

Effects of Mobile-Assisted Language Learning on EFL Learners' Listening Skill Development

Rui Li

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

(Submitted March 24, 2022; Revised August 16, 2022; Accepted August 18, 2022)

ABSTRACT: Although an increasing number of studies have focused on the use of mobile-assisted language learning (MALL) for English as a foreign language (EFL) learners' listening skill development, there is a lack of comprehensive meta-analysis regarding the effect sizes of these studies. To fill the gap, 20 selected experimental studies involving 1218 participants were included for a meta-analysis based on the proposed inclusion and exclusion criteria. Results showed that the overall effect size was moderate-to-large, $g = 0.792$, 95% CI [0.536, 1.047], suggesting that MALL for EFL learners' listening skill development is more effective than traditional methods. Regarding moderators for the overall effect, different moderator effects of educational levels, software types, control conditions, intervention settings, measured outcome types and intervention durations were reported. Specifically, educational levels were found to be a significant moderator, while software types, control conditions, intervention settings, measured outcome types and intervention durations were not the significant moderators. The implications for practice were discussed as well.

Keywords: English as a foreign language (EFL), Evidence-based applied linguistics (EBAL), Listening skill, Meta-analysis, Mobile-assisted language learning (MALL)

1. Introduction

The development of English as a foreign/second language (EFL/ESL) listening skill “is seen not only as something valuable for its own sake but as something that supports the growth of other aspects of language use, such as speaking and reading” (Richards, 2005, p. 85). Currently, there have been a positive change in the teaching of EFL listening comprehension, but learners still confront with such difficulties as insufficient classroom instructions and a lack of sustained listening practice to guide them to be successful listeners (Hwang et al., 2016; Vandergrift & Goh, 2012). In recent years, the popularity of mobile-assisted language learning (MALL) technologies has reshaped the traditional FL listening instructions, since “listening comprehension skill is an invisible mental process that is more difficult to develop” (Mendoza et al., 2020, p. 61), and compared to other language skills MALL is mostly used to autonomously and ubiquitously develop learners' listening skill with sufficient exposure of multimodal listening materials (Li, 2022a; Li, 2022b). Motivated by the possibility of overcoming the barriers of insufficient classroom instructions and the poverty of listening practice, pedagogical potentials of MALL for EFL learners' listening skill development have been well-recognized to create ubiquitous learning environments.

While researchers (Alabsi, 2020; Liu et al., 2018; Saedakhtar et al., 2021; Tai & Chen, 2021) have claimed that MALL can effectively improve learners' listening skill, a quantitative meta-analysis of the overall effect size among these studies is still in paucity. Drawing on evidence-based applied linguistics (EBAL), the effects of a pedagogical intervention should be supported with sound evidence available in foreign language education. In other words, by combining the results detailed in multiple studies with larger sample sizes (Li, 2022a), a meta-analysis from the drive of evidenced-based practice will afford more accurate estimates of the effects of MALL for EFL learners' listening skill development, and offer a deeper understanding of related moderators that potentially affect the overall effects. More specifically, this study aims to (a) generalize empirical findings of MALL for listening skill development research while (b) tackling the variability of the aggregated effects with a meta-analytic approach.

2. Literature review

2.1. Review of MALL for EFL listening skill development

With the rapid development of mobile technologies, such as smartphones, tablet PCs and other portable devices, pedagogical affordances of MALL have received considerable attention in the existing studies (e.g., Holden & Sykes, 2011; Lai et al., 2022; Thorne et al., 2021; Zheng et al., 2017; Zheng et al., 2018). Researchers have

maintained that MALL supports seamless or ubiquitous learning (Burston & Giannakou, 2022; Chen et al., 2020), overcomes a lack of individualized learning (Sung et al., 2016), and affords instant and timely feedback (Sung et al., 2015), etc. For years, while MALL has been documented to facilitate EFL learners' four main language skills (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), researchers have gradually realized the importance of using MALL to develop EFL learners' listening skill. More specifically, since listening comprehension has been considered the most difficult of the four main language skills (Siegel, 2014), the ubiquitous feature of MALL could afford learners with increased practice and multimodal materials. Furthermore, the individualized and interactive features of MALL could encourage learners to overcome the limitation of traditional classroom instructions (Li, 2021a).

MALL for EFL listening skill development is defined as the use of MALL devices, e.g., PDAs (Chang et al., 2018), MP3 (Rahimi & Soleymani, 2015), iPods (de la Fuente, 2014), mobile phones (Alabsi, 2020; Al-Shamsi et al., 2020) and mobile virtual reality (VR) gear (Tai & Chen, 2021), among others, to facilitate EFL learners' listening skill development. Researchers have conducted extensive experimental or quasi-experimental studies to compare the experimental group using MALL for EFL listening, such as mobile captions and subtitles (Alabsi, 2020), micro dialogues (de la Fuente, 2014), mobile English listening system (Liu et al., 2018), mobile VR (Tai & Chen, 2021), and portable media players (Rashtchi & Mazraehno, 2019) and the control group using traditional approaches, e.g., conventional paper-and-pencil (Azar & Nasiri, 2014), DVDs and audio CDs from the book (Rashtchi & Mazraehno, 2019) and indoor computer classroom (Chang et al., 2018).

While an emerging array of empirical studies on MALL for EFL listening skill development, researchers have resulted in mixed findings. Some have found that MALL can effectively improve learners' listening comprehension skill (Alabsi, 2020; Liu et al., 2018; Saeedakhtar et al., 2021; Tai & Chen, 2021). Despite the promising results, not everyone could be convinced, as other researchers (e.g., Hwang et al., 2016; Rashtchi & Mazraehno, 2019; Seo & Choi, 2014) have obtained the limited effects of MALL for EFL listening skill development. Besides the aforementioned empirical studies, researchers have also conducted reviews in an attempt to gain a holistic understanding of the current state. To our knowledge, only Coşkun and Marlowe (2020) began to adopt a narrative review and investigate the use of MALL for EFL listening skill development. From their review, they summarized that most studies reported the facilitative effects.

