ReCALL (2026), 38: 1, 75–93
doi:10.1017/S0958344025100335

RESEARCH ARTICLE

A meta-analysis on mobile-assisted vocabulary learning: Do mobile applications help?

Yonghong Zhou
University of Macau, Macau SAR (yc27110@um.edu.mo)

Mingming Zhou
University of Macau, Macau SAR (zhmingming@gmail.com)

Abstract

This study synthesized 65 (quasi-)experimental studies published between 2010 and 2024 that examined the use of mobile applications to develop language learners' vocabulary learning. Bayesian meta-analysis was adopted to assess (1) overall effect size; (2) subgroup analyses (i.e. education level, vocabulary knowledge, aspects of vocabulary learning, learning environment, sample size, mobile application type, gender, and cultural background); and (3) publication bias. A large effect size of 1.28 was found for the overall effectiveness of using mobile applications for vocabulary learning when we restricted the studies to long-term treatment duration of 10 weeks or above. Each moderator was analyzed and discussed, and implications for language teaching and research were provided.

Keywords: mobile-assisted language learning; vocabulary learning; meta-analysis

1. Introduction

The important role of vocabulary learning in language learning has been emphasized by researchers and educators. Indeed, as published studies attest, a lack of vocabulary usually leads to poor language skills, which further results in decreased motivation in vocabulary learning (Hao, Wang & Ardasheva, 2021) and subsequently in poor learning performance (Li & Tong, 2020). Moreover, EFL learners often struggle with vocabulary learning because of the complexity of knowing a word and the large number of core words (Yu & Trainin, 2022). To address this challenge, educators have increasingly turned to mobile technology as a potential remedy to improve vocabulary learning processes and outcomes.

The use of mobile applications in language learning has been shown not only to enrich language learners' learning experiences but also to customize learning processes according to individual needs (Lee, 2023). However, learning outcomes have not been entirely positive. Some recent studies have not found significant differences between the use of mobile and traditional forms of instruction (Rachels & Rockinson-Szapkiw, 2018). In addition, Li and colleagues (2021) observed a decline in students' vocabulary performance after three weeks of a WeChat-assisted lexical learning program, suggesting that mobile applications may be ineffective and could even

Corresponding author: Mingming Zhou; Email: zhmingming@gmail.com

Cite this article: Zhou, Y. & Zhou, M. (2026). A meta-analysis on mobile-assisted vocabulary learning: Do mobile applications help?. *ReCALL* 38(1): 75–93. https://doi.org/10.1017/S0958344025100335

© The Author(s), 2025. Published by Cambridge University Press on behalf of EUROCALL, the European Association for Computer-Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

hinder learning. These inconsistent results attest to the need to synthesize existing findings to reach a more convincing conclusion. Therefore, the purpose of this study was to examine whether mobile applications were more effective than traditional face-to-face methods through meta-analyzing existing primary studies.

2. Literature review

2.1. Vocabulary knowledge

Vocabulary is a central component of language proficiency (Schmitt & Schmitt, 2020). Nation (2013) stresses that language learners have to not only learn many words (vocabulary breadth) but also master a range of knowledge aspects related to individual words (vocabulary depth). Using a component approach, Nation (2013) proposes three subconstructs of vocabulary learning: form, meaning, and use. Each can be further broken into two levels of mastery: receptive knowledge mastery and productive knowledge mastery. Of course, there are other important areas such as semantic or morphological aspects of vocabulary learning, which have been frequently examined in linguistic studies (see Levin & Hovav, 2017). As we are particularly interested in vocabulary learning in this study, discussions of those aspects are beyond the scope of the current research.

When it comes to teaching and learning vocabulary, researchers tend to separate vocabulary knowledge into receptive and productive knowledge (Milton, 2009). However, overemphasis on the receptive–productive distinction could lead to the misconception that receptive knowledge always precedes productive knowledge, overlooking the fact that productive knowledge of one aspect could precede receptive knowledge of another, and a learner could write a word correctly before fully understanding all its form–meaning connections. Further, this receptive versus productive view tends to generate inconsistent results when examining the effectiveness of technology use on vocabulary learning: some have confirmed the benefit of technology use (Hao *et al*., 2021), whereas others disagree (Lin & Lin, 2019; Yu & Trainin, 2022). As Nation's (2013) framework considers multiple dimensions of vocabulary learning and levels of vocabulary mastery, vocabulary knowledge has more often been described as multifaceted, rather than a simple binary construct, including spelling, word collocation, words' grammatical functions, and so on. Therefore, it is necessary to consider levels of mastery as well as aspects of vocabulary knowledge to assess the effectiveness of technology use on vocabulary learning.

2.2. Mobile-assisted vocabulary learning

MAVL (mobile-assisted vocabulary learning) can be defined as vocabulary learning with the help of mobile technologies, including mobile short message service/multimedia message service, mobile learning apps, and context-aware mobile technologies (Lin & Lin, 2019). It stands in contrast with conventional teaching, which involves face-to-face interaction in the classroom through structured resources such as textbooks, blackboards, and slides (Yap, 2016). According to the cognitive theory of multimedia learning (Mayer, 2005), the usefulness of using mobiles in vocabulary learning rests on the fact that multimedia learning combines both words and pictures, which enhances understanding by facilitating the development of verbal and visual mental models. While working with two separate channels, learners can process information differently, and each set of information can be organized into models that help learners understand and remember.

Recent years have witnessed a significant growth of research on MAVL. The majority of studies have focused on the effectiveness of mobile use on vocabulary learning motivation (Chen, Wang, Zou, Lin & Xie, 2019) and knowledge development (Li & Hafner, 2022). While scholars have concurred on the benefits of mobile devices in vocabulary learning, other researchers have pointed out the challenges associated with MAVL. Indeed, some studies report no significant differences in

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

delayed vocabulary tests between learning with mobiles and learning with print materials (e.g. Lai, 2016). Further, studies that support the effectiveness of MAVL typically have short-term treatments, which leaves unanswered the question of whether MAVL could have a long-term positive impact on vocabulary learning (Burston & Giannakou, 2022). Hence, the effectiveness of using mobile applications in vocabulary learning remains inconclusive.

2.3. Present study

The present study aimed to examine the overall effectiveness of MAVL and potential moderators in this relationship. Specifically, two research questions guided this study:

1. To what extent is mobile technology effective in improving foreign language learners' vocabulary learning?
2. Is the effect of mobile technology on foreign language learners' vocabulary learning significantly moderated by gender, education level, cultural background, vocabulary knowledge, vocabulary aspect, learning environment, sample size, type of mobile application, or language proficiency?

To achieve a holistic understanding of this issue, several meta-analyses have been conducted, as summarized in Table 1. The current review differed from these studies in several respects. First, Lin and Lin (2019), Mahdi (2018), and Hao *et al.* (2021) only included studies with small sample sizes and short durations. Given that the perceived effectiveness of the intervention tends to be substantially exaggerated in such experiments (Burston & Giannakou, 2022), we only considered studies with larger sample sizes and longer duration of the intervention to reach more substantiated conclusions. Further, a precise calculation of moderator effects should involve at least seven studies in each subgroup of moderators (Borenstein, Hedges, Higgins & Rothstein, 2010). However, many previous reviews did not meet this requirement (e.g. Lin & Lin, 2019; Mahdi, 2018). Finally, previous meta-analyses tend to explore vocabulary learning outcomes as a whole (e.g. Chen, Chen, Jia & An, 2020) or emphasize vocabulary acquisition based on vocabulary types without considering the vocabulary aspect (e.g. Lin & Lin, 2019; Yu & Trainin, 2022). They consequently fail to present a full picture on this topic.

