and collaborative MALL applications would sustain learners' attention and motivate their interest, self-efficacy and engagement (e.g., Andujar & Salaberri-Ramiro, 2021; Chien et al., 2020; Ko & Lim, 2022). The non-significant result of duration intensity moderator is inconsistent with Xu and colleagues (2019), who found that low intensity tended to have larger moderating effect than high intensity in computer-assisted writing skill development. This may be due to the difference between mobile-assisted and computer-assisted learning tasks. However, considering the limited number of studies involved in the meta-regression analysis, the result should be interpreted with caution, warranting further research in this regard.

Implications for Practice

Implications for practice could also be inferred from the results. First, the use of MALL for FL learners' speaking skill development was confirmed to be more effective than traditional methods, indicating that the pedagogical potential of MALL for FL learners' speaking skill development should be encouraged for further investigations, especially for learners of the primary and secondary levels. Although it is extremely difficult to do any research that involves under 18-year-old samples for some reasons (e.g., personal intimacy protection, institutional and parental permission), the significant moderating effects of educational levels suggest that FL learners of all levels should receive equal and balanced attention with respect to the application of MALL for their speaking skill development. Second, the use of MALL for FL learners' speaking skill development should be integrated with effective instructional approaches, e.g., flipped classroom, multimedia learning and game-based learning. Further, learners should also be trained on self-monitoring and self-management to sustain their motivation and maximize the pedagogical effects of MALL for FL learners' speaking skill development (Hsu & Liu, 2021). Third, since longer durations achieved higher effect sizes, researchers should design long-term intervention durations in future studies to increase the reliability of the findings (Sung et al., 2016).

Conclusion and Limitations

This study reported results based on a meta-analysis of 20 effect sizes from 18 experimental and quasi-experimental studies published during 2000–2021. The results showed that the overall effect size was significantly large, suggesting the effectiveness of MALL for FL learners' speaking skill development over traditional methods. The moderator analysis results suggest that drill and practice, flipped classroom, multimedia learning and game-based approaches are effective instructional approaches, and long-term teaching interventions that have better pedagogical benefits should be taken. Regarding the non-significant moderators, pedagogical benefits can still be obtained regardless of the learner-related factors (e.g., proficiency levels, educational levels, and language types) and research design-related factors (e.g., intervention settings, software types, measured outcome types and duration intensity).

Despite the aforementioned findings, limitations should be addressed as well. First, due to the limit of keyword combinations and strict inclusion/exclusion criteria, some of the 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. Meanwhile, as our reviewer has critically commented, arguing that although the quality of selected studies has been confirmed for this meta-analysis, "the generic nature of MERSQI might result in critical language learning parameters to be overlooked, e.g., age, sex, L2 language competence, durations on task, etc." (Sung et al., 2019). Consequently, future studies should include greater selection of empirical studies which use appropriate experimental research design, and account for more keyword combinations and wider time range. Second, this study only considered speaking performance as the outcome. With more MALL for FL learners' speaking skill development studies being conducted that involved other affective outcomes, future research should consider investigating the effect of MALL for FL learners' speaking skill development on learners' affective outcomes (e.g., motivation, self-efficacy and flow experiences). Third, although some preliminary results were obtained from the moderator analysis, the moderating effects of other potential moderators should be further examined in the future meta-analytic research.

Acknowledgements

This research was supported by the Scientific Research Funds from Hunan Provincial Ministry of Education (grant number 23A0040). The author would like to thank the three anonymous reviewers, Professor Dorothy Chun and Professor Lara Lomicka Anderson for their constructive and insightful comments and suggestions.

References

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

Ahn, T. Y., & Lee, S. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778–786. https://doi.org/10.1111/bjet.12354

*Ali, A. Z. M., Segaran, K., & Hoe, T. W. (2015). Effects of verbal components in 3D talking-head on pronunciation learning among non-native speakers. *Educational Technology & Society*, *18*(2), 313–322. https://www.jstor.org/stable/jeductechsoci.18.2.313

Andujar, A., & Salaberri-Ramiro, M. S. (2021). Exploring chat-based communication in the EFL class: Computer and mobile environments. *Computer Assisted Language Learning*, *34*(4), 434–461. https://doi.org/10.1080/09588221.2019.1614632