Taken together, although the aforementioned studies are valuable to shed light on MALL for EFL listening research, several issues remain open for debate. On the one hand, given the growing number of empirical studies that have progressively accumulated sufficient data sources, the mixed results may justify the need to scrutinize the pedagogical effects of MALL on EFL learners' listening skill development and which moderators may underline such discrepancy. Compared to those empirical studies, meta-analytic results based on multiple studies and increased sample sizes are more reliable and generalizable (Li, 2022d; Li, in press). On the other hand, besides the empirical studies, while the qualitative review (Coşkun & Marlowe, 2020) might provide a comprehensive vision, little is still known about the effect size with a quantitative meta-analytic approach, and how the effect of MALL for listening skill development was affected by some possible moderators. Given that understanding the overall effect and moderator effect could not only index the effectiveness of an intervention, but also inform pedagogy, it is necessary to conduct a meta-analysis that examines the effectiveness and potential moderators on MALL for EFL learners' listening skill development.

2.2. Moderators of MALL for EFL listening skill development

The meta-analysis regarding the effectiveness of MALL for listening skill development is not a simple black-and-white issue, rather rigorous analysis of potential factors that can moderate the effectiveness is also required. The selection of moderators was based on the existing literature and widely referred to previous studies (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), resulting in the following six moderators: educational levels, software types, control conditions, intervention settings, intervention durations and measured outcome types.

Educational levels. Participants of different educational levels tend to have different learning outcomes. For instance, Costabile and colleagues (2003) employed 9 to 10-year-old children as participants of primary educational level to evaluate the effects of a tutoring hypermedia and found children's affective perceptions predict their learning outcomes. By contrast, in a quasi-experiment, Li (2021a) recruited 17 to 21-year-old college EFL learners to understand the effects of game-based vocabulary learning on their vocabulary achievement, motivation and self-confidence. Results indicated that those participants of tertiary educational level might not care too much about the entertainment feature of games that tend to arouse their affective perceptions, implying learner educational level might be an important moderator. In this study, we took

educational levels (viz. primary, secondary and tertiary education) as a moderator to test whether there exists moderator effect of educational levels on the MALL for EFL learners' listening skill development.

Software types. Depending on the specific educational purposes of MALL technologies, the software could be further categorized into general purposes (technologies that were designed for non-educational purposes, e.g., MP3 and iPods) and educational purposes (technologies that were designed for educational purposes, e.g., listening management system). It is generally believed that domain-specific MALL technologies developed for educational purposes are better tailored to EFL learners' individualized needs compared to those of general purposes (Chen et al., 2020; Sung et al., 2016). However, it remains largely unclear whether the research findings of the domain-general MALL technologies could be generalizable to the domain-specific type (Li, 2022c), viz. MALL for listening skill development in particular.

Control conditions. While participants of the experimental group who adopted MALL technologies have been extensively described, those of the control group should be considered with caution as well. Thus, two main control conditions are classified: paper-and-pencil (participants of the control group who used traditional paper-and-pencil method to practice listening) and computer-based methods (participants of the control group who used traditional computer-based method, e.g., watching videos, DVDs and CDs, to practice listening). Importantly, as all effect sizes are potentially caused by a comparison with a control condition, it is very important to understand what the control conditions look like in the studies reviewed. As such, the moderator effect of control conditions deserves further scrutiny in this study.

Intervention settings. Since EFL learning or teaching activities might occur in the classroom or outside of the classroom, intervention settings of MALL for listening skill development are classified into two kinds: classroom and outdoor (Sung et al., 2015). The moderator effect of intervention settings receives considerable attention in the existing MALL studies (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016). While these studies found the stronger effect of learning in outdoor settings than classroom settings, it is still unclear whether similar results could be applicable to the domain-specific EFL listening comprehension. In this study, we examine the moderator effect of intervention settings on the effectiveness of MALL for EFL listening skill development.

Intervention durations. The consensus regarding the moderator effect of intervention durations has not yet been reached thus far. For instance, in a systematic review of MALL research, Hwang and Fu (2019) argued that longer treatment durations can fully represent the real effects of MALL on learning outcomes. However, in a recent meta-analysis of MALL, Chen and colleagues (2020) claimed that shorter-term interventions yielded larger effect sizes than longer-term ones, because EFL learners' curiosity might not be sustain for a long period of time. As such, we examine the moderator effect of intervention durations in this study.

Measured outcome types. According to Xu and colleagues (2019), the heterogeneity of measured outcomes might impact the results of an intervention. In this study, we categorize two measured outcome types, viz. standardized vs. researcher-designed, to understand the moderator effect of measured outcome types.

2.3. Research statements and questions

This study is designed to provide a brand-new quantitative perspective on MALL for EFL listening skill development, using a meta-analysis of rigorously peer-reviewed empirical research. The research purposes are twofold: First, drawing on the data collected from the primary studies, we conduct a meta-analysis to calculate the aggregated overall effect regarding MALL for EFL listening skill development. Second, according to similar meta-analyses (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), the moderator effects, such as educational levels, software types, control conditions, intervention settings, intervention durations and measured outcome types, are analyzed as well. Consequently, two research questions to be addressed are as follows.

Research question 1: What is the overall effect size of MALL vs. non-MALL for EFL learners' listening skill development?

Research question 2: How do moderators, such as educational levels, software types, control conditions, intervention settings, intervention durations and measured outcome types, affect the aggregated effect on listening skill development?