As a result, the current meta-analysis targeted studies with larger sample sizes, longer duration of intervention, and more studies in each subgroup of moderators by anchoring between a longer time span of publications (15 years) and a wider research scope (considering studies of vocabulary knowledge types and aspects). It is noteworthy that much attention has been paid to English learning in previous reviews, neglecting other languages (Oakes & Howard, 2022). Hence, the current review did not limit the search to English as the target learning language.

3. Methods

3.1. Data collection

The literature search was performed in June 2025 in two steps. First, to conduct a comprehensive literature review, an annotated bibliography of previous mobile-assisted language learning (MALL) review studies was examined for possible keywords (Lin & Lin, 2019). Together, three sets of search terms were applied to locate relevant literature (see Supplementary Material A): mobile-related keywords, language learning–related keywords, and vocabulary learning–related keywords. Second, major research databases were thoroughly searched, including Web of Science, the Education Resources Information Center, Education Search Complete, Scopus, Google Scholar, and CNKI. Keywords in English and Chinese were used, with no language restrictions in the search results. In addition to publications in English and Chinese, research published in other languages that provided an English title and/or abstract was also retrieved. Due

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

Table 1. Summary of prior meta-analysis studies in mobile-assisted/technology-assisted vocabulary learning

Year inclusion	Author(s)	Year of publication	No. of studies	Sample size	Intervention duration	Moderators	Learning outcome	Overall effect size (research design)
2006 to 2017	Yu & Trainin	2022	34	2,511	NA	• Type of instruction (incidental, intentional) • Type of assessment (recognition, production) • Linguistic distance (non-Indo-European, Indo-European) • Grade level (university, K-12) • Type of technology (computer-assisted language learning & mobile-assisted language learning)	Receptive & productive knowledge	.64 (between groups)
2005 to 2018	Lin & Lin	2019	33	NA	1–16 weeks	• Research settings (formal, informal, mixed) • Treatment duration (short, long) • Level of autonomy • Types of vocabulary learning (receptive, productive vocabulary learning, mixed)	Second language (L2) word retention	1.005 (between groups and within groups)
Up to 2017	Mahdi	2018	16	NA	1–16 weeks	• Types of vocabulary learning (receptive, productive vocabulary learning) • Age of learners • Aspects of vocabulary (form, meaning, use)	Vocabulary learning	.67 (between groups)
2012 to 2018	Hao *et al.*	2021	45	2,374	1–16 weeks	• Participant's education level • Participant's L2 proficiency • Contextual features (devices, game, study duration, setting, test format) • Methodology (research design, reliability, measures)	Receptive & productive & mixed	.845 (between groups)

to technical issues, we were unable to access all of them. As a result, only studies in English, Korean, Turkish, and Chinese were included in the data set. We also hand-searched key educational and applied linguistics journals, peer-reviewed book chapters, and tracked reference lists from relevant papers. Moreover, we also consulted existing MALL bibliographies (i.e. Burston, n.d.; Burston, 2025).

3.2. Eligibility criteria

Duman, Orhon and Gedik (2015) documented a sharp growth in MALL publications from 2008 onward. Thus, we chose to review studies published from 2008 onwards to ensure that studies under review would reflect the influence of mobile devices on language learning. The following criteria were used to select articles for analysis in the study:

- The study adopted either an experimental or a quasi-experimental design and reported empirical data of between-group comparisons with a control group. Studies with only pre-/post-test designs were excluded.
- The outcome variables in the study should be specific to vocabulary learning, rather than to language learning in general.
- The technology used in the treatment must be exclusively mobiles, not other technologies such as laptops or desktops.
- The study employed second language (L2) learner samples from K-12 to higher education.
- Studies with sample sizes of 15 or more in each experimental/control group (i.e. a total sample size of 30 or more) and a treatment duration of at least 10 weeks were included.
- The study provided detailed sample information and sufficient statistical information (i.e. standard deviation [*SD*], mean [*M*], statistical significance [*p*], and sample size [*N*]).

Studies were excluded for the following reasons:

- Studies that focused on the perceptions, attitudes, or motivation towards MAVL (e.g. Sun & Gao, 2020).
- Studies with insufficient data for statistical analyses (e.g. Chen, Jia & Li, 2021).
- The full text of the study was unavailable.

3.3. Selection process

The first author independently screened the titles and abstracts of identified papers for eligibility. Studies that were not published in English or Chinese but showed potential relevance to the focus of this review were further evaluated to see if they met the inclusion criteria, using Google Translate. After narrowing down the number of articles, both authors examined the full texts of potentially suitable papers to ensure their eligibility for inclusion. The interrater agreement was initially 90% and was then brought to 100% agreement after further discussion. The PRISMA flow diagram (Liberati *et al.*, 2009) showed that the initial database search yielded a total of 10,163 records, and 6,461 publications were excluded after reviewing the abstracts. Another 1,591 records were removed due to duplication. This led to a total of 2,111 publications being assessed for eligibility, from which 68 studies were selected for further review (Figure 1). A full list of the studies included in this review is presented in Supplementary Material B.

3.4. Coding scheme

We performed data coding in EndNote 20. The coding was conducted by the two authors independently. For interrater reliability, Cohen's kappa coefficient was .80, with an interrater agreement of 90%. When disagreement arose, the authors cross-checked the accuracy of data

