

An Overview of Mobile Language Learning in Selected Review Studies: 2010-2018¹

Chaka Chaka², University of South Africa, Pretoria, Gauteng, South Africa

Abstract

There has been a growing number of reviews of mobile language learning (MLL) conducted on various aspects of this field of study over the last two decades. However, unlike in other fields, there is a paucity of overviews of reviews that assess and integrate findings from multiple reviews, and that compare and contrast such findings, or establish gaps in existing reviews. The current overview examined 17 review studies published between 2010 and 2018, and investigated twenty-three characteristics extracted from these reviews with a view to comparing them and synthesising the main findings of these reviews. The selection criteria used were: reviews published between 2010 and June 2018, reviews published in peer-reviewed journals, use of at least one bibliographic database, reviews focusing on mobile language learning (MLL) or having MLL as one of the learning domains, and reviews published in English. Some of the findings of this overview are: most reviews were qualitative in nature and few of them had clearly and precisely formulated purposes; there is a lack of theoretical frameworks by the majority of these review studies; there were 25 precise and specific mobile device typologies reported by eleven reviews, of which personal digital assistants (PDAs) and mobile phones were the two most reported mobile device types, respectively; language was the most reported subject area followed by mathematics; and fewer reviews mentioned and specified their research designs, their coding procedures, their inter-rater reliability agreement, and their quality assessment. Finally, English as a foreign language (EFL) was the major target language and vocabulary learning was the main focus of MLL.

Resumen

Un número creciente de revisiones del aprendizaje de idiomas móvil (MLL) se ha realizado sobre varios aspectos de este campo de estudio durante las últimas dos décadas. Sin embargo, a diferencia de otros campos, hay una escasez de resúmenes de revisiones que evalúan e integran hallazgos de múltiples revisiones, y que comparan y contrastan tales hallazgos, o establezcan brechas en revisiones existentes. El resumen actual examinó 17 estudios de revisión publicados entre 2010 y 2018, e investigó veintitrés características extraídas de estas revisiones con el fin de compararlas y sintetizar los principales hallazgos de estas revisiones. Los criterios de selección utilizados fueron: revisiones publicadas entre 2010 y junio de 2018, revisiones publicadas en revistas revisadas por pares, uso de al menos una base de datos bibliográfica, revisiones centradas en el aprendizaje de idiomas móvil (MLL) o tener MLL como uno de los dominios de aprendizaje, y revisiones publicadas en inglés. Algunas de las conclusiones de este resumen son: la mayoría de las revisiones fueron de naturaleza cualitativa y pocas de ellas tenían propósitos formulados de manera clara y precisa; faltan marcos teóricos en la mayoría de estos estudios de revisión; hubo 25 tipologías de dispositivos móviles precisas y específicas reportadas por once revisiones, de las cuales los asistentes digitales personales (PDA) y los teléfonos móviles fueron los dos tipos de dispositivos móviles más reportados, respectivamente; el lenguaje fue el área temática más reportada seguida por las matemáticas; y menos revisiones mencionaron y especificaron sus diseños de investigación, sus procedimientos de codificación, su acuerdo de confiabilidad entre evaluadores y su evaluación de la calidad. Finalmente, el inglés como lengua extranjera (EFL) fue el principal idioma de destino y el aprendizaje de vocabulario fue el enfoque principal de MLL.

Introduction

The last decade has witnessed a progressive increase in reviews of mobile learning in various fields, and especially in the field of mobile language learning (MLL). In this regard, mobile learning is used here to refer to any form of learning mediated through mobile devices (Rikala, 2013; Persson & Nouri, 2018). As such, mobile language learning (hereafter MLL) is intended as a generic reference to mobile-mediated language learning in English language, English as a second language (ESL) and English as a foreign language (EFL). This generic sense also subsumes mobile-assisted language learning (MALL). Such reference differentiates MLL from mobile learning focusing generally on other subject areas. This means that the current overview focused on reviews, literature reviews, systematic reviews and meta-analyses of MLL, and on those reviews dealing with other subject areas, but which had language learning (e.g., English language, ESL or ESL) as one of their key subject domains. However, in the present study, the term *review study* (RS) – and in certain instances, *review* – is used as a convenient generic term to collectively refer to all articles reviewed in the study itself. Appropriate and relevant terminological differentiators (e.g., reviews, literature reviews, systematic reviews and meta-analyses) will be used whenever there is a need to do so.

Mostly, review studies have been conducted as either literature reviews, reviews, systematic reviews, meta-analyses, or research syntheses (Evans & Popova, 2015; Grant & Booth, 2009; Uddin & Arafat, 2016). Each of these review types has a different focus and a different purpose. For example, a literature review examines evidence or aspects related to published literature on a given topic or area. It uses one or more

¹ This is a refereed article. Received: 20 March, 2021. Accepted: 7 June, 2021.