3. Research design

This study aimed to meta-analyze the experimental or quasi-experimental studies on MALL for EFL listening skill development. Specifically, it firstly dealt with the overall effect of MALL for EFL listening; and secondly it calculated the moderator effects of the related moderators, such as educational levels, software types, control conditions, intervention settings, intervention durations and measured outcome types, on the overall effect size. As a result, practical implications could be obtained from the meta-analytic findings.

3.1. Data collection source

To exhaustively retrieve the related primary studies on MALL for EFL listening skill development, data collection procedures were observed as follows: First, data were searched from several electronic online databases (Web of Science, ProQuest, Scopus and ERIC), publishers (e.g., ScienceDirect, Springer, SAGE, Taylor & Francis and Wiley) and search engines (Google Scholar and Baidu Scholar) by using a combination of the following MALL-related and listening-related keywords integrated with Boolean operators, i.e., (mobile-assisted language learning OR MALL OR mobile applications OR portable 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 iPod OR iPhone OR tablets OR MP3 OR personal digital assistants OR PDAs OR podcasts) AND (listen OR listening comprehension OR listening skill OR listening ability OR listening performance OR listening score OR listening test). Then, to further avoid the insufficient search of a significant portion of the relevant literature in the first-round, a second-round backward and forward citation search was conducted with snowballing technique (Biernacki & Waldorf, 1981) by scanning references in some review articles (e.g., Chen et al., 2020; Coşkun & Marlowe, 2020; Jia & Hew, 2021). Third, each of the following EdTech journals (*Computers & Education*, *Internet and Higher Education*, *British Journal of Educational Technology*, *Computers in Human Behavior*, *Interactive Learning Environments*, *Educational Technology Research and Development*, *Educational Technology & Society*, *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) and language education journals (*Computer Assisted Language Learning*, *Language Learning & Technology*, *ReCALL*, *System* and *CALICO Journal*) was manually searched to avoid the incomplete inclusion. The initial retrieval resulted in 74 primary studies after duplicates removed.

3.2. Inclusion and exclusion criteria

To exclude irrelevant literature, two researchers independently and manually narrowed down the search to only cover the articles related to MALL for EFL listening skill development based on the inclusion and exclusion criteria:

(1) *Timespan constraints*. Given that studies of MALL for EFL listening skill development remained few in number before 2010 (Hwang & Fu, 2019), the studies published in English during 2010 to 2021 were considered.

(2) *Topic and publication type constraints*. The studies should adopt a form of MALL devices/tools (e.g., mobile phones, PDAs and iPods) on EFL/ESL listening skill development. Those studies that failed to use educational technologies, or used CALL technologies (e.g., computers and projectors), used MALL technologies on first language or other FL listening skill development were excluded. The studies should be rigorously peer-reviewed publications, including journal articles, book chapters and conference proceedings. Those non-peer-reviewed publications were excluded. Furthermore, other review publications (e.g., review articles, book reviews, and editorial materials) were excluded. As a result, those irrelevant ($k = 31$) were excluded, resulting in $k = 43$ left for further analysis.

(3) *Material constraints*. For the data transformation or effect size calculation, only the experimental or quasi-experimental studies that reported means, *SD*, and number of the participants were included. More specifically, the independent variables should include different interventional modes (e.g., traditional learning method as a control group vs. MALL method as a treatment group), and the dependent variable should include a measure of the researcher-designed or standardized pre- and post-tests on EFL listening comprehension performance between different modes. Among them, $k = 25$ studies were excluded for the following reasons: eight publications were about qualitative research on listening pedagogical or theoretical recommendations, five on

EFL learners' attitudes or perceptions and twelve without sufficient information for effect size calculation. As a result, 20 eligible publications (18 articles and 2 conference proceedings) were finalized.

3.3. Study quality evaluation

Study quality is one of the most important issues to consider, as it “can affect study results, which can in turn affect the conclusions drawn. It is thus necessary to develop an explicit, transparent, and reproducible instrument of assessing study quality” (Valentine, 2019, p. 130). Informed by Valentine (2019), the commonly used study quality instrument, viz. the Medical Education Research Study Quality Instrument (MERSQI) (Reed et al., 2007), was adopted to evaluate the methodological quality of the selected studies. 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 evaluating the study quality of educational research (e.g., Jensen & Konradsen, 2018). 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 potential range 5 to 18. The average total score of 11 or so could be taken as the benchmark of satisfied study quality. For instance, Jensen and Konradsen (2018), in their meta-analysis, obtained the average total score of 10.9, with a range of 6 to 14.5. Likewise, Cook and Reed (2015) achieved the average score of 11.3, with a range of 8.9 to 15.1. Our meta-analysis showed that the average total score is 14.684 ($SD = 0.682$), with a range of 13.5 to 15, suggesting that the selected primary studies had the high quality for the coming meta-analysis.

3.4. Coding scheme

Drawing on the existing meta-analytic studies on MALL (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016; Xu et al., 2019), all the 20 selected publications were coded with the following major categories (Table 1).

After the code scheme was developed, the following coding procedures were observed. First, given the recommended practice for data dependencies (Plonsky & Oswald, 2014), multiple effect sizes reported in a single publication involved different participants or different treatment interventions were coded separately to ensure the reliability of the analyses, resulting in 20 eligible primary studies that yielded 21 effect sizes as experimental studies. Second, two coders negotiated with each other to ensure the consistent understanding of each coding items. The discrepancies were resolved by consensus through discussions, along with any necessary reviews of the existing coding scheme.