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

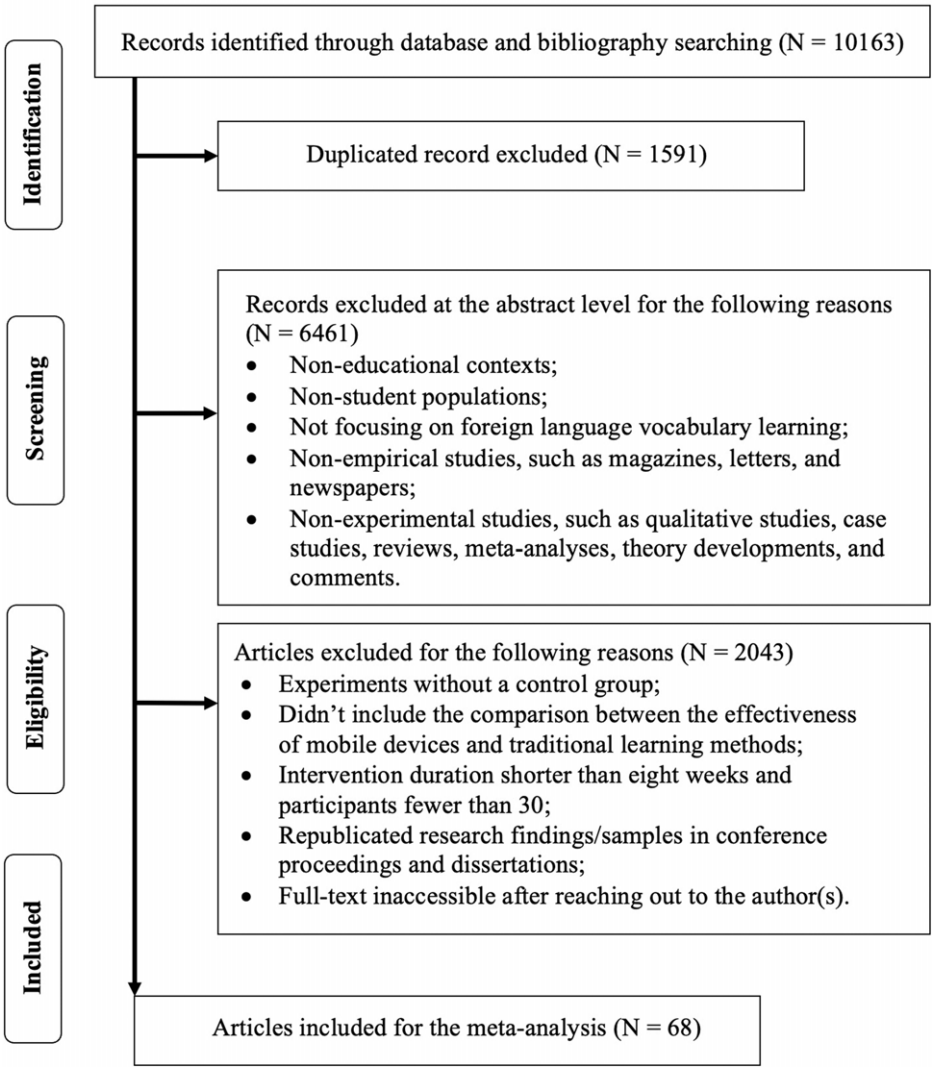


Figure 1. The PRISMA flow diagram for the review.

coding to ensure no errors were involved and came to a final agreement. A total of nine moderator variables were coded for each study.

Gender. Past studies have shown that learners of different genders interact with mobile devices for learning activities differently (Şad, Özer, Yakar & Öztürk, 2022). Hence, gender was considered as a moderator and coded as a continuous variable using the percentage (%) of female participants in each study, ranging from 0% to 100%.

Education level. Education level has been found to be a significant moderator in MAVL in previous analyses (Mahdi, 2018), with senior students benefiting more than junior students. Therefore, in this study, education level was considered as a moderator and coded as a categorical moderator in three levels: primary, secondary, and tertiary.

Cultural background. Learners from diverse cultural backgrounds have tended to hold varying attitudes toward MAVL in terms of technological affordances and applicability (Hsu, 2013). Hence, cultural background was considered as a moderator and was coded as a continuous

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

variable according to Hofstede and colleagues' (2010) individualism index, ranging from 0 to 100. Countries with individualism index values under 50 were characterized as collectivistic countries.

Vocabulary knowledge. According to previous meta-analyses (Lin & Lin, 2019), learners typically gained more receptive vocabulary knowledge than other types of vocabulary knowledge when learning with mobile technology. We thus considered it as a moderator and coded vocabulary knowledge based on the types of vocabulary knowledge (Nation, 2013). If the study measured both receptive and productive vocabulary knowledge, it was coded as "receptive + productive knowledge."

Vocabulary aspect. Vocabulary aspects have been found to serve as a moderator in the relationship between the use of mobile applications and vocabulary learning, wherein learners gained most in the meaning aspect of vocabulary learning (Mahdi, 2018). In this review, learners' vocabulary knowledge was coded as "form," "meaning," or "use." If the study examined learners' vocabulary knowledge in multiple aspects, then it is coded as a combination of these three aspects.

Mobile application type. Different types of mobile applications could significantly affect the level of effectiveness of mobile use on language learning, with educational-purpose technologies generating more learning gains than general-purpose ones (Chen *et al*., 2020). In this study, we considered mobile applications as a moderator and coded it as "educational-purpose application" (programs that are specifically designed for educational purposes), "general-purpose application" (programs that are not specifically designed for educational purposes), and "general-purpose + educational-purpose application" (programs that are designed for mixed purposes) (Chen *et al*., 2020).

Learning environment. Prior studies have shown that MAVL seems to be more efficient in unrestricted settings, as both formal and informal learning opportunities exposed learners to various resources and thus reinforced their vocabulary knowledge (Chen *et al*., 2020). In this study, we followed Chen *et al*.'s (2020) practice to categorize learning environments into three groups: classroom (learning in formal classrooms), out-of-class (learning outside formal classrooms), and unrestricted learning settings (learning in and outside formal classrooms).

Sample size. Past studies (Li, 2023) have shown that large sample sizes tend to achieve greater effect sizes than small sample size studies. Based on the classification of the number of participants in prior research (Hwang & Fu, 2019), sample size was coded as a categorical moderator in three categories: small (30–51), medium (51–100), and large (100 and above).

Language proficiency. Past studies have shown that learners of higher L2 proficiency perform better than learners of lower L2 proficiency when using mobile apps to assist vocabulary learning (Ou-Yang & Wu, 2017). We thus included L2 proficiency as a potential moderator in MAVL and classified participants' language proficiency levels as "elementary," "intermediate," and "advanced" based on the Common European Framework of Reference for Languages (Council of Europe, 2001). Only studies wherein participants' proficiency was assessed through standardized tests were coded for this moderator.

3.5. Data analysis methods

The data analysis procedure consisted of three steps: overall effect size calculation, moderator analyses, and publication bias calculation. Standardized mean difference was used to compute the overall effectiveness of MAVL by extracting standard deviation, mean, and sample size from each study. We employed a Bayesian meta-analysis method as it addresses publication bias from small-study effects and allows for the inclusion of more variables, even with small sample sizes (Thompson & von Gillern, 2020). Specifically, we used the "bayesmeta" package to calculate effect sizes. The heterogeneity was evaluated by between-study variance τ^2. Due to an anticipated higher level of heterogeneity among effect sizes (Harrer, Cuijpers, Furukawa & Ebert, 2021), we used a random-effects random model (more descriptions are presented in Supplementary Material C). Cohen's *d* (effect size) and 95% credibility interval were reported to interpret the Bayesian analysis

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

results. The benchmark of effect sizes set by Plonsky and colleagues (2023) was adopted, whereby *d* values around .40 were considered a small effect, .70 a medium effect, and 1.00 a large effect.

We adopted weakly informative priors, which was especially effective for small-sample subgroup analyses (Gelman, 2009). For the effect μ, we adopted a normal distribution with a vague effect prior centered at 0. For the heterogeneity τ, we used a half-normal distribution of .50 that assumes $\tau \leq .98$ with 95% probability (Röver, 2020). Furthermore, we conducted sensitivity analyses by changing the priors to observe how the posterior results responded to different priors (Norouzian, De Miranda & Plonsky, 2019). Various prior distributions (i.e. DuMouchel, Jeffrey, uniform) were compared and no significant difference was observed. Hence, the aforementioned prior distributions were considered appropriate.

Categorical moderators were examined with subgroup analyses, and continuous moderators (i.e. gender composition and cultural background) were examined with meta-regression analyses. For subgroup analyses, we examined each moderator separately, and then computed the differences between subgroups, using posterior means, credible intervals, and posterior probabilities. For continuous moderators, meta-regression functionality was processed by the function "bmr," and "bmr" was used for computing the coefficient β and 95% credibility interval to examine the effect of moderators.

Additionally, following van Doorn and colleagues' (2021) guidelines, we selected a statistical model and a prior distribution in the planning stage, then conducted sensitivity analysis to check prior assumptions and a Bayesian meta-analysis. With an accurate understanding of mean posterior distribution and credibility intervals, we plotted the prior and posterior distribution, reported the posterior mean and 95% credible interval, set and justified prior settings, and discussed the robustness of the analysis in the final reporting.