² chakachaka8@gmail.com, 0000-0003-3558-4141

databases, and makes non-generalizable findings (Grant & Booth, 2009). A review is a generic review that offers a state-of-the-art analysis of recent literature on one or more areas. It provides a narrative summary of original articles. Owing to its selective nature, its findings are not generalizable (Uddin & Arafat, 2016; Paré et al., 2015). For its part, a systematic review is more rigorous than the first two types of reviews in that it has focused questions, and employs pre-specified eligibility criteria for its search strategies and for its analysis and appraisal of primary articles (Uddin & Arafat, 2016; Grant & Booth, 2009; Moher et al., 2009). A meta-analysis also uses pre-specified eligibility criteria for its search strategies, but, in addition, it applies statistical techniques to analyse and summarize results from several studies (Grant & Booth, 2009; Moher et al., 2009; Paré et al., 2015; Xiao & Watson, 2019). Lastly, research synthesis, also known as thematic synthesis, extracts themes from literature or from original articles, clusters them, and synthesizes them into analytical themes (Xiao & Watson, 2019).

For instance, Haßler et al.'s (2016) and Hwang and Tsai's (2011) literature reviews report on mobile learning studies conducted on multiple subject disciplines. With specific reference to MLL, review studies have been conducted by researchers such as Duman et al. (2015), Kukulska-Hulme and Viberg (2018), Shadiev et al. (2017), and Viberg and Grönlund (2013). Of these reviews, Kukulska-Hulme and Viberg's (2018) have an exclusive focus on mobile collaborative language learning. Reviews that have language learning as one of their subject domains include those conducted by Baran (2014), Frohberg et al. (2009), Hwang and Tsai (2011), Korkmaz (2015), and Sönmez et al. (2018). On the other hand, authorities like Crompton et al. (2017), Krull and Duarte (2017), and Pimmer et al. (2016) have carried out systematic reviews on mobile learning in which language learning was one of the key subject domains. In this case, meta-analyses with language learning components have been conducted by scholars such as Chee et al. (2017), Mahdi (2017), Sung et al. (2016), Tingir et al. (2017), and Wu et al. (2012), while Taj et al.'s (2016) meta-analysis focuses exclusively on MLL within an EFL context. Some of these meta-analytic studies such as Sung et al.'s (2016) comprises meta-analysis and critical synthesis.

Contextualising Issues

While reviews, systematic reviews, and meta-analyses have been conducted on MLL as delineated above, there is currently a paucity of overviews of reviews conducted in the field of MLL in contrast to other disciplines or areas, such as: education (Polanin et al., 2016); psychology (Cooper & Koenka, 2012); technology and learning (Tamim et al., 2011); computer-assisted language learning and second language acquisition interface (Plonsky & Ziegler, 2016); and as opposed to other areas of second language learning like literacy learning (Torgerson, 2007). In an overview of reviews, characteristics of review studies become key foci or units of analysis and not characteristics or aspects related to primary studies (see Kim et al., 2018; Polanin et al., 2016). Various names are used to refer to an overview of reviews. Among these are a tertiary review, a second-order review, a review of reviews, an umbrella review, a systematic review of systematic reviews, a meta-meta-analysis, or a synthesis of meta-analysis (Grant & Booth, 2009; Hunt et al., 2018; Kim et al., 2018; Pieper et al., 2012; Polanin et al., 2016; Tight, 2018; Uddin & Arafat, 2016). In this context, Pieper et al. (2012) argue that there is no universal definition of an overview and that the concept itself is often not definitively articulated whenever it is used.

There are benefits of conducting overviews. Some of these benefits are the following: a) identifying, assessing and integrating findings from multiple review studies; b) tapping into past research syntheses; c) broadening evidence synthesis questions that cannot be explored through standard reviews (McKenzie & Brennan, 2017), or framing research problems in broader parameters; d) aggregating the results of several reviews or contrasting multiple results on the same topic (Pieper et al., 2012; Polanin et al., 2016); e) mapping trends and changes in research over time; f) enhancing the knowledge base confined to current reviews; and g) exploring the possibility of up-dating existing reviews (Polanin et al., 2016), or establishing gaps in existing reviews.

Nonetheless, there are pitfalls associated with overviews. Some of these pitfalls are bias, out-datedness, overlap and lack of methodological rigour. Bias is related to biased reporting, which may involve a bias towards certain review studies. It may also entail over-reporting, under-reporting, or inconsistent reporting of certain aspects at the expense of others. Out-datedness has to do with review studies that are not up-to-date, or with the preference given to outdated review studies over the recent ones (Polanin et al., 2016). Overlapping occurs when two or more similar studies feature in one or more reviews, while a lack of methodological rigour has to do with the absence of methodological standards and reporting guidelines (Pieper et al., 2012).

Theoretical Framework

This overview employed elements of Dennen and Hao’s (2014) M-COPE (mobile, conditions, outcomes, pedagogy and ethics) and elements of Chaka (2015) mobile presence learning (see Chaka et al., 2020) as its theoretical framing. For instance, Dennen and Hao (2014) differentiate between learning tasks that are mobile-dependent and those that are mobile-enhanced. For its part, mobile presence learning is presence learning enabled by mobile devices (Chaka et al., 2020). In this overview, these two theoretical lenses were employed as a composite theoretical framing.