*Ataeifar, F., Sadighi, F., Bagheri, M. S., Behjat, F., & Wang, S. (2019). Iranian female students' perceptions of the impact of mobile-assisted instruction on their English speaking skill. *Cogent Education*, *6*(1), 1–19. https://doi.org/10.1080/2331186X.2019.1662594

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, *14*, 395–411. https://doi.org/10.1002/sim.4780140406

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. (2013). Mobile-assisted language learning: A selected annotated bibliography of implementation studies 1994–2012. *Language Learning & Technology*, *17*(3), 157–225. http://hdl.handle.net/10125/44344

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

Burston, J., & Giannakou, K. (2022). MALL language learning outcomes: A comprehensive meta-analysis 1994–2019. *ReCALL*, *34*(2), 147–168. https://doi.org/10.1017/S0958344021000240

*Chang, M., & Lan, S. (2021). Flipping an EFL classroom with the LINE application: Students' performance and perceptions. *Journal of Computers in Education*, *8*(2), 267–287. https://doi.org/10.1007/s40692-020-00179-0

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

*Chew, S. W., Fang, W., Hsu, C., & Chen, N. (2018). Effects of reciprocal feedback on EFL learners' communication strategy use and oral communicative performance. *Smart Learning Environments*, *5*(11), 1–16. https://doi.org/10.1186/s40561-018-0061-2

*Chien, S., Hwang, G., & Jong, M. S. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English speaking performance and learning perceptions. *Computers & Education*, *146*, 1–20. https://doi.org/10.1016/j.compedu.2019.103751

Cho, K., Lee, S., Joo, M.-H. & Becker, B. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences*, *8*(3), 105–121. https://doi.org/10.3390/educsci8030105

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: The medical education research study quality instrument and the Newcastle-Ottawa scale-education. *Academic Medicine*, *90*(8), 1067–1076. https://doi.org/10.1097/ACM.0000000000000786

Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, *27*(2), 197–216. https://doi.org/10.1017/S0958344014000287

Elaish, M. M., Shuib, L., Hwang, G., Ghani, N. A., Yadegaridehkordi, E., & Zainuddin, S. Z. (2021). Mobile English language learning: A systematic review of group size, duration, and assessment methods. *Computer Assisted Language Learning*, 1–28. https://doi.org/10.1080/09588221.2021.1931341

Eutsler, L., Mitchell, C., Stamm, B. & Kogut, A. (2020). The influence of mobile technologies on preschool and elementary children's literacy achievement: A systematic review spanning 2007–2019. *Educational Technology Research and Development*, *68*(4), 1739–1768. https://doi.org/10.1007/s11423-020-09786-1

*Fang, W., Yeh, H., Luo, B., & Chen, N. (2021). Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. *ReCALL*, *33*(1), 71–87. https://doi.org/10.1017/S0958344020000208

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

*Grimshaw, J., & Cardoso, W. (2018). Activate space rats! Fluency development in a mobile game-assisted environment. *Language Learning & Technology*, *22*(3), 159–175. http://hdl.handle.net/10125/44662

Guanuche, A., Eiriz, O. & Espí, R. (2020). Corrective feedback through mobile apps for English learning: A review. In F. R. Narváez, D. F. Vallejo, P. A. Morillo, & J. R. Proaño, (Eds.), *Smart technologies, systems and applications: First international conference, SmartTech-IC 2019: Quito, Ecuador, December 2–4, 2019, Proceedings* (pp. 229–242). Springer. https://doi.org/10.1007/978-3-030-46785-2_19

Hou, Z., & Aryadoust, V. (2021). A review of the methodological quality of quantitative mobile-assisted language learning research. *System*, *100*, 1–15. https://doi.org/10.1016/j.system.2021.102568

Hsu, K., & Liu, G. (2021). A systematic review of mobile-assisted oral communication development from selected papers published between 2010 and 2019. *Interactive Learning Environments*, 1–18. https://doi.org/10.1080/10494820.2021.1943690

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/10.1080/10494820.2018.1486861

*Hwang, W., Huang, Y., Shadiev, R., Wu, S., & Chen, S. (2014). Effects of using mobile devices on English listening diversity and speaking for EFL elementary students. *Australasian Journal of Educational Technology*, *30*(5), 503–516. https://doi.org/10.14742/ajet.237

*Hwang, W., Shih, T. K., Ma, Z., Shadiev, R., & Chen, S. (2016). Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts. *Computer Assisted Language Learning*, *29*(4), 639–657. https://doi.org/10.1080/09588221.2015.1016438