Publication bias would occur where studies with significant results are more likely to be published than ones with non-significant results (Harrer *et al.*, 2021). We examined publication bias using a contour-enhanced funnel plot. Specifically, a symmetrical funnel plot and a small number of detected imputed studies indicated a lower risk of publication bias, which provided robust support for the meta-analysis results (Lam & Zhou, 2022).

4. Results

4.1. Descriptions of studies

To ensure the robustness of the meta-analysis results, we followed Röver (2020) and conducted an $\widehat{R}$ value check and posterior predictive checks. We first used posterior predictive checks to identify outliers that fell outside the 95% posterior predictive intervals. Three studies (i.e. Lye, 2022; Roussel & Galan, 2018; Wu, 2015) were detected as outliers, which were subsequently excluded from the following analyses. Our data set showed an $\widehat{R}$ value below 1.01 (Harrer *et al.*, 2021), demonstrating good convergence validity. All studies were then considered to determine the association between mobile use and vocabulary learning. Among the 65 studies, quasi-experimental design was adopted in 49 studies (i.e. studies with no random assignment to the experimental group and control group of participants), and 16 studies used mixed methods (i.e. studies that used both quantitative and qualitative approaches). Most studies ($N = 59$) were published in English, four in Chinese, one in Turkish, and one in Korean. A full description of the studies in our review is presented in Figure 2.

4.2. Overall analyses

The summary is presented in Figure 3. Meta-analysis of all studies showed an overall large effect of mobile use on vocabulary learning, with a weighted mean correlation across all effect sizes of 1.28 (95% CI = [1.03, 1.52]), suggesting strong evidence in support of the hypothesis that students who

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

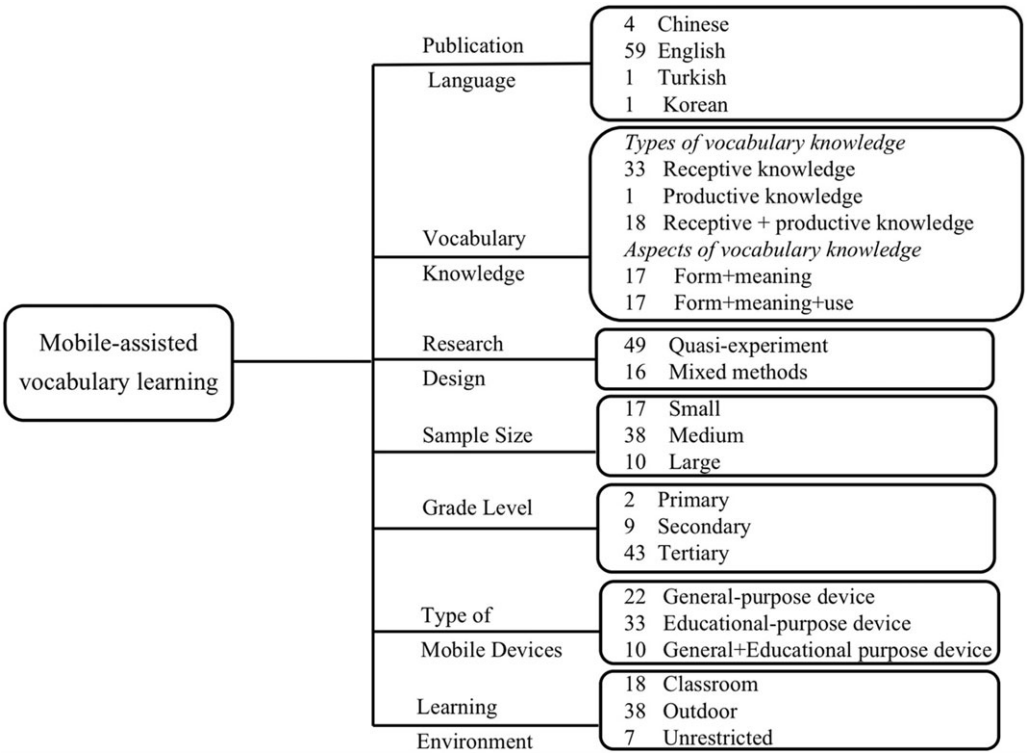


Figure 2. Descriptions of included studies (Note: The sum may be lower than 65 due to insufficient information for coding).

used mobile applications for language learning demonstrated more gains than those who did not. Further, we found a high heterogeneity of 1.07 (95% CI = [.91, 1.27]), indicating a variance in effect size and warranting further moderator analyses. Only 11 studies reported delayed post-test results, with an overall effect size of 1.41 (95% CI = [.86, 1.95]), and the heterogeneity was 1.06 (95% CI = [.75, 1.45]). Figure 4 demonstrates the prior and posterior distribution plot of effect size and heterogeneity.

4.3. Subgroup analyses

Subgroup analyses were conducted for seven moderators (excluding gender and cultural backgrounds). Table 2 provides a summary of the effect sizes, heterogeneity, and moderator analysis results. The between-group comparisons in these moderator analyses are presented in Table 3.

Gender. Only 34 studies reported gender information of the sample, and the regression weight was small ($\beta = -.45$, 95% CI = [−1.61, .71]). Regarding the heterogeneity, gender showed a high heterogeneity ($\tau = 1.00$, 95% CI = [.76, 1.25]), suggesting a moderate between-study heterogeneity in effect sizes.

Education level. Two studies included primary school students, nine studies included secondary school students, and the remaining 43 studies included university/college students. Eleven studies failed to report this information. The effect sizes tended to be larger for studies with tertiary samples ($d = 1.31$, 95% CI = [.99, 1.61]), followed by secondary school ($d = 1.26$, 95% CI = [.77, 1.73]) and primary school samples ($d = .60$, 95% CI = [−.47, 1.54]).