Method

As Hunt et al. (2018) contend, overview studies are, comparatively, a new and emerging tool of summarising and synthesising findings from multiple review studies. As such, there is not yet generically acceptable and established guidelines for conducting them. Owing to this lack of universal guidelines, overviews depend on some of the currently used review methods such as a systematic literature review (SLR) approach (Abu Saa et al., 2019; Baillie, 2015; Okoli, 2015) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) that are often applied in systematic reviews and meta-analyses. This overview employed these two approaches as will become clear below.

With reference to a search and identification protocol for review studies, SLR has many permutations of search strategies (Abu Saa et al., 2019; Cooper et al., 2018; Kim et al., 2018; Okoli, 2015; Ramírez-Montoya & Lugo-Ocando, 2020). Based on this, the ensuing four main stages, with their attendant steps, as modified from Okoli (2015), were followed: planning (identifying the purpose and formulating research questions); selection (identifying keywords, identifying databases, inclusion/exclusion criteria, and searching for and selecting studies); extraction (assessing the quality of studies and data extraction strategy); and execution (data analysis and data synthesis).

Planning

Identifying the Purpose of the Overview Study

As stated earlier, currently, there is a paucity of overview studies conducted in the field of MLL in general, despite the fact that there are many review studies that have been carried out in this field over the last two decades. As such, the current overview is intended to fill this gap and contribute to the knowledge generated from aggregating the foci and the findings of multiple reviews on MLL published from 2010 to 2018. On this basis, the present overview has a dual purpose: To compare twenty-three different characteristics investigated by 17 review studies reviewed in this overview and to synthesize the main findings of these 17 review studies. The twenty-three key characteristics that were compared are as shown in Table 1 and as phrased in the research questions below. These characteristics are the key elements extracted from the 17 reviews. They include the type of review, the number of databases each review used, a research design of each review, etc. (see Kim et al., 2018; Lin & Lin, 2019; Ramírez-Montoya, 2020; Ramírez-Montoya & Lugo-Ocando, 2020). They are specific features assessed in each review article.

Twenty-three key aspects investigated in each review study	
Coverage period	Research design
Review type	Sample group(s)
Types of reviewed publications	Data collection method(s)
Number of publications included	Coding procedure and data analysis
Databases searched	Inter-rate reliability (agreement)
Purpose	Quality assessment
Reported types of mobile device(s) used	Publication bias
Educational level(s)	Moderator variables
Educational context(s)	Effect size
Learning domains of the reviewed articles	Main findings
Theoretical framework	

Table 1: Twenty-three key aspects of the review studies investigated by the current overview study

Research Questions

The following served as research questions (RQs) for this study:

RQ 1: What types of review studies has the overview identified?

RQ 2: What are the research purposes and theoretical frameworks of the review studies?

RQ 3: What coverage periods, what number of databases and articles, what number of mobile device types, what educational levels and educational contexts, and how many subject areas do these review studies focus on?

RQ 4: What mobile device types and what types of subject areas are reported by review studies?

RQ 5: What are the research designs, sample groups and data collection methods of these review studies?

RQ 6: What is the coding procedure, data analysis, inclusion/exclusion criteria, inter-rater reliability agreement, and quality assessment of these review studies?

RQ 7: What are the publication biases, moderator variables and effect sizes (where applicable) of the review studies?

RQ 8: What are the main findings of these review studies?

Selection

Identifying Keywords

The search keywords selected were formulated based on some of the research questions and the purpose stated in the preceding sub-sections. In particular, the following keywords were created: review; mobile language learning; mobile devices; theoretical framework; and 2010-2018. Search strings involving these keywords were then created and combined with the two Boolean operators, AND and OR. Where necessary, parentheses and double quotation marks were used in the search strategy. Two examples of such search strings were as follows:

- review AND mobile language learning AND mobile devices AND theoretical framework AND 2010-2018.
- (review) AND (mobile language learning) AND (mobile devices) AND (theoretical framework) AND (2010-2018).

Different permutations of these search strings were employed with *review*, *mobile language learning* and *mobile devices* replaced with their relevant synonyms, in certain cases.

Identifying Databases

Eight online databases, together with two online academic social platforms, were identified for conducting the search strategy. This was done by querying the search strings mentioned above and their related permutations. The eight online databases comprised: Google Search; Google Scholar; Semantic Scholar; Education Resources Information Center (ERIC); SpringerLink; ScienceDirect; Taylor & Francis Online; and Wiley Online Library. They were complemented by ResearchGate and Academia.edu, which are online academic social networking sites (see Figure 1).

Inclusion/Exclusion Criteria, and Searching for and Selecting Studies

Inclusion/Exclusion criteria were formulated to determine the eligibility of review studies. Two of the inclusion criteria were that reviews had to focus on mobile language learning and needed to have been published between 2010 and June 2018 (see Table 2).