Table 1. The descriptive information of coding scheme

Coding types	Subtypes	Operational definitions	References
Educational levels	• primary education	Kindergarten or primary school students.	Chen et al. (2020)
	• secondary education	Junior middle school or senior high school students.	
	• tertiary education	College students.	
Control conditions	• paper-and-pencil	Participants of the control group used traditional paper-and-pencil method to practice listening.	Researcher-designed
	• computer-based	Participants of the control group used traditional computer-based method (e.g., watching videos, DVDs and CDs) to practice listening.	
Intervention settings	• classroom	Teaching/learning activities occurred in the classroom.	Chen et al. (2020)
	• outdoor	Teaching/learning activities occurred outside of the classroom.	
Software types	• general purposes	Apps that were NOT designed for educational purposes, e.g., WhatsApp, captions and subtitles, podcasts and portable media players.	Chen et al. (2020)
	• educational purposes	Apps that were designed for educational purposes, e.g., mobile English listening system and mobile VR-assisted listening system.	
Measured outcome types	• standardized	Standardized TOEIC/IELTS/OPT listening comprehension scores.	Xu et al. (2019)

	<ul style="list-style-type: none"> • researcher-designed 	Listening comprehension tests made by researchers.	
Intervention durations	<ul style="list-style-type: none"> • short (one session, ≤ 1 week) 	Durations fewer than one week or only one session.	Chen et al. (2020)
	<ul style="list-style-type: none"> • intermediate (> 1 week, ≤ 4 weeks) 	Durations over one week, but fewer than four weeks.	
	<ul style="list-style-type: none"> • long (> 4 weeks, ≤ one semester) 	Durations over four weeks, but within one semester.	

Note. Apps = Applications; VR = virtual reality; TOEIC = Test of English for International Communication; IELTS = International English Language Testing System; OPT = Oxford Placement Test.

3.5. Calculation and outlier diagnosis of the effect sizes

Hedges' g that could "provide a simple correct for the bias of small sample size" was taken as the effect size index (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, to calculate effect sizes. For instance, we used means, sample sizes and p -values for both the experimental and control group (Hsu et al., 2013), and means, sample sizes and t -values (Hwang et al., 2014; Liu et al., 2018), to compute the effect sizes. Furthermore, a between-study Q test was executed to ensure whether the moderators played a role in the between-study heterogeneity (Li, 2022c).

Furthermore, according to Lipsey and Wilson (2001, p.108), potential outliers with the extreme effect sizes that were "more than 3 standard deviations from the mean of all the effect sizes" should be excluded. As such, one selected research ($g = 3.926$, Saeedakhtar et al., 2021) out of the 20 eligible studies that yielded extremely large effect sizes was excluded, resulting in 19 remaining primary studies that yielded 20 effect sizes for the final analysis.

3.6. Publication bias analysis

The publication bias refers to the phenomenon that nonsignificant results are unlikely to be published, and the unpublished studies might differ from the published studies (Borenstein et al., 2005; Borenstein et al., 2009). According to some existing studies (Borenstein et al., 2005; Borenstein et al., 2009; Lipsey & Wilson, 2001), the inspection of publication bias was based on a fail-safe N method. In other words, by evaluating how many unpublished studies with nonsignificant results would change the meta-analytic results from significant to nonsignificant, Rosenthal (1991) proposed a classical fail-safe N (N_{fs}) method to avoid the file-drawer problem. The calculation is as follows:

$$N_{fs} = \left(\frac{N_0}{Z_c^2}\right)(N_0\bar{Z}_0^2 - Z_c^2)$$

where N_0 is the number of studies, Z_c is the critical value of Z , and \bar{Z}_0 is the mean of Z obtained for the N_0 studies. It indicated that there existed no publication bias, since the result of fail-safe N was 896, which was significantly higher than the respective observed number 20 ($Z = 13.259$, $p < .001$), that Rosenthal (1991) suggested for the file-drawer problem.

4. Results

4.1. Overall analysis results

Twenty experimental studies involving 1218 participants were selected for the meta-analysis. The overall effect size was estimated using a *random effect model*, which "assumes that each observed effect size differs from the sampling error plus a value that represents other sources of variability" (Lipsey & Wilson, 2001, p. 119). As shown in Table 2, the overall effect size computed from 20 effect sizes is moderate-to-large, $g = 0.792$, 95% CI [0.536, 1.047] and significant, $Z(19) = 6.072$, $p < .001$, indicating the use of MALL for EFL learners' listening skill development is more effective than traditional methods.

Table 2. Overall effect size results of the included studies

<i>k</i>	<i>g</i>	95% CI		Z-value	Heterogeneity	
		LL	UL		<i>Q</i> -value	<i>I</i> ²
20	0.792	0.536	1.047	6.072	93.318***	79.640

Note. *k* = number of effect sizes; *g* = Hedges' *g*; LL = lower limit, UL = upper limit; 95 % CI = 95% confidence interval; ****p* < .001.

4.2. Moderator analysis results

As apparent in Table 2, *Q*-value was 93.318 with *p* < 0.001, indicating that there were between-group differences among the effect sizes resulting from factors other than subject-level sampling error. The *I*² for the overall model showed high heterogeneity (*I*² = 79.640), indicating the need for moderator analysis (Borenstein et al., 2005; Borenstein et al., 2009). To achieve this aim, moderator analysis results were presented in Table 3.