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*(4), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Kamasak, R., Özbilgin, M., Atay, D. & Kar, A. (2020). The effectiveness of mobile-assisted language learning (MALL): A review of the extant literature. In A. S. Moura, P. Reis, & M. N. D. S. Cordeiro (Eds.), *Handbook of research on determining the reliability of online assessment and distance learning* (pp.194–212). IGI Global. https://doi.org/10.15642/NOBEL.2020.11.2.188-202

Klimova, B. & Zamborova, K. (2020). Use of mobile applications in developing reading comprehension in second language acquisition – A review study. *Education Sciences*, *10*, 391–411. https://doi.org/10.3390/educsci10120391

*Ko, E. G., & Lim, K. Y. (2022). Promoting English learning in secondary schools: Design-based research to develop a mobile application for collaborative learning. *The Asia-Pacific Education Researche*r, *31*, 307–319. https://doi.org/10.1007/s40299-021-00562-0

*Lan, Y., & Lin, Y. (2016). Mobile seamless technology enhanced CSL oral communication. *Educational Technology & Society*, *19*(3), 335–350. https://www.jstor.org/stable/jeductechsoci.19.3.335

Lee, S. M. (2022). A systematic review of context-aware technology use in foreign language learning. *Computer Assisted Language Learning*, *35*(3), 294–318. https://doi.org/10.1080/09588221.2019.1688836

Lee, Y. S., Sung, Y. T., Chang, K. E., Liu, T. C. & Chen, W. C. (2014). A meta-analysis of the effects of learning languages with mobile devices. In Y. W. Cao, & T. Väljataga, & J. K. T. Tang, H. Leung, & M. Laanpere (Eds.), *New horizons in web based learning* (pp.106–113), Springer. https://doi.org/10.1007/978-3-319-13296-9_12

Li, R. (2021a). Modeling the continuance intention to use Automated Writing Evaluation. *SAGE Open*, *11*(4), 1–13. https://doi.org/10.1177/21582440211060782

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. (2022a). Research trends of blended language learning: A bibliometric synthesis of SSCI-indexed journal articles during 2000–2019. *ReCALL*, *34*(3), 309–326. https://doi.org/10.1017/S0958344021000343

Li, R. (2022b). Foreign language reading anxiety and its correlates: A meta-analysis. *Reading and Writing*, *35*(4), 995–1018. https://doi.org/10.1007/s11145-021-10213-x

Li, R. (2022c). Effects of mobile assisted language learning on EFL/ESL reading comprehension. *Educational Technology & Society*, *25*(3), 15–29. https://www.jstor.org/stable/48673721

Li, R. (2022d). Effects of blended language learning on EFL learners' language performance: An activity theory approach. *Journal of Computer Assisted Learning*, *38*, 1273–1285. https://doi.org/10.1111/jcal.12697

Li, R. (2023a). Investigating effects of computer-mediated feedback on L2 vocabulary learning. *Computers & Education*, *198*, 1–11. https://doi.org/10.1016/j.compedu.2023.104763

Li, R. (2023b). Effects of mobile-assisted language learning on EFL learners' listening skill development. *Educational Technology & Society*, *26*(2), 36–49. https://doi.org/10.30191/ETS.202304_26(2).0003

Li, R. (2023c). Still a fallible tool? Revisiting effects of Automated Writing Evaluation from Activity Theory perspective. *British Journal of Educational Technology*, *54*(3), 773–789. https://doi.org/10.1111/bjet.13294

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

*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), 325–342. https://doi.org/10.1080/09588221.2017.1312463

*Lin, C., & Hwang, G. (2018). A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Educational Technology & Society*, *21*(2), 205–219. https://www.jstor.org/stable/26388398

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.

*Liu, T., & Chu, Y. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, *55*(2), 630–643. https://doi.org/10.1016/j.compedu.2010.02.023

Mahdi, H. (2017). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research*, *55*(3), 1–21. https://doi.org/10.1177/0735633117698826

*Mu, Z. (2017). Affordance-based mobile learning in a flipped classroom to enhance CFL learners' oral proficiency: A comparative study. *DEStech Transactions on Social Science, Education and Human Science*, *129*, 106–111. https://doi.org/10.12783/dtssehs/eemt2017/14417

Peng, H., Jager, S., & Lowie, W. (2021). Narrative review and meta-analysis of MALL research on L2 skills. *ReCALL*, *33*(3), 278–295. https://doi.org/10.1017/S0958344020000221