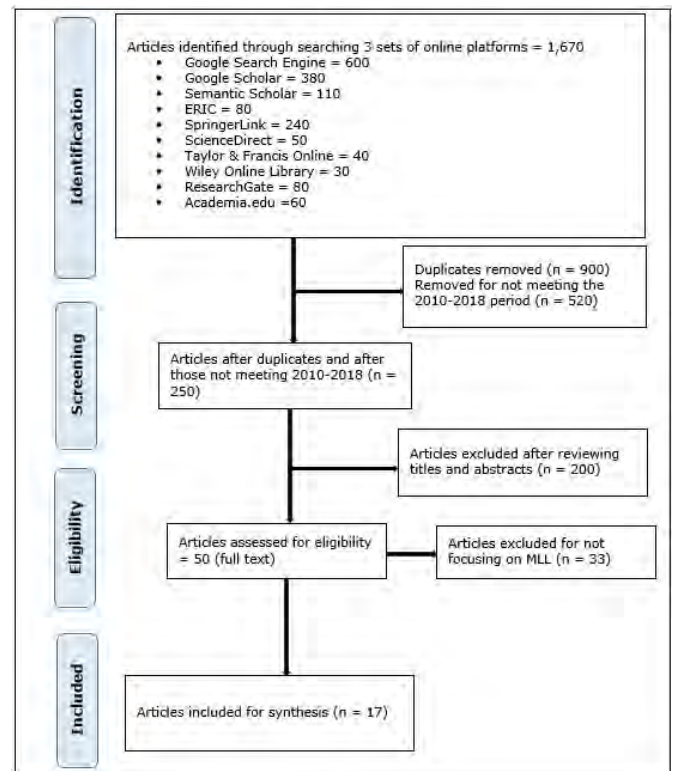


Figure 1. PRISMA flowchart for screening articles

Criteria	Inclusion	Exclusion
Time period	Reviews published between 2010 and June 2018	Articles not published between 2010 and 2018
Types of articles	Reviews published in peer-reviewed journals	Reviews not published in peer-reviewed journals
Databases	Use of at least one bibliographic database	No use of any bibliographic database
Content and focus of articles	Reviews focusing on mobile language learning (MLL) or having MLL as one of the learning domains	Reviews not focusing on mobile language learning or not having MLL as one of the learning domains. Reviews focusing purely on non-content aspects such as perceptions, attitudes, usability, etc.
Language of publication	Reviews published in English	Reviews not published in English

Table 1: Inclusion/Exclusion criteria

Eligible studies were identified through searching and screening between July 2018 and January 2019. Sets of search strings – and their attendant permutations - mentioned earlier, were iteratively queried into the ten online databases identified above. Numerous searches conducted on these databases yielded 1,550 articles (see Figure 1). These resultant articles were screened after which, 900 articles were removed as duplicates and 520 articles were eliminated for not meeting the 2010-2018 coverage period. Of the remaining 250 articles, 200 were excluded after reviewing their titles and their abstracts. A full-text screening of the 50 remaining articles was done. Both forward snowballing and backward snowballing³ were conducted on these articles (Jalali & Wohlin, 2012; Wohlin, 2014). This dual snowballing search yielded 6 more articles. Of the 56 articles, 39 articles were excluded for not focusing on MLL. After this screening process, 17 articles (see Figure 2) were selected as the units of analysis for the current overview.

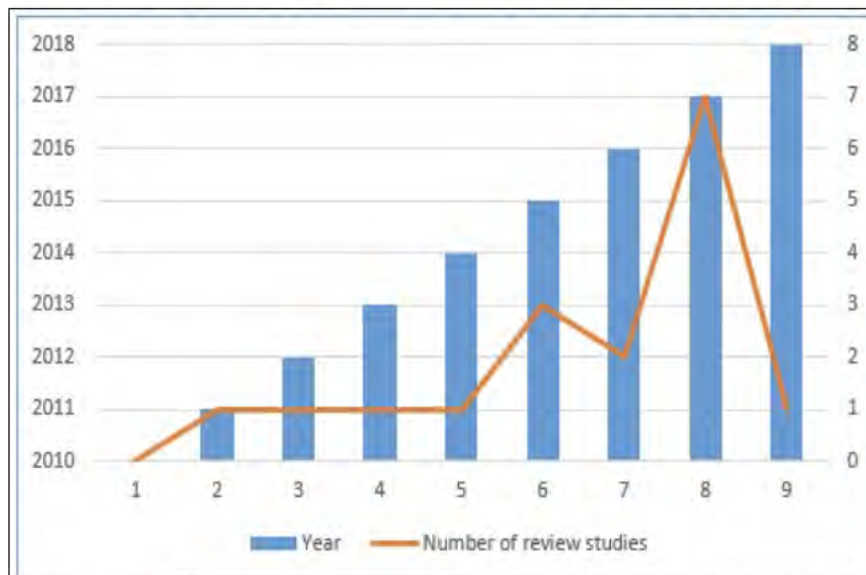


Figure 2. Number of review studies and their years of publication

Extraction

Assessing the Quality of Studies and Data Extraction Strategy

To ensure that the included reviews met a minimum threshold of methodological quality, the reviews were assessed by employing a combined set of criteria derived from the Assessment of Multiple Systematic Reviews tool (AMSTAR) (Shea et al., 2009; see Bobrovitz et al., 2015) and from the quality assessment questions formulated by Kitchenham and Charters (2007) and Kitchenham et al. (2009; see Abu Saa et al. 2019). In all, eighteen combined criteria were created (see Table 3).

³ Forward snowballing refers to locating relevant articles that cite the articles discovered during the search process, and backward snowballing refers to locating relevant articles from reference lists of given journals (Jalali & Wohlin, 2012).