Table 3. Moderator analysis results

Moderators	<i>k</i>	<i>g</i>	95% CI		Z-value	Heterogeneity		
			LL	UL		<i>Q</i> -value	<i>df</i>	<i>p</i>
<i>Educational levels</i>	20	0.590	0.418	0.762	6.717***	6.386*	2	0.041
1. primary education	4	0.485	0.247	0.724	3.990***	primary vs. secondary, <i>p</i> = 0.748		
2. secondary education	3	0.415	0.059	0.772	2.282*	secondary vs. tertiary, <i>p</i> = 0.028		
3. tertiary education	13	0.974	0.628	1.320	5.512***	tertiary vs. primary, <i>p</i> = 0.023		
<i>Control conditions</i>	20	0.802	0.547	1.056	6.184***	1.210	1	0.271
1. paper-and-pencil	10	0.645	0.269	1.022	3.357**			
2. computer-based	10	0.932	0.588	1.276	5.309***			
<i>Intervention settings</i>	20	0.793	0.532	1.054	5.961***	0.051	1	0.821
1. classroom	11	0.821	0.469	1.172	4.576***			
2. outdoor	9	0.760	0.371	1.149	3.826***			
<i>Software types</i>	20	0.724	0.513	0.935	6.736***	0.599	1	0.439
1. general purposes	8	0.666	0.408	0.923	5.072***			
2. educational purposes	12	0.843	0.476	1.210	4.500***			
<i>Measured outcome types</i>	20	0.765	0.534	0.996	6.484***	0.056	1	0.813
1. standardized	11	0.805	0.401	1.208	3.907***			
2. researcher-designed	9	0.745	0.463	1.027	5.180***			
<i>Intervention durations</i>	20	0.771	0.517	1.025	5.952***	0.341	2	0.843
1. short	2	1.017	-0.158	2.193	1.696			
2. intermediate	6	0.709	0.033	0.355	3.921***			
3. long	12	0.817	0.195	0.038	4.185***			

Note. *k* = number of effect sizes; *g* = Hedges' *g*; LL = lower limit, UL = upper limit; 95 % CI = 95% confidence interval; to further locate the sources of variation, post-hoc comparisons were reported for the significant heterogeneity; **p* < .05; ***p* < .01; ****p* < .001.

Educational level. As shown in Table 3, the effect size was largest for college students of tertiary education (*g* = 0.974, *p* < .001), followed by the moderate effect size of primary (*g* = 0.485, *p* < .001) and secondary (*g* = 0.415, *p* < .050) educational levels. Between-group comparisons reached a statistical significance (*Q* = 6.386, *p* = .041), indicating that the effect sizes of three categories differ significantly from each other. Post-hoc comparisons were further executed. College students of tertiary education who used MALL for EFL listening skill development were found to have higher beneficial effects than those of primary (*Q* = 5.188, *p* = .023) and secondary (*Q* = 6.386, *p* = .028) educational levels. However, no significant difference was obtained between those of the primary and secondary educational levels (*Q* = 0.103, *p* = .748).

Control conditions. Regarding the moderator effect of control conditions, both paper-and-pencil (*g* = 0.645, *p* < .001) and computer-based (*g* = 0.932, *p* < .001) conditions were found to be significant. The moderator effect of control conditions did not reach a significant level (*Q* = 1.210, *p* = .271).

Intervention settings. Intervention settings consist of instructional activities occurred in the classroom (classroom setting) and outside of the classroom (outdoor setting). The classroom setting (*g* = 0.821, *p* < .001) had a high effect size, while the outdoor setting (*g* = 0.760, *p* < .001) had a moderate effect size. The moderator effect of intervention settings did not reach a significant level (*Q* = 0.051, *p* = .821).

Software types. According to Rosell-Aguilar (2017), MALL software could be taxonomized into two categories: educational and general purposes. The MALL software for educational purposes achieved a significantly high effect size ($g = 0.843, p < .001$), and general purposes had a moderate effect size ($g = 0.666, p < .001$). No significant between-group difference was found between them ($Q = 0.599, p = .439$).

Measured outcome types. The standardized tests achieved a large effect size ($g = 0.805, p < .001$), and researcher-designed tests had a moderate-to-large effect size ($g = 0.745, p < .001$). The effect size did not vary based on whether the standardized or researcher-designed tests were used ($Q = 0.056, p = .813$), indicating the robustness of research outcomes between studies.

Intervention durations. When integrated with MALL for EFL listening skill development, long-term (> 4 weeks, \leq one semester) durations had a high effect size, and intermediate-term (> 1 week, ≤ 4 weeks) durations had a moderate effect size (Table 3). However, the significant effect size of short-term (one session, ≤ 1 week) durations was not found, $Z = 1.696$, 95% CI $[-0.158, 2.193]$, $p = .090$. Between-group comparisons did not achieve a statistical significance level ($Q = 0.341, p = .843$), suggesting the effect size did not significantly differ among the three categories.

5. Discussion

This study represents a meta-analytic approach to the effectiveness of MALL for EFL listening skill development over the past decade. Through the meta-analysis, we found 20 experimental studies involving a total of 1218 EFL learners that met the proposed inclusion and exclusion criteria. Results of the meta-analysis showed that the overall effect size was moderate-to-large, suggesting that MALL for EFL listening skill development is more effective than traditional methods. In addition to the overall effect of MALL for EFL listening skill development, this study also reported the different moderator effects of educational levels, software types, control conditions, intervention settings, measured outcome types and intervention durations. Research findings were discussed to answer the two research questions in the remainder of this section accordingly.