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA*, *298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

*Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2019). The impact of mobile-based dynamic assessment on improving EFL oral accuracy. *Education and Information Technologies*, *24*(5), 3091–3105. https://doi.org/10.1007/s10639-019-09915-1

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, *34*(2), 243–258. https://doi.org/10.1558/cj.27623

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

*Saran, M., Seferoglu, G., & Cagıltay, K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research*, *34*, 97–114. http://hdl.handle.net/20.500.12416/944

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

*Sun, Z., Lin, C., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning*, *30*(3-4), 304–324. https://doi.org/10.1080/09588221.2017.1308384

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

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., Lee, H., Yang, J., & Chang, K. (2019). The quality of experimental designs in mobile learning research: A systemic review and self-improvement tool. *Educational Research Review*, *28*, 1–21. https://doi.org/10.1016/j.edurev.2019.05.001

Valentine, J. C. (2019). Incorporating judgments about study quality into research syntheses. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.). *The handbook of research synthesis and meta-analysis (3rd ed.)* (pp.129–140). Russell Sage Foundation. https://doi.org/10.7758/9781610448864.10

Wang, C., Lan, Y., Tseng, W., Lin, Y. R., & Gupta, K. C. (2020). On the effects of 3D virtual worlds in language learning - a meta-analysis. *Computer Assisted Language Learning*, *33*(8), 891–915. https://doi.org/10.1080/09588221.2019.1598444

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 ed.)* (pp.153–172). Russell Sage Foundation. https://doi.org/10.7758/9781610448864.12

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, *51*(6), 1991–2005. https://doi.org/10.1111/bjet.13023

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

Appendix A. The Selected Studies on MALL for FL Speaking Development

Study	Technologies	Designs	General Findings
Ali et al. (2015)	3D Talking head with audio and text MALL	TE	The facilitative effects of three multimedia presentation strategies in 3D talking-head MALL on pronunciation skills were confirmed.
Ataeifar et al. (2019)	Twitter	TE	MALL was found to improve Iranian female EFL learners' speaking skill. The paper explored their perceptions of the learners' experience.
Chang & Lan (2021)	LINE	QE	LINE improved college learners' English performance and their perceptions of MALL.
Chew et al. (2018)	Speaking Yo with peer feedback	TE	Mobile-supported peer feedback improved learners' speaking performance rather than communication strategies.
Chien et al. (2020)	spherical video-based VR with peer feedback	TE	Mobile-supported peer assessment improved learners' English speaking, motivation, critical thinking skills, and reduced their English learning anxiety.
Fang et al. (2021)	mobile-supported TBLT	TE	Mobile-supported TBLT group outperformed the traditional TBLT group on the vocabulary and conversation comprehension tests.
Grimshaw & Cardoso (2018)	Spaceteam ESL	TE	Mobile game-assisted environment improved learners' oral fluency and positively influence anxiety and WTC.
Hwang et al. (2014)	mobile speaking system	QE	MALL improved EFL elementary learners' speaking performance.
Hwang et al. (2016)	mobile system	TE	Game-based activities can improve learners' speaking skills if driven by a MALL system.
Ko & Lim (2022)	WikiTalki	TE	MALL recording and sharing function increased learner participation in large-sized English classes; peer feedback promoted collaborative learning; continuous refinement of the application design enhanced the quality of instructional design and usability.
Lan & Lin (2016)	MOSE	TE	No significant difference found in learners' speaking performance between traditional classroom learning and the mobile learning.
Liakin et al. (2017)	TTS	TE	The pedagogical use of mobile TTS technology can complement and enhance the teaching of L2 pronunciation.
Lin & Hwang (2018)	Facebook	TE	MALL improved EFL learners' English oral presentation skill.
Liu & Chu (2010)	Handheld English Language Learning Organization (HELLO)	TE	Incorporating ubiquitous games into the English learning process could achieve a better learning outcomes and motivation than using non-gaming method.
Mu (2017)	WeChat	TE	MALL improved Chinese EFL learners' speaking performance.
Rezaee et al. (2019)	WhatsApp	TE	MALL improved EFL learners' speaking accuracy.
Saran et al. (2009)	MMS	TE	Using mobile phones had positive effects on students' pronunciation learning.
Sun et al. (2017)	mobile SNS	TE	Mobile social networking the English-speaking skills of young learners.