Quality assessment questions
1. Was an "a priori" design offered?
2. Was there duplicate publication selection and data extraction?
3. Were three or more electronic databases searched?
4. Was a comprehensive literature search conducted?
5. Were the review aims/purposes clearly stated?
6. Was the status of publication (e.g., peer-reviewed or grey literature) employed as an inclusion criterion?
7. Was a list of reviewed publications provided?
8. Were the characteristics of the included publications provided?
9. Was the scientific quality of the included publications assessed and reported?
10. Is the sample size (e.g., the number of included publications) justified?
11. How clearly defined is the sample design/target selection of cases/publications?
12. Was the possibility of publication bias reported and assessed?
13. Does the review explore diverse perspectives and contexts?
14. How clearly defined are the assumptions/theoretical perspectives/values that have shaped the form and output of the publication/research?
15. Were the methods used to integrate the findings of studies appropriate?
16. Are the findings credible?
17. What contribution do the findings make to the existing body of knowledge?
18. Was the scientific quality of the included studies used appropriately to formulate conclusions?

Table 3: Quality assessment questions (Bobrovitz et al., 2015; Shea et al., 2009)

Two coders autonomously assessed the quality of each review study using these combined criteria. Each of the 18 criteria were allotted a "yes", "no/partially" or "unclear" rating. On this basis, the overall agreement was calculated using Cohen's kappa (κ) values. Disagreements in rating these criteria was resolved by discussion and consensus between coders (see Belur et al., 2018; Bobrovitz et al., 2015; Cohen, 1960; Landis & Koch, 1977; McHugh, 2012; Mengist et al., 2020; Schlosser et al., 2007; Sim & Wright, 2005). The inter-coder agreement, which is the extent of independence coders display in evaluating a feature of an item and in reaching the same conclusion (Lombard et al. 2002), was .86. It was based on Landis and Koch's (1977) κ scoring and interpretation: <0 = poor; 0.00-0.20 = slight; 0.21-0.40 = fair; 0.41-0.60 = moderate; 0.61-0.80 = substantial = 0.81-1.00 = near perfect. Thus, according to this κ scoring, 0.86 is deemed to be near perfect.

Data were extracted from the 17 reviews in keeping with the aspects featuring in Table 1. The information related to each of these aspects was extracted as provided from each review.

Execution

Data Analysis and Data Synthesis

The extracted data were analysed using thematic analysis and quantitative content analysis. Categories were developed and themes related to these categories were formulated (see Table 1). In addition, content items from the review content were matched to their corresponding categories as displayed in Table 1. These data sets, represented through themes and content items, were compared and synthesised in response to the research questions of the present overview.