Regarding research question 1, the result showed that EFL learners benefit more from MALL for their listening skill development than from traditional methods, and the effect size of 0.792 could be comparable to several recently published meta-analyses of MALL for language learning ($g = 0.720$, Burston & Giannakou, 2022; $g = 0.722$, Chen et al., 2021). After a scrutiny of those studies included, three possible reasons could be tentatively concluded to explicate the pedagogical benefits of MALL for EFL listening skill development. First, the mobility feature of MALL in its own right supports ubiquitous and autonomous learning. In other words, EFL learners can use MALL to autonomously practice listening comprehension from anywhere and at any time (Li, 2022b), which will increase the input exposure rates of learning materials compared to the highly temporal and spatial-constrained traditional methods, e.g., indoor computer classroom (Chang et al., 2018) and conventional paper and pencil (Azar & Nasiri, 2014). The increased input exposure rates of MALL also lend support from the old saying—practice makes perfect. Second, the multimodal materials of MALL reduce working memory loads and facilitate listening comprehension process. According to the Dual Coding Theory (Paivio, 1991), there are verbal and non-verbal channels of working memory that process information independently from one another. When the two channels are interconnected with each other, working memory loads will be decreased and the learning outcomes will be improved (Li, 2021a). Compared to the unimodal presentation of the traditional methods, EFL learners who used MALL can make full use of multimodal materials to practice listening comprehension (Mayer, 2009). For instance, EFL learners' listening skill was greatly improved with a mobile VR due to the multimodal virtual presence and the high degree of immersion (Tai & Chen, 2021). Third, the interactivity of MALL affords enjoyable listening experiences, which increases EFL learners' flow experiences (Li et al., 2021), motivation and engagement (Li, 2021a; Li et al., 2019), and self-efficacy (Li, 2021b). For instance, within a mobile VR environment, the interaction between realistic environment and spatial audio allows learners to feel “being there” and “being participants”, which triggers their flow experiences, motivation and engagement, and facilitates listening skill development in turn (Tai & Chen, 2021).

Research question 2 dealt with the moderator effects of educational levels, software types, control conditions, intervention settings, measured outcome types and intervention durations. Moderator analysis results are discussed as follows.

Educational levels. While pedagogical benefits of MALL for EFL listening skill development are obtained among EFL learners of all educational levels, larger effect regarding college students of tertiary education over

those of primary and second education may be attributed to the restricted use of mobile devices for children of primary and secondary education, rather than adults of tertiary education (Wiederhold, 2019). Another possible explanation might lie in researchers' inadequate attention to participants of primary and secondary education, warranting further research in this regard. A closer look at the unbalanced distribution of educational levels indicates that, among the 20 eligible experimental studies, only 35% of the participants ($k = 7$ out of 20) who adopted MALL for EFL listening skill development are children of primary and secondary education. By contrast, 65% of the participants ($k = 13$) are college students.

Control conditions. While both paper-and-pencil and computer-based conditions have significant effect sizes, the moderator effect of control conditions did not reach a significant level, suggesting that MALL for listening skill development is robustly effective, regardless of the difference in control conditions. To elucidate the reasons, it will come as no surprise to find that, although some researchers (e.g., Alabsi, 2020; Kargozari & Tafazoli, 2012; Shiri, 2015) designed the control group with the paper-and-pencil method, while others (e.g., Rashtchi & Mazraehno, 2019 using DVDs and audio CDs; Zhang, 2016 using traditional CD Room; Rahimi & Soleymani, 2015 using desktop computers) with the computer-based method, all the studies included are under strict (quasi-)experimental design, warranting the reliability and robustness of the results reported.

Intervention settings. While both the informal outdoor and the formal classroom settings are beneficial, the pedagogical benefit of classroom setting is higher than that of outdoor setting, suggesting that EFL learners who used MALL for formal classroom listening practice would outperform those did informally. This result is, however, inconsistent with domain-general meta-analyses on MALL (Chen et al., 2020; Sung et al., 2015; Sung et al., 2016), advocating larger effect size in informal outdoor setting than in classroom setting. A plausible explanation for the discrepancy might rest on the different domains of investigation: domain-general vs. domain-specific. In other words, different from those domain-general meta-analyses of MALL research, this study meta-analyzed the domain-specific MALL for EFL listening skill development that requires formal and intensive classroom instruction to warrant its pedagogical benefit (Sung et al., 2015).

Software types. The higher effect of MALL devices for educational purposes than for general purposes corroborates the result of Sung and colleagues (2016), who posited that compared to MALL devices for general purposes, MALL devices for educational purposes integrated functionalities well with the content of curriculums (Li, 2022b). Thus, MALL devices for educational purposes, e.g., mobile English listening systems (Liu et al., 2018) and subtitles for English listening (Alabsi, 2020), are better tailored to EFL learners' personalized needs than those of general purposes (Chen et al., 2020).

Measured outcome types. The effect size did not vary between the standardized or researcher-designed tests, which is consistent with previous finding that compared the moderator effect between standardized and researcher-designed tests in a meta-analysis of writing devices (Xu et al., 2019), suggesting both standardized tests and researcher-designed tests could warrant a robust reliability to measure EFL learners' listening comprehension performance.

Intervention durations. Although there is no significant between-group difference of short-term, intermediate-term and long-term durations, long-term durations achieved a high effect size of 0.817, showing a consistent result compared to researchers who advocating that "long-term teaching interventions are important for obtaining reliable results" (Sung et al., 2016, p. 264). Moreover, unlike other skills (e.g., vocabulary learning) that could be improved over a short period of time, EFL listening skill development requires long-term practice and training (Sung et al., 2015). It is promising to observe that 60% of the studies ($k = 12$ out of 20) involved listening tasks are carried out for long-term (> 4 weeks, \leq one semester) durations, while only 10% of the studies ($k = 2$) are for short-term (one session, ≤ 1 week) durations, suggesting that appropriate long-term intervention durations and opportunities for mobile listening practice are most favored. This may explain the reason why the beneficial effects of MALL for EFL listening skill development for short-term is not significant.

6. Implications

With regard to the aforementioned findings, the following practical implications could also be inferred for instructors, designers and researchers.

6.1. Implications for instructors

First, since MALL for EFL listening skill development has been confirmed to be more effective than traditional methods, its pedagogical potentials should be encouraged to explore among learners of different educational levels, especially among children of primary and secondary educational levels who are often banned from mobile devices use (Wiederhold, 2019). Second, while robust effects of MALL for listening skill development could be obtained regardless of the difference in control conditions, an experimental-control group comparison regarding “MALL vs. computer-assisted” design achieves higher effect size ($g_{\text{computer-based}}$ vs. $g_{\text{paper-and-pencil}} = 0.932$ vs. 0.645) than that of “MALL vs. paper-and-pencil” design, suggesting that instructors should adopt the former design in their EFL listening instruction to achieve better pedagogical effectiveness. Third, given the higher effect of MALL for educational purposes than for general purposes, instructors should help EFL learners select and tailor appropriate MALL devices and prepare useful multimodal listening materials based on their personalized needs.