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

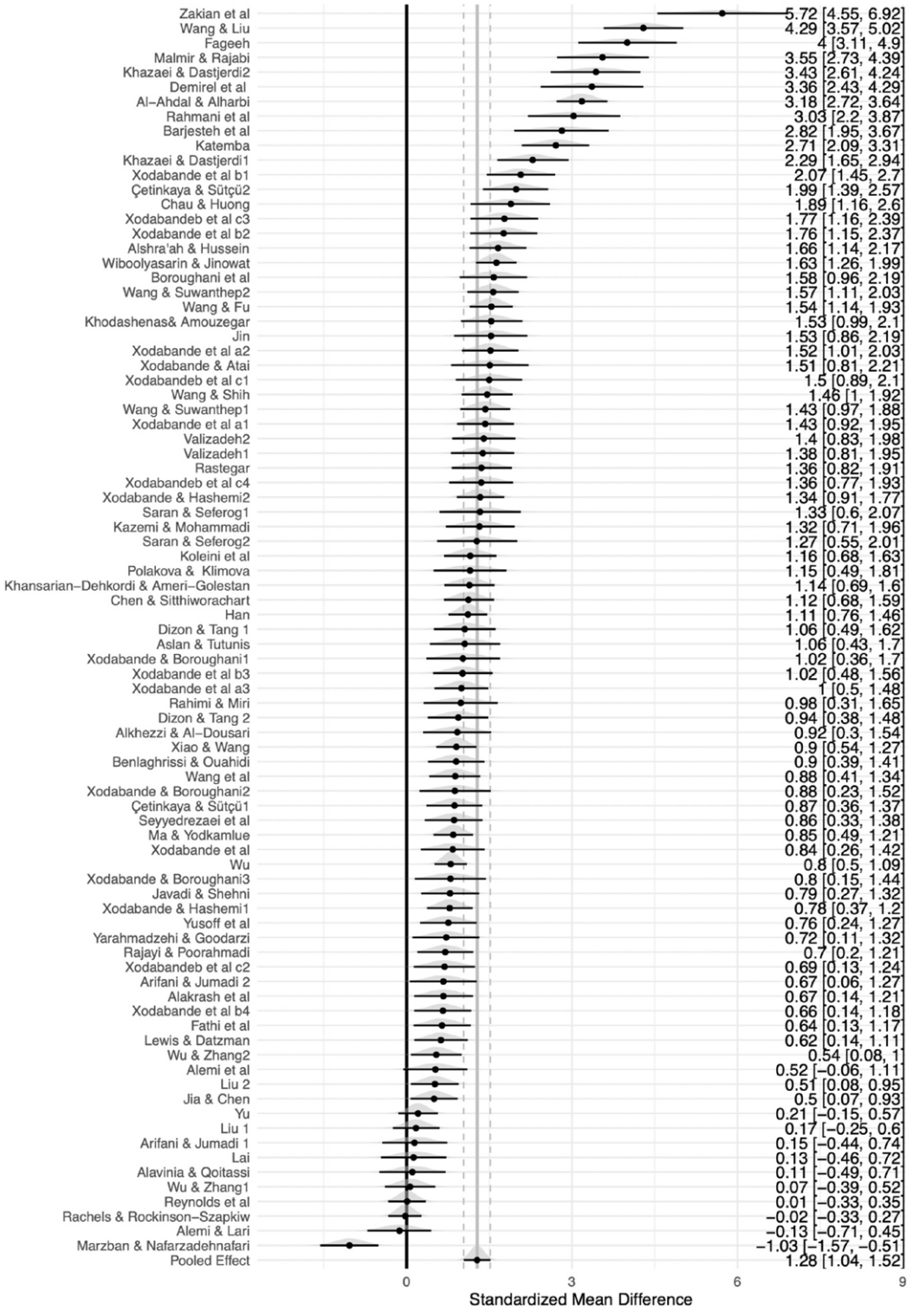


Figure 3. Forest plot for all included effect sizes.

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

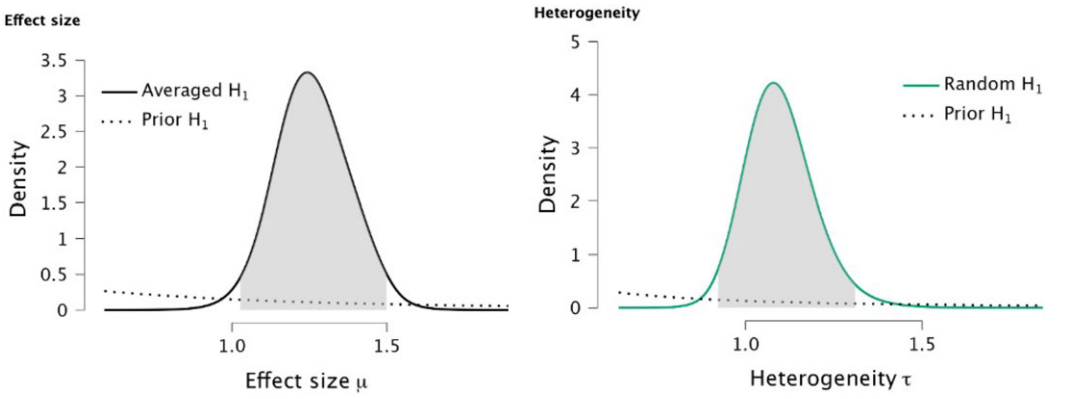


Figure 4. Prior and posterior distribution of overall effect size (μ) and between-study heterogeneity (τ) regarding all included studies.

Cultural background. All 65 studies provided cultural background information. The regression weight was very low (β = .00, 95% CI = [−.02, .02]), with a high heterogeneity (τ = 1.08, 95% CI = [.91, 1.27]), suggesting that the observed effect did not vary much across cultural backgrounds.

Vocabulary knowledge. Thirty-three studies examined receptive vocabulary knowledge, one examined productive knowledge, and another 18 examined both receptive and productive vocabulary knowledge; 13 studies did not report such information. MAVL was most efficient in receptive vocabulary knowledge ($d = 1.38$, 95% CI = [1.05, 1.69]), followed by productive vocabulary knowledge ($d = 1.25$, 95% CI = [.48, 1.92]). A combination of receptive and productive vocabulary knowledge showed the smallest effect size ($d = .66$, 95% CI = [.27, 1.04]).

Vocabulary aspects. Seventeen studies focused on the vocabulary knowledge aspect of form and meaning, and 17 examined form and meaning and use; 31 studies failed to report this information. Learning vocabulary with mobiles was found to be more efficient when learning the form and meaning of a word ($d = 1.33$, 95% CI = [.97, 1.76]) than learning all aspects of a word (i.e. form, meaning, and use) ($d = .65$, 95% CI = [.33, .97]).

Learning environment. Eighteen studies were conducted in classrooms, 38 out of class, and seven in unrestricted learning settings; two studies did not specify the learning environments. MAVL was most efficient in out-of-class contexts ($d = 1.42$, 95% CI = [1.10, 1.73]), followed by unrestricted settings ($d = 1.22$, 95% CI = [.51, 1.86]). Learning in class was found to have the smallest effect size ($d = .82$, 95% CI = [.48, 1.16]).

Sample size. Seventeen studies included a small sample size (30–51 participants), 38 included a medium sample size (51–100), and 10 involved a large sample size (over 100). MAVL was found to be the most efficient in studies with a medium sample size ($d = 1.40$, 95% CI = [1.07, 1.74]), followed by the small sample size studies ($d = 1.07$, 95% CI = [.70, 1.43]) and the large sample size studies ($d = .94$, 95% CI = [.38, 1.47]).

Mobile application type. Twenty-two studies employed general-purpose applications, 33 used educational-purpose applications, and 10 had both general-purpose and educational-purpose applications. The learning effect was found to be the largest for mobile applications designed for both general and educational purposes ($d = 1.36$, 95% CI = [.73, 1.96]), compared to mobile applications developed for educational purposes ($d = 1.31$, 95% CI = [.99, 1.66]). Mobile applications designed for general purposes revealed a smaller effect size ($d = 1.08$, 95% CI = [.69, 1.47]).