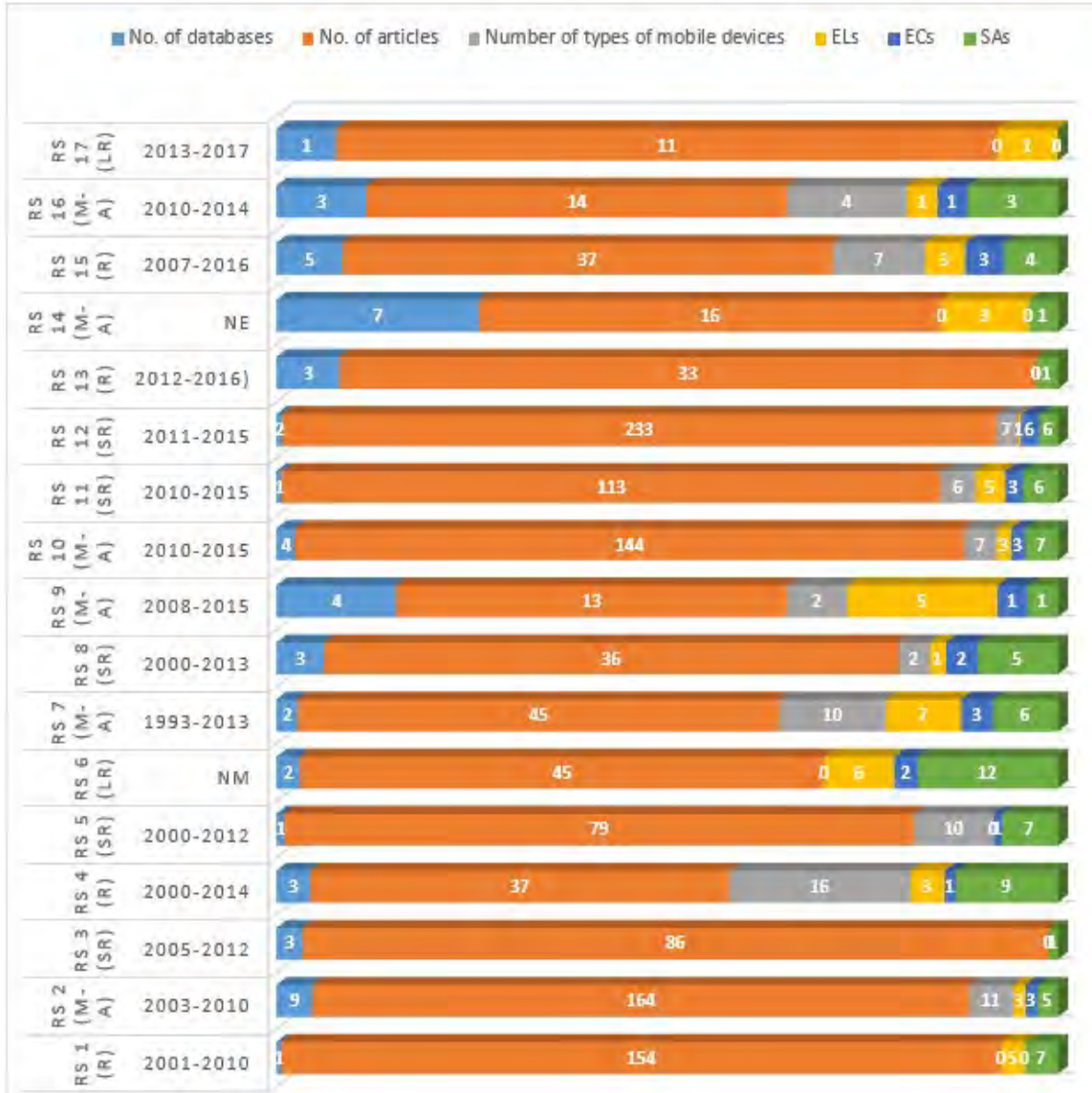
Findings

The findings presented in this section of the overview are grounded on the data sets extracted from the 17 review articles and are guided by the manner in which the data sets were codified and analysed as mentioned

in the preceding section. In addition, these findings are in response to the eight research questions (RQs) formulated above.

Types of review studies identified, and research purposes and theoretical frameworks of review studies

As illustrated in Figure 3, the 17 reviewed articles consisted of the following typologies: four reviews; two literature reviews; five systematic reviews; and six meta-analyses. Collectively, these review studies were published between 2010 and 2018 as shown in Figure 2.



Note: LR = literature review; M-A = meta-analysis; R = review; SR = systematic review; NE = not explicit; NM = not mentioned

Figure 3: Types of review studies, coverage periods, number of databases, number of articles, mobile device types, number of educational levels (ELs), number of educational contexts (ECs) and subject areas (SAs)

Pertaining to purposes, seven review studies had clear and specific purposes, and two had clear purposes (see Table 4). By contrast, the remaining eight review studies had vague purposes. Even though the 17 studies had diverse purposes, ten of them had overlapping purposes. In relation to theoretical frameworks, only two review studies (RS 4 and RS 7) employed theoretical frameworks. These were mobile learning and theoretical perspectives and the activity-theory based framework for MALL. The other fifteen review studies did not mention nor did they employ any theoretical framework.

Review study	Purpose(s)
RS 1	To examine the mobile or ubiquitous learning papers published in the Social Science Citation Index (SSCI) database from 2001 to 2010. (V)
RS 2	To systematically review and synthesise the relevant literature through a meta-analysis so as to offer a more comprehensive analysis of previous studies. (V)
RS 3	To present an overview of contemporary MALL research. (C)
RS 4	To address trends and gaps identified in the literature regarding the integration of mobile learning into teacher education. (C&S)
RS 5	To analyse MALL studies published from the years 2000-2012 in order to investigate the characteristics and general research trends, and to demonstrate to what extent mobile devices are being employed to support language learning. (C&S)
RS 6	To examine the research in the literature within the last three years, and to identify new approaches. (V)
RS 7	(a) to present an overview of the use of mobile devices in language learning, together with the target population, types of hardware and software, teaching/learning methods, the settings in which the teaching/learning, language skills were implemented, target languages, first language (L1) vs. second/foreign language (L2), and duration of the intervention; (b) to determine the overall effectiveness of employing mobile technology in education on students' language achievement; and (c) to examine whether moderator variables influence the effects of mobile devices on language learning. (C&S)
RS 8	To address the following broadly focused research question: How can mobile ubiquitous learning formats in higher education be synthesised according to their theoretical underpinnings and to what extent are these categories tied to different educational outcomes? (V)
RS 9	Not explicitly stated (To synthesize the lesson learnt so far in the field of MALL). (V)
RS 10	To provide insights into the research trends and issues in the studies of m-learning through content analysis of selected journals from January 2010 to December 2015. (C)
RS 11	To provide a current synthesis of mobile learning research from 2010-2015 in PK-12 education. (V)
RS 12	To analyse the research topics, methods, settings, and technologies used in mobile learning research in higher education, and published from January 2011 to December 2015. (C&S)
RS 13	To improve understanding of how mobile technologies and applications (both generic and language-specific) have been used to support CL among second and foreign language learners. (C&S)
RS 14	To investigate and synthesise research on the overall effectiveness of mobile devices on vocabulary learning. (C&S)
RS 15	To understand publications' trend, research focus, technology used, methodology, and current issues. (V)
RS 16	Not explicitly stated (Investigating the effects of mobile devices on student achievement in science, mathematics and reading in grades K-12). (C&S)
RS 17	To summarize the research findings in the literature by employing content analysis. (V)

Note: V = vague; C = clear; C&S = clear and specific

Table 4: Purposes of review studies

Coverage periods, number of databases, articles and mobile device types, educational levels and educational contexts, and subject areas

Two studies (RS 6 and RS 14) did not specify their coverage periods (years of coverage) (see Figure 3). The remaining fifteen studies that specified their coverage periods had these coverage periods varying considerably. For instance, all but one study (RS 7, 1993-2013) had their coverage periods from or after 2000. Of these, three studies (RS 4, RS 8 and RS 5) had the longest coverage periods spanning 14, 13 and 12 years, respectively. But, overall, RS 7 had the longest coverage period that traversed 20 years. The shortest coverage period was 4 years and was shared by four review studies: RS 12, RS 13, RS 16 and RS 17.