6.2. Implications for designers

First, MALL technology designers should consider EFL learners’ educational levels. That is to say, the difficulty of listening tasks could be dynamically tailored with reference to the ongoing detection of educational levels. Second, designers should collect instructors and learners’ opinions regarding the use of MALL applications to adjust the different intervention settings. Third, to strengthen the educational purposes, designers should try to integrate curriculums into the MALL applications depending on the educational needs and purposes.

6.3. Implications for researchers

First, as scant attention has been paid on participants of primary and secondary educational levels, researchers should focus on how the use of MALL can facilitate the listening skill development among learners of primary and secondary education. Second, given that longer durations achieve higher effect sizes regarding the use of MALL for EFL listening skill development, researchers should design the research with long-term durations to increase the reliability of findings (Sung et al., 2016). Third and importantly, while this study sheds light on the effectiveness of MALL for EFL learners’ listening skill development, there is a desperate need for researchers to focus on MALL for other foreign language listening comprehension, e.g., Chinese as a foreign language (CFL).

7. Conclusion

To gain a more comprehensive understanding into the effects of MALL for EFL learners’ listening skill development, this study conducted a meta-analysis of 20 selected experimental studies involving 1218 EFL learners. Results in response to the two main research questions were concluded: First, the overall effect size was moderate-to-large, suggesting the use of MALL for EFL listening skill development is more effective than traditional methods. Second, moderator analysis results indicated that educational levels were found to be a significant moderator, while no significant moderator effects of software types, control conditions, intervention settings, measured outcome types and intervention durations were obtained.

While results of this study may contribute to advancing our understanding regarding the effects of MALL for EFL listening skill development and the potential moderators that may affect such effects, there are some limitations with the study: First, given the limited studies of MALL for other FL listening skill development, it is rather premature to consider other foreign languages in the meta-analysis this time. As such, we only considered EFL learners who used MALL for listening skill development. Future research should not only consider the moderator effect of different foreign languages, but also meta-analyze the effectiveness of MALL for a particular FL listening comprehension. Second, due to the limited keyword combinations and strict inclusion/exclusion criteria, only 20 experimental studies that met the criteria were analyzed. Future study may consider the inclusion of more publications on MALL for EFL listening skill development. Third, due to small number of experimental studies with affective outcomes that are premature for meta-analysis, this study only involved standardized or researcher-designed test scores for listening outcome. Future study should take EFL learners’ affective outcomes into account when the number of primary studies is sufficient for analysis. Lastly, drawing on existing meta-analyses for MALL, we only reported the moderator effects of educational levels, software types, control conditions, intervention settings, measured outcome types and intervention durations for the overall effect size. Future attempt should consider other potential moderators with sufficient information for calculation.

Acknowledgement

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

References

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

- *Alabsi, T. (2020). Effects of adding subtitles to video via Apps on developing EFL students' listening comprehension. *Theory and Practice in Language Studies*, 10(10), 1191–1199. <https://doi.org/10.17507/tpls.1010.02>
- *Al-Shamsi, A., Al-Mekhlafi, A. M., Busaidi, S. A., & Hilal, M. M. (2020). The Effects of mobile learning on listening comprehension skills and attitudes of Omani EFL adult learners. *International Journal of Learning, Teaching and Educational Research*, 19(8), 16–39. <https://doi.org/10.26803/IJLTER.19.8.2>
- *Andujar, A., & Hussein, S. A. (2019). Mobile-mediated communication and students' listening skills: A Case study. *International Journal of Mobile Learning and Organisation*, 13(3), 309–332. <https://doi.org/10.1504/IJMLO.2019.100432>
- *Azar, A. S., & Nasiri, H. (2014). Learners' attitudes toward the effectiveness of mobile assisted language learning (MALL) in L2 listening comprehension. *Procedia - Social and Behavioral Sciences*, 98, 1836–1843. <https://doi.org/10.1016/j.sbspro.2014.03.613>
- 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., & 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, C., Warden, C. A., Liang, C., & Chou, P. (2018). Performance, cognitive load, and behaviour of technology-assisted English listening learning: From CALL to MALL. *Journal of Computer Assisted Learning*, 34(2), 105–114. <https://doi.org/10.1111/jcal.12218>
- 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>
- 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>
- Coşkun, A., & Marlowe, Z. (2020). The Place of technology-assisted language learning in EFL listening: A Review of literature and useful applications. In M. Durnali & İ. Limon (Eds.), *Enriching Teaching and Learning Environments with Contemporary Technologies* (pp. 102–116). IGI Global.
- Costabile, M. F., De Angeli, A., Roselli, T., Lanzilotti, R. & Plantamura, P. (2003). Evaluating the educational impact of a tutoring hypermedia for children. *Information Technology in Childhood Education Annual*, 1, 289–308. <https://www.learntechlib.org/primary/p/18867/>
- *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. <https://doi.org/10.1080/09588221.2012.733710>
- Holden, C. L., & Sykes, J. M. (2011). Leveraging mobile games for place-based language learning. In P. Felicia (Ed.), *Developments in Current Game-Based Learning Design and Deployment* (pp. 27–45). IGI Global.
- *Hsu, C., Hwang, G., Chang, Y., & Chang, C. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society*, 16(1), 403–414.
- Hwang, G., & Fu, Q. (2019). Trends in the research design and application of mobile language learning: A Review of 2007–2016 publications in selected SSCI journals. *Interactive Learning Environments*, 27(4), 567–581. <https://doi.org/10.1016/j.compedu.2018.01.004>