Language proficiency. Twenty-eight studies reported participants' language proficiency level, among which 21 included intermediate L2 learners, five included elementary L2 learners, and two

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

Table 2. Summary of effect sizes, heterogeneity, and moderator analyses

Model	Number of observations (*k*)	Effect size (*d*)	95% CI lower	95% CI upper	Heterogeneity (τ)	95% CI lower	95% CI upper
Overall	85	1.28	1.03	1.52	1.07	.91	1.27
	Number of observations (*k*)	Regression coefficient (β)	95% CI lower	95% CI upper	Heterogeneity (τ)	95% CI lower	95% CI upper
Gender [a]	44	−.45	−1.61	.71	1.00	.76	1.25
Cultural background [a]	85	.00	−.02	.02	1.08	.91	1.27
	Number of observations (*k*)	Effect size (*d*)	95% CI lower	95% CI upper			
Educational level							
Primary	2	.60	−.47	1.54			
Secondary	15	1.26	.77	1.73			
Tertiary	56	1.31	.99	1.61			
Vocabulary knowledge							
Receptive	51	1.38	1.05	1.69			
Productive	6	1.25	.48	1.92			
Receptive + Productive	14	.66	.27	1.04			
Vocabulary aspect							
Form + Meaning	30	1.33	.97	1.76			
Form + Meaning + Use	18	.65	.33	.97			
Learning environment							
Classroom	19	.82	.48	1.16			
Out-of-class	56	1.42	1.10	1.73			
Unrestricted	8	1.22	.51	1.86			

(*Continued*)

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

Table 2. (*Continued*)

Model	Number of observations (*k*)	Effect size (*d*)	95% CI lower	95% CI upper	Heterogeneity (τ)	95% CI lower	95% CI upper
Sample size							
Small	25	1.07	.70	1.43			
Medium	50	1.40	1.07	1.74			
Large	10	.94	.38	1.47			
Mobile application							
Educational	44	1.31	.99	1.66			
General	27	1.08	.69	1.47			
General + Educational	14	1.36	.73	1.96			
Language proficiency							
Elementary	5	.79	−.10	1.65			
Intermediate	26	1.57	1.05	2.08			
Advanced	2	1.31	.56	1.91			

Note. CI = confidence interval.[a]These two variables were coded as continuous variables, and meta-regression was used for moderator analyses.

Table 3. Summary of between-group comparisons in moderator analyses

Between-group comparison	Mean difference (*d*)	95% CI	Posterior probability of superiority
Educational level			
Secondary vs. Primary	.44	[.14, .74]	.99
Tertiary vs. Secondary	.06	[−.22, .10]	.24
Tertiary vs. Primary	.39	[.12, .64]	1.00
Vocabulary knowledge			
Productive vs. Receptive	.20	[−.05, .47]	.94
Receptive + Productive vs. Productive	−.68	[−.96, −.40]	.00
Receptive + Productive vs. Receptive	−.48	[−.62, −.33]	.00
Vocabulary aspect			
Form + Meaning + Use vs. Form + Meaning	−.48	[−.63, −.33]	.00
Learning environment			
Classroom vs. Unrestricted	−.48	[−.68, −.29]	.00
Classroom vs. Out-of-class	−.51	[−.65, −.38]	.00
Out-of-class vs. Unrestricted	−.03	[−.15, .20]	.62
Sample size			
Large vs. Small	−.08	[−.25, .08]	.17
Medium vs. Small	.21	[.06, .36]	.99
Large vs. Medium	−.29	[−.43, −.16]	.00
Mobile application			
Educational vs. General	.03	[.08, .15]	.69
General + Educational vs. General	.07	[−.06, .24]	.80
General + Educational vs. Educational	.04	[−.08, .18]	.70
Language proficiency			
Intermediate vs. Elementary	.83	[.61, 1.04]	1.00
Advanced vs. Elementary	.88	[.45, 1.32]	1.00
Intermediate vs. Advanced	−.05	[−.46, .35]	.40

Note. CI = confidence interval.

studies included advanced L2 learners; the remaining 37 studies failed to report participants' language proficiency level through standardized tests. MAVL was more efficient in intermediate learners ($d = 1.57$, 95% CI = [1.05, 2.08]), followed by advanced learners ($d = 1.31$, 95% CI = [.56, 1.91]) and elementary learners ($d = .79$, 95% CI = [−.10, 1.65]).

4.4. Publication bias

The publication bias analyses were first performed via a contour-enhanced funnel plot (Figure 5). The observed asymmetry, particularly the absence of studies in the lower left contour, could suggest the presence of publication bias or small-study effects. This pattern was further examined

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

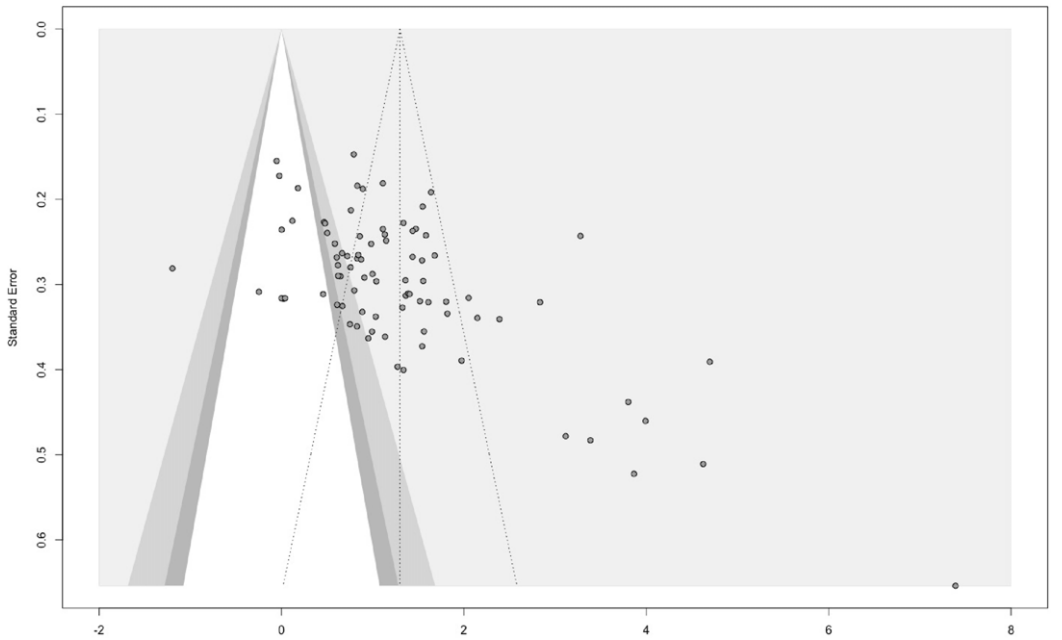


Figure 5. Contour-enhanced funnel plot of study-level standardized mean differences.

through Egger's regression test for overall vocabulary learning with mobile apps ($\beta = 6.56$, $t = 5.62$, $p < .001$). Subsequent trim-and-fill test results showed that 26 more studies were needed to nullify publication bias, and the overall adjusted effect size was .74 (95% CI = [.44, 1.03]). Thus, after adjusting, we still observed a medium effect size of mobile use in enhancing vocabulary learning. Moreover, the adjusted effect size was .87 for receptive vocabulary knowledge (95% CI = [.44, 1.30]), 1.46 for productive vocabulary knowledge (95% CI = [.57, 2.36]), and .68 (95% CI = [.30, 1.06]) for a combination of receptive and productive vocabulary knowledge.

5. Discussion

The present meta-analysis aggregated studies since 2010 and provided a comprehensive examination of the effectiveness of MAVL while considering a series of potential moderators. Overall, the findings were consistent with previous meta-analyses that MAVL was generally an effective approach for vocabulary learning, as shown by a large effect size. Specifically, learners who used mobile applications demonstrated a sustained advantage in long-term vocabulary retention compared to those who did not.

Moderator analyses showed the effectiveness of MAVL tended to vary across different aspects of vocabulary learning. First, mobile applications enhanced more vocabulary learning via a combination of form and meaning than vocabulary learning involving all three aspects. This was also corroborated by recent research on the order of acquisition of word knowledge components, as shown in the hierarchical acquisition of vocabulary aspects, starting with the form–meaning link and ending with the multiple-meaning (Sukying & Nontasee, 2022).