Notes. 3D = 3 dimensional; VR = virtual reality; TBLT = task-based language teaching; WTC = willingness to communicate; TTS = text to speech; MMS = Multimedia Messaging Service; SNS = Social Network Software; QE = quasi-experiment; TE = true experiment; EFL = English as a foreign language; L2 = second language.

Appendix B. Related Reviews and Meta-Analyses on MALL

Types	Study	*k*	Timespan	Skills	General Findings
Qualitative reviews	Burston (2013)	345	1994–2012	Language	MALL was involved in L1 and L2/FL with mobile technologies.
	Burston (2014)	NA	NA	Language	The pedagogical challenges of MALL included constrained access to internet and cost of mobile technologies.
	Burston (2015)	315	1994–2012	Language	MALL studies were reviewed on durations, sample sizes, design and learning outcomes.
	Duman et al. (2015)	69	2000–2012	Language	MALL technologies were frequently used in vocabulary learning.
	Elaish et al. (2021)	151	2010–2017	Language	MALL studies were reviewed on group size, duration and methods.
	Eutsler et al. (2020)	61	2007–2019	Language	MALL technologies within preschool and elementary children's literacy domains were reviewed.
	Guanuche et al. (2020)	39	2010–2019	Language	The correct feedback of MALL technologies contributed to English learning process.
	Hou & Aryadoust (2021)	174	2007–2020	Language	A review of methodological quality of quantitative MALL research was conducted.
	Hsu & Liu (2021)	28	2010–2019	Speaking	MALL-related speaking studies were reviewed on presage, context, learner, process and product variables.
	Hwang & Fu (2019)	93	2007–2016	Language	MALL studies were reviewed on research methods, research issues, language and learner types, and learning outcomes.
	Kamasak et al. (2020)	13	2010–2020	Language	It posited the complex nature of language learning and cautioned about the promise and risks of using MALL.
	Lee (2022)	88	2000–2017	Language	Context-aware MALL technologies were reviewed on publication year, methodology, participants, languages, device types and the effectiveness.
	Klimova & Zamborova (2020)	9	2018–2020	Reading	MALL-related reading studies were reviewed on objectives, intervention durations, measures and main findings.
	Rosell-Aguilar (2017)	NA	NA	Language	The effective design and evaluation of MALL technologies were presented.

	Shadiev et al. (2020)	22	2009–2018	Language	Pedagogical approaches, data collection, locations, affordances of authentic environments and issues in MALL research were reviewed.
Meta-analyses	Burston & Giannakou (2022)	84	1994–2019	Language	Main effect was 0.72 for between-group (independent, experimental) and 1.16 for within-group (quasi-experimental) treatment studies; four critical moderators (language learner competence level, language area focus, institutional environment, and intervention duration) revealed similarly large effect sizes.
	Chen et al. (2020)	84	2008–2018	Language	Main effect was 0.722 for the overall effectiveness; among nine moderators, target language skill, target language and L1/L2 were found to be significant moderators.
	Cho et al. (2018)	20	2005–2017	Language	Main effect was 0.51 for the overall effectiveness; among six moderators, test type and source of the study were found to be significant moderators.
	Peng et al. (2021)	17	2008–2017	Language	Main effect was 0.95 for the overall effectiveness; three moderators (type of activities, modality of delivery, and duration of treatment) were reported.
	Lee et al. (2014)	44	1993–2013	Language	Main effect was 0.54 for the overall effectiveness; six moderators (school level, source, context, language skills, test types, and learner types) were reported.
	Mahdi (2017)	16	2008–2017	Vocabulary	Main effect was 0.67 for the overall effectiveness; moderator analyses of learners' age, type of vocabulary learning (receptive or productive), and aspects of vocabulary used (form, meaning, and use) were also reported.
	Sung et al. (2015)	44	1993–2013	Language	Main effect was 0.55 for the overall effectiveness; moderator analyses of learning stages, hardware use, software used, intervention settings, teaching methods, intervention durations, learning skills, target languages, and L1/L2 were also reported.

	Sung et al. (2016)	110	1993–2013	Language	Main effect was 0.523 for the overall effectiveness; moderator analyses of learning stages, hardware use, software used, intervention settings, teaching methods, intervention durations, and domain subjects were also reported.

Notes. *k* = number of studies included; NA = not available; L1 = first language; FL = foreign language; L2 = second language.