- *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>
- Jia, C., & Hew, K. F. (2021). Toward a set of design principles for decoding training: A Systematic review of studies of English as a foreign/second language listening education. *Educational Research Review*, 33, 1–18. <https://doi.org/10.1016/j.edurev.2021.100392>
- *Kargozari, H. R., & Tafazoli, D. (2012). Podcasting: A Supporting tool for listening. In *Proceedings of International Technology, Education and Development (INTED2012) Conference* (pp. 3870–3873). International Association of Technology, Education and Development (IATED).
- *Kim, H. (2013). Emerging mobile apps to improve English listening skills. *Multimedia-Assisted Language Learning*, 16(2), 11–30. <https://doi.org/10.15702/mall.2013.16.2.11>
- Lai, Y., Saab, N., & Admiraal, W. (2022). Learning strategies in self-directed language learning using mobile technology in higher education: A Systematic scoping review. *Education and Information Technology*, 27, 7749–7780. <https://doi.org/10.1007/s10639-022-10945-5>
- Li, R. (2021a). 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. (2021b). Modeling the continuance intention to use automated writing evaluation. *SAGE Open*, 11(4), 1–13. <https://doi.org/10.1177/21582440211060782>
- Li, R. (2022a). Effects of mobile assisted language learning on EFL/ESL reading comprehension. *Educational Technology & Society*, 25(3), 15–29.
- Li, R. (2022b). 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. (2022c). 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. (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. (in press). Effects of mobile-assisted language learning on foreign language learners' speaking skill development. *Language Learning & Technology*.
- 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>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE.
- *Liu, G., Chen, J., & Hwang, G. (2018). Mobile-based collaborative learning in the fitness center: A Case study on the development of English listening comprehension with a context-aware application. *British Journal of Educational Technology*, 49(2), 305–320. <https://doi.org/10.1111/bjet.12581>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Mendoza, V., la Torre, A. V., & Páez-Quinde, C. (2020). Mobile Apps (English listening and 6 Minutes English) and the listening skill. In M. E. Auer, H. Hortsch & P. Sethakul (Eds.), *The Impact of the 4th Industrial Revolution on Engineering Education* (pp. 60–66). Springer.
- *O, K.-M. (2015). The Effectiveness of mobile assisted language learning on L2 listening comprehension. *Multimedia-Assisted Language Learning*, 18(2), 135–158. <https://doi.org/10.15702/mall.2015.18.2.135>
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Journal of Psychology*, 45(3), 255–287. <https://doi.org/10.1037/h0084295>

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

*Rahimi, M., & Soleymani, E. (2015). The Impact of mobile learning on listening anxiety and listening comprehension. *English Language Teaching*, 8(10), 152–161. <https://doi.org/10.5539/elt.v8n10p152>

*Rashtchi, M., & Mazraehno, M. R. T. (2019). Exploring Iranian EFL learners’ listening skills via TED talks: Does medium make a difference? *Journal of Language and Education*, 5(4), 81–97. <https://doi.org/10.17323/jle.2019.9691>

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>

Richards, J. C. (2005). Second thoughts on teaching listening. *RELC Journal*, 36(1), 85–92. <https://doi.org/10.1001/jama.298.9.1002>

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.

Saeedakhtar, A., Haqju, R., & Rouhi, A. (2021). The Impact of collaborative listening to podcasts on high school learners’ listening comprehension and vocabulary learning. *System*, 101, 102588. <https://doi.org/10.1016/j.system.2021.102588>

*Seo, W., & Choi, I. (2014). The Effect of using a smart-phone speaking application on Korean middle school students’ English expression learning and satisfaction. *Multimedia-Assisted Language Learning*, 17(1), 34–57. <https://doi.org/10.15702/mall.2014.17.1.34>

*Shiri, S. (2015). The Application of podcasting as a motivational strategy to Iranian EFL learners of English: A View toward listening comprehension. *Advances in Language and Literary Studies*, 6(3), 155–165. <https://doi.org/10.7575/aiac.all.v.6n.3p.155>

Siegel, J. (2014). Exploring L2 listening instruction: Examinations of practice. *ELT Journal*, 68(1), 22–30. <https://doi.org/10.1093/elt/cct058>

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>

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A Second-order meta-analysis and validation study. *Review of Educational Research*, 81, 4–28. <https://doi.org/10.3102/0034654310393361>

*Tai, T., & Chen, H. H. (2021). The Impact of immersive virtual reality on EFL learners’ listening comprehension. *Journal of Educational Computing Research*, 59(7), 1272–1293. <https://doi.org/10.1177/0735633121994291>

Thorne, S. L., Hellermann, J., & Jakonen, T. (2021). Rewilding language education: Emergent assemblages and entangled actions. *The Modern Language Journal*, 105(S1), 106–125. <https://doi.org/10.1111/modl.12687>

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.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Wiederhold, B. K. (2019). Should smartphone use be banned for children? *Cyberpsychology, Behavior, and Social Networking*, 22(4), 235–236. <https://doi.org/10.1089/cyber.2019.29146.bkw>

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>

*Zhang, Y. (2016, December). *The Impact of mobile learning on ESL listening comprehension* [Paper presentation]. 3rd International Conference on Advanced Education and Management (ICAEM 2016). <https://doi.org/10.12783/dtssehs/icaem2016/4290>

Zheng, D., Dai, Y., & Liu, Y., (2017). Place, identity and becomings with mobile technologies: Feminist and Chinese philosophies. *International Communication of Chinese Culture*, 4(3), 319–337. <https://doi.org/10.1007/s40636-017-0106-1>

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. <https://doi.org/10.1016/j.linged.2017.10.004>