Next, secondary and tertiary learners showed more learning gains than primary school learners when using mobiles for vocabulary learning. Underlying factors for the lower learning gains by primary students included lower language learning motivation, limited language proficiency, and underdeveloped self-regulation capacity (Yu & Trainin, 2022). Third, applications for educational

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

purposes yielded a larger effect size in vocabulary learning than applications for general purposes (also see Chen *et al.*, 2020). Plausible reasons included the availability of interesting and diversified learning content (Roohani & Vincheh, 2023) and individualized learning systems in educational mobile applications (Song & Xiong, 2023). General applications for vocabulary learning could also present distracting factors such as online videos and irrelevant messages, which adversely affected the learning process and outcome.

Further, unrestricted environment and out-of-class settings contributed to more gains than classroom settings, which echoes Lin and Lin's (2019) findings. It seems that learners in the unrestricted and informal settings had more opportunities to explore and self-regulate their own learning experience after class (Criollo-C, Guerrero-Arias, Jaramillo-Alcázar & Luján-Mora, 2021), which contributed to increased learning gains. Fifth, mobile devices were likely to enhance more vocabulary gains for intermediate-level learners than elementary-level learners, which was consistent with Burston and Giannakou's (2022) findings. Nonetheless, it should be noted that only three studies involved elementary levels of learners, with no studies examining advanced learners. Future research is thus warranted to re-examine the results with varied levels of language learners.

Finally, mobile use on learning vocabulary was found to be more efficient in enhancing productive vocabulary knowledge than receptive vocabulary knowledge (Hao *et al.*, 2021). Previous linguistic studies have confirmed that grammatical functions and collocations were learned after form–meaning links were created (Sukying & Nontasee, 2022). However, mobile applications only provide limited or even no opportunities for acquisition of grammatical functions and collocations (Barjesteh, Movafaghardestani & Modaberi, 2022). The difficulty of the target word components and limited learning resources could explain why the combined vocabulary knowledge benefited the least from mobile application usage.

6. Conclusion and implications

This meta-analysis adds to existing MAVL literature by confirming a positive and large effect size of mobile applications on both vocabulary learning and delayed vocabulary retention. In other words, the use of mobile applications for vocabulary learning was more effective than conventional methods with treatment duration of over 10 weeks. Among the eight potential moderators in this relationship, only aspects of vocabulary knowledge, mobile application type, and education level were significant moderators, suggesting that the effect of using mobile applications in improving vocabulary learning was independent of many individual and other attributes.

This study does, however, have some limitations. First, the contour-enhanced funnel plot and Egger's test signified the existence of publication bias, and this analysis was thus at the risk of the "file-drawer" effect in which studies with significant findings were more likely to be accepted than those with non-significant results (Dechartres, Atal, Riveros, Meerpohl & Ravaud, 2018). As a result, the computed effect size might be inflated. Second, we identified only two studies with primary school student samples; future research is recommended in this area. Third, separate scores for different aspects of vocabulary learning were not reported in some studies. This did not allow us to disentangle the effect of MAVL on independent aspects of vocabulary learning. Last, some moderators in our study (e.g. vocabulary aspect and vocabulary knowledge) tended to be interconnected and thus might lead to a confounding effect. However, the limited number of studies included in our analyses did not allow for further analysis to address this concern.

Despite these limitations, our findings have implications for future MAVL research and practice. First, although teenagers were found to be keener on using mobile applications, studies with this age group were surprisingly few. As vocabulary is a fundamental factor in learning a language, especially at an early age (Sitompul, 2020), further investigations to explore the

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

pedagogical potential of MAVL should be extended to various age groups. Furthermore, our study shows that using mobile applications designed for educational purposes produced a significantly stronger effect than those designed for general purposes when learning vocabulary. To realize mobile applications' maximum benefits, language teachers are encouraged to incorporate educational mobile applications into both in-class instructions and out-of-class learning, while granting sufficient autonomy. For designers, more advanced and integrated features need to be developed to foster greater opportunities to strengthen learners' vocabulary learning.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344025100335

Data availability statement. Data available within the supplementary materials. Data also openly available in a public repository, with a permanent identifier (https://doi.org/10.7910/DVN/RMVEUS)

Authorship contribution statement. Yonghong Zhou: Writing – original draft, Methodology, Investigation, Visualization, Formal analysis, Data curation; Mingming Zhou: Conceptualization, Writing – review & editing, Methodology, Formal analysis, Supervision.

Funding disclosure statement. This research did not receive any specific funding.

Competing interests statement. The authors declare no competing interests.

Ethical statement. Ethical approval was not required.

GenAI use disclosure statement. The authors declare no use of generative AI.

References

Barjesteh, H., Movafaghardestani, E. & Modaberi, A. (2022) COVID-19's impact on digitalization of education: Incorporating visual vocabulary learning application to foster vocabulary knowledge. *Asian Education and Development Studies*, 11(1): 172–187. https://doi.org/10.1108/AEDS-05-2020-0111

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2): 97–111. https://doi.org/10.1002/jrsm.12

Burston, J. (n.d.) *General MALL bibliography 1994–2020*. Cyprus University of Technology. https://sites.google.com/site/eurocallmallsig/home

Burston, J. (2025) *A selected annotated bibliography of experimental MALL implementation studies 1999–2024. EUROCALL MALL SIG*. https://sites.google.com/site/eurocallmallsig/home

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

Chen, M.-P., Wang, L.-C., Zou, D., Lin, S.-Y. & Xie, H. (2019) Effects of caption and gender on junior high students' EFL learning from iMap-enhanced contextualized learning. *Computers & Education*, 140: Article 103602. https://doi.org/10.1016/j.compedu.2019.103602

Chen, Z., Chen, W., Jia, J. & An, H. (2020) The effects of using mobile devices on language learning: A meta-analysis. *Educational Technology Research & Development*, 68(4): 1769–1789. https://doi.org/10.1007/s11423-020-09801-5

Chen, Z., Jia, J. & Li, W. (2021) Learning curriculum vocabulary through mobile learning: Impact on vocabulary gains and automaticity. *International Journal of Mobile Learning and Organisation*, 15(2): 149–163. https://doi.org/10.1504/IJMLO.2021.114518

Council of Europe (2001) *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit. https://www.coe.int/en/web/common-european-framework-reference-languages

Criollo-C, S., Guerrero-Arias, A., Jaramillo-Alcázar, Á. & Luján-Mora, S. (2021) Mobile learning technologies for education: Benefits and pending issues. *Applied Sciences*, 11(9): Article 4111. https://doi.org/10.3390/app11094111

Dechartres, A., Atal, I., Riveros, C., Meerpohl, J. & Ravaud, P. (2018) Association between publication characteristics and treatment effect estimates: A meta-epidemiologic study. *Annals of Internal Medicine*, 169(6): 385–393. https://doi.org/10.7326/M18-1517

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

Gelman, A. (2009) Bayes, Jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*, 24(2): 176–178. https://doi.org/10.1214/09-STS284D

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

Hao, T., Wang, Z. & Ardasheva, Y. (2021) Technology-assisted vocabulary learning for EFL learners: A meta-analysis. *Journal of Research on Educational Effectiveness*, 14(3): 645–667. https://doi.org/10.1080/19345747.2021.1917028