Appendix C. MERSQI items and rating criteria

Domain	MERSQI Items	Rating criteria
Study design	Single group cross-sectional or single group posttest only	1
	Single group pretest and posttest	1.5
	Nonrandomized, 2 group	2
	Randomized controlled trial	3
Sampling	No. of institutions studied: 1	0.5
	No. of institutions studied: 2	1
	No. of institutions studied: >2	1.5
	Response rate: <50% or not reported	0.5
	Response rate: 50-74%	1
	Response rate: >75%	1.5
Type of data	Subjective assessment by participant	1
	Objective measurement	3
Validity of instrument	Internal structure: Not reported	0
	Internal structure: Reported	1
	Content: Not reported	0
	Content: Reported	1
	Relationships to other variables: Not reported	0
	Relationships to other variables: Reported	1
Data analysis	Data analysis inappropriate for study design or type of data	0
	Data analysis appropriate for study design or type of data	1
	Complexity of analysis: Descriptive analysis only	1
	Complexity of analysis: Beyond descriptive analysis	2
Outcomes	Satisfaction, attitudes, perceptions, opinions, general facts	1
	Knowledge, skills	1.5
	Behaviors	2
	Mixed outcomes	3

Appendix D. The Proposed Coding Scheme of Moderators

Moderators	Coding types	Subtypes	Operational definitions	References
Learner-related	Proficiency levels	1. low	Studies that reported learners as foreign language beginners.	Li (2022b)
		2. intermediate	Studies that reported learners as intermediate learners.	
		3. high	Studies that reported learners as highly proficient learners.	
	Educational levels	1. primary education	Kindergarten or primary school students.	Li (2022c)
		2. secondary education	Junior middle school or senior high school students.	
		3. tertiary education	College students.	
	Language types	1. FL	Studies that reported learning a foreign language.	Li (2022b)
		2. L2	Studies that reported learning a second language.	
Instruction-related	Instructional approaches	1. drill and practice	Studies that used MALL devices to practice FL speaking ability.	Wang et al. (2020)
		2. flipped classroom	Studies that used flipped classroom approach.	
		3. game-based	Studies that used game-based approach.	
		4. multimedia learning	Studies that involved different multimedia information for speaking skill development.	
		5. peer feedback	Studies that involved peer assessment or interactions.	
		6. task-based	Studies that involved task-based learning activities.	
	Intervention settings	1. classroom	Teaching activities that occurred in the classroom.	Chen et al. (2020)
		2. outside classroom	Teaching activities that occurred outside of the classroom.	
Methodology-related	Software types	1. general purposes	Applications that were NOT designed for educational purposes, e.g., WeChat, WhatsApp, telegram and LINE.	Chen et al. (2020)
		2. educational purposes	Applications that were designed for educational purposes, e.g., WikiTalki, mobile speaking system and spherical virtual reality.	
	Intervention durations	1. one session, ≤1 week	Durations fewer than one week or only one session.	Chen et al. (2020)
		2. >1 week, ≤4	Durations over one week, but fewer than	

		weeks	four weeks.	
		3. >4 weeks, ≤one semester	Durations over four weeks, but within one semester.	
	Duration intensity	Min/week	Studies that reported the number of minutes per week.	Xu et al. (2019)
	Measured outcome types	1.standardized	Standardized TOEIC/IELTS/OPT speaking comprehension test scores.	Xu et al. (2019)
		2.researcher-designed	Speaking comprehension test scores made by researchers.	

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

About the Author

Rui Li is a full professor, also serving as a doctoral and post-doctoral supervisor at School of Foreign Languages, Hunan University. His research interests include CALL, applied linguistics and psycholinguistics. He published articles in top-tier SSCI-indexed journals, such as *Computers & Education*, *Computer Assisted Language Learning*, *Language Learning & Technology*, *ReCALL*, *British Journal of Educational Technology*, *Journal of Computer Assisted Learning*, *Interactive Learning Environments*, *Journal of Educational Computing Research*, *Journal of Computing in Higher Education*, *Educational Technology & Society*, *Reading and Writing*, *Journal of Psycholinguistic Research*, *Porta Linguarum*, *RELC Journal*, among others. He serves as an editorial board member of *Educational Technology & Society* or anonymous reviewer for over 30 SSCI-indexed journals in CALL, applied linguistics and psycholinguistics.

E-mail: liruidianzi@hotmail.com

ORCiD: https://orcid.org/0000-0001-9764-9954