Four studies employed one database each to conduct their searches, while three studies utilised two databases each (see Figure 3). The rest of the other studies employed more than two databases. Of these, RS 2 and RS 14 employed the most databases, with nine and seven databases, respectively. Four studies had fewer included articles, of which RS 17 (n =11) and RS 9 (n = 13) had the least included articles. The study with the most included articles was RS 12 (n = 233). It is followed by RS 2 (n = 164) and RS 1 (n = 154).

Of the 17 review studies, six did not mention nor report any number of the mobile devices of the publications they had reviewed. Two studies (RS 8 and RS 9) reported only two types of mobile devices of the publications they had included. The study that reported the highest number of the types of mobile devices used by its publications is RS 4 (n = 16). It is followed by RS 2 (n = 11), and RS 5 (n = 10) and RS 7 (n =10). With respect to educational levels, three studies reported no educational levels of their reviewed publications. As opposed to this, three studies reported one educational level each, while five studies reported three educational levels. Contrarily, five studies reported more educational levels, with two of them (RS 7, n = 7 and RS 6, n = 6) reporting the most educational levels.

In regards to educational contexts, five studies did not mention any educational context of their reviewed publications, while one study mentioned the most educational contexts (n = 6) related to its reviewed publications. Lastly, two studies reported the most subject areas (n = 12 and n = 9) from their reviewed articles, whereas one study did not report any subject area from its reviewed articles (see Figure 3).

Mobile device types and types of subject areas reported in each review study

As illustrated in Figure 4, barring the six studies that did not report any mobile devices of their reviewed publications, the eleven other studies collectively reported 34 mobile technologies. However, only 25 of these technologies had precise and specific typologies, with 2 of these technologies (cell phones and iPhones) subsumed under either mobile phones or smartphones. Five of the 34 reported technologies had imprecise typologies, while the remaining four were a combination of two (n = 3) and three (n = 1) of the specified 25 typologies.

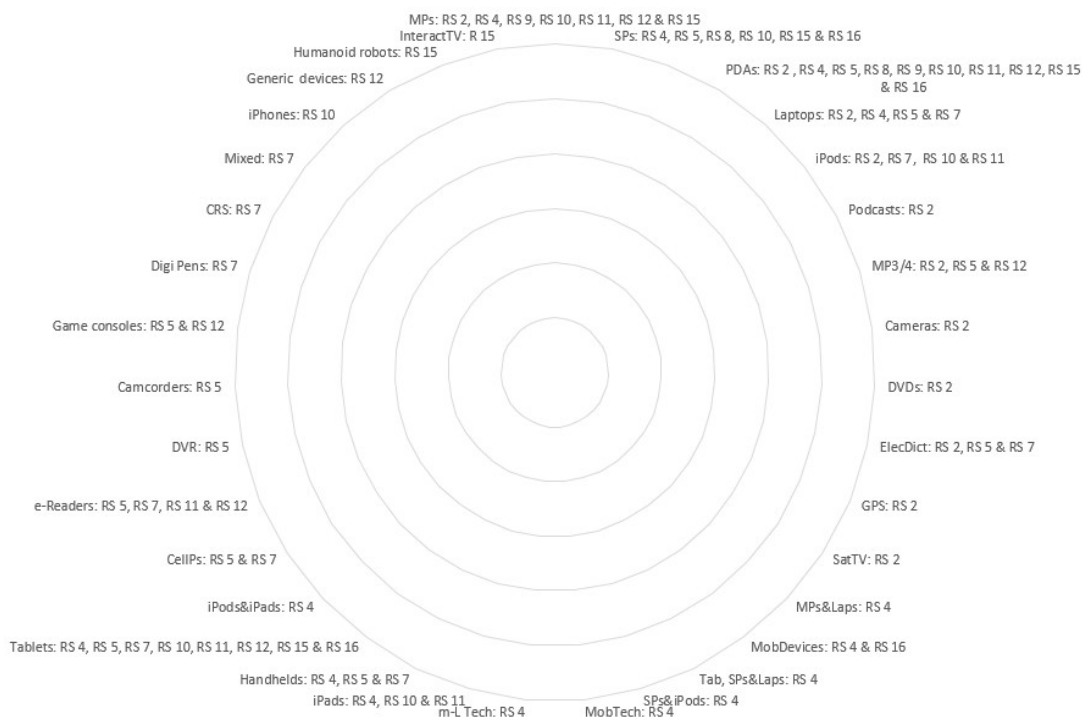


Figure 4. Review studies and different mobile device types

Note: RS = review study; MPs = mobile phones; InteractTV = interactive television; SPs = smart phones; PDAs = personal digital assistants; CRS = full name not provided in the original study text; Digi Pens = digital pens; DVD = digital video disc; DVR = digital video recorder; ElecDict= electronic dictionary; CellIPs – cell phones; GPS = The Global Positioning System; m-L Tech = mobile learning technology; MobDevices = mobile devices; MobTech = mobile technology; Laps = laptops; Tab = tablet; SatTV= satellite television