Harrer, M., Cuijpers, P., Furukawa, T. A. & Ebert, D. D. (2021) *Doing meta-analysis with R: A hands-on guide*. Boca Raton: CRC Press. https://doi.org/10.1201/9781003107347

Hofstede, G., Hofstede, G. J. & Minkov, M. (2010) *Cultures and organizations: Software of the mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Hsu, L. (2013) English as a foreign language learners' perception of mobile assisted language learning: A cross-national study. *Computer Assisted Language Learning*, 26(3): 197–213. https://doi.org/10.1080/09588221.2011.649485

Hwang, G.-J. & Fu, Q.-K. (2019) Trends in the research design and application of mobile language learning: A review of 2007–2016 publications in selected SSCI journals. *Interactive Learning Environments*, 27(4): 567–581. https://doi.org/10.1080/10494820.2018.1486861

Lai, A. (2016) Mobile immersion: An experiment using mobile instant messenger to support second-language learning. *Interactive Learning Environments*, 24(2): 277–290. https://doi.org/10.1080/10494820.2015.1113706

Lam, K. K. L. & Zhou, M. (2022) Grit and academic achievement: A comparative cross-cultural meta-analysis. *Journal of Educational Psychology*, 114(3): 597–621. https://doi.org/10.1037/edu0000699

Lee, Y.-J. (2023) Language learning affordances of Instagram and TikTok. *Innovation in Language Learning and Teaching*, 17(2): 408–423. https://doi.org/10.1080/17501229.2022.2051517

Levin, B. & Hovav, M. R. (2017) Morphology and lexical semantics. In Spencer, A. & Zwicky, A. M. (eds.), *The handbook of morphology*. Hoboken: Wiley, 248–271. https://doi.org/10.1002/9781405166348.ch12

Li, F., Fan, S., Wang, Y. & Lu, J. (2021) Chinese university students' experience of WeChat-based English-language vocabulary learning. *Education Sciences*, 11(9): Article 554. https://doi.org/10.3390/educsci11090554

Li, J.-T. & Tong, F. (2020) The effect of cognitive vocabulary learning approaches on Chinese learners' compound word attainment, retention, and learning motivation. *Language Teaching Research*, 24(6): 834–854. https://doi.org/10.1177/1362168819829025

Li, R. (2023) Investigating effects of computer-mediated feedback on L2 vocabulary learning. *Computers & Education*, 198: Article 104763. https://doi.org/10.1016/j.compedu.2023.104763

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

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J. & Moher, D. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10): e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Lin, J.-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

Lye, L. (2022) Effect of texting on Spanish learners' vocabulary acquisition and motivation. *CALICO Journal*, 39(3): 326–349. https://doi.org/10.1558/cj.23992

Mahdi, H. S. (2018) Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research*, 56(1): 134–154. https://doi.org/10.1177/0735633117698826

Mayer, R. E. (ed.) (2005) *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816819

Milton, J. (2009) *Measuring second language vocabulary acquisition*. *Bristol: Multilingual Matters*. https://doi.org/10.21832/9781847692092

Nation, I. S. P. (2013) *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Norouzian, R., De Miranda, M. & Plonsky, L. (2019) A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal*, 103(1): 248–261. https://doi.org/10.1111/modl.12543

Oakes, L. & Howard, M. (2022) Learning French as a foreign language in a globalised world: An empirical critique of the L2 motivational self system. *International Journal of Bilingual Education and Bilingualism*, 25(1): 166–182. https://doi.org/10.1080/13670050.2019.1642847

Ou-Yang, F. C. & Wu, W.-C. V. (2017) Using mixed-modality vocabulary learning on mobile devices: Design and evaluation. *Journal of Educational Computing Research*, 54(8): 1043–1069. https://doi.org/10.1177/0735633116648170

Plonsky, L., Hu, Y., Sudina, E. & Oswald, F. L. (2023) Advancing meta-analytic methods in L2 research. In Mackey, A. & Gass, S. M. (eds.), *Current approaches in second language acquisition research: A practical guide*. Hoboken: Wiley, 304–333. https://doi.org/10.1002/9781394259670.ch14

Rachels, J. R. & Rockinson-Szapkiw, A. J. (2018) The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1–2): 72–89. https://doi.org/10.1080/09588221.2017.1382536

Roohani, A. & Vincheh, M. H. (2023) Effect of game-based, social media, and classroom-based instruction on the learning of phrasal verbs. *Computer Assisted Language Learning*, 36(3): 375–399. https://doi.org/10.1080/09588221.2021.1929325

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press

Roussel, S. & Galan, J.-P. (2018) Can clicker use support learning in a dual-focused second language German course? *Language Learning & Technology*, 22(3): 45–64. https://doi.org/10125/44656

Röver, C. (2020) Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, 93(6): 1–51. https://doi.org/10.18637/jss.v093.i06

Şad, S. N., Özer, N., Yakar, Ü. & Öztürk, F. (2022) Mobile or hostile? Using smartphones in learning English as a foreign language. *Computer Assisted Language Learning*, 35(5–6): 1031–1057. https://doi.org/10.1080/09588221.2020.1770292

Schmitt, N. & Schmitt, D. (2020) *Vocabulary in language teaching*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108569057

Sitompul, A. R. (2020) EFL young learners' vocabulary learning strategies: An overview. *Retain*, 8(2): 31–40.

Song, B. & Xiong, D. (2023) A comparative study of the effects of social media and language learning apps on learners' vocabulary performance. *Asia Pacific Education Review*. Advance online publication. https://doi.org/10.1007/s12564-023-09871-z

Sukying, A. & Nontasee, W. (2022) The acquisition order of vocabulary knowledge aspects in Thai EFL learners. *World Journal of English Language*, 12(5): 306–319. https://doi.org/10.5430/wjel.v12n5p306

Sun, Y. & Gao, F. (2020) An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research & Development*, 68(3): 1181–1198. https://doi.org/10.1007/s11423-019-09733-9

Thompson, C. G. & von Gillern, S. (2020) Video-game based instruction for vocabulary acquisition with English language learners: A Bayesian meta-analysis. *Educational Research Review*, 30: Article 100332. https://doi.org/10.1016/j.edurev.2020.100332

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G. & Wagenmakers, E.-J. (2021) The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3): 813–826. https://doi.org/10.3758/s13423-020-01798-5

Wu, Q. (2015) Pulling mobile assisted language learning (MALL) into the mainstream: MALL in broad practice. *PloS ONE*, 10(5): Article e0128762. https://doi.org/10.1371/journal.pone.0128762

Yap, W. L. (2016) Transforming conventional teaching classroom to learner-centred teaching classroom using multimedia-mediated learning module. *International Journal of Information and Education Technology*, 6(2): 105–112. https://doi.org/10.7763/IJIET.2016.V6.667

Yu, A. & Trainin, G. (2022) A meta-analysis examining technology-assisted L2 vocabulary learning. *ReCALL*, 34(2): 235–252. https://doi.org/10.1017/S0958344021000239

About the authors

Yonghong Zhou is a PhD student in the Faculty of Education at the University of Macau. Her research interests focus on second language acquisition, educational technology, and emotion regulation.

Mingming Zhou is a professor in the Faculty of Education at the University of Macau. Her research interests focus on educational technology, emotion, and positive psychology.

https://doi.org/10.1017/S0958344025100335 Published online by Cambridge University Press