For the types of subject areas reported, 38 different subject areas were mentioned by sixteen of the seventeen studies that reported subject areas of their reviewed publications (see Figure 5). Nonetheless, four of these subject areas were academic disciplines, whereas three of them were composite subject areas. Natural science was reported by seven studies and is followed by mathematics and social sciences that were mentioned by six and five studies, respectively. But overall, language as a subject area featured in thirteen studies in different permutations or sub-areas.

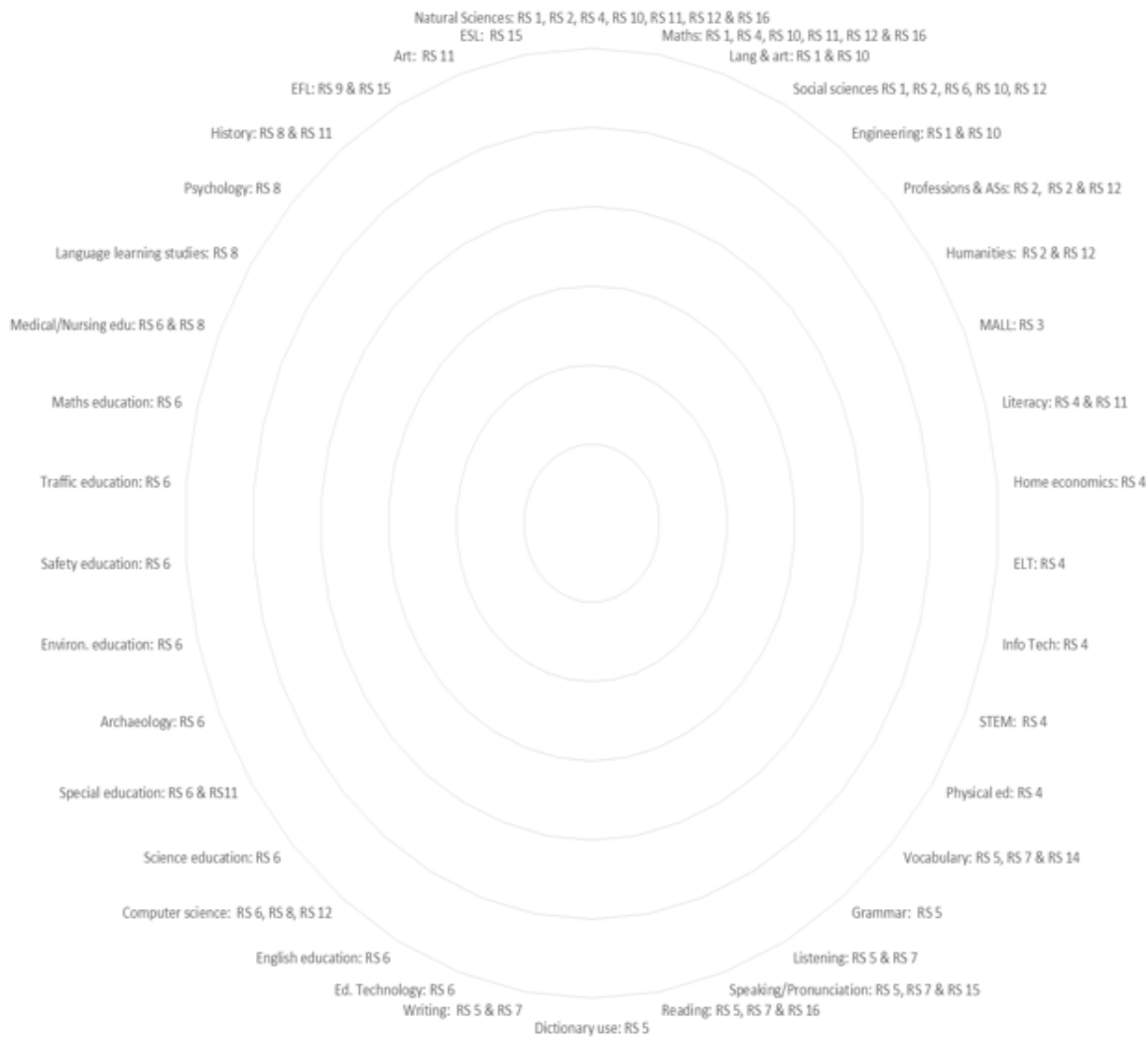


Figure 5: Review studies and subject areas

Note: RS = review study; SSs = social sciences; Professions & ASs = professions and applied sciences; ELT = English language teaching; Info Tech = information technology; STEM = science, technology, engineering and math; Physical ed = physical education; EFL = English as a foreign language; ESL = English as a second language

Research Designs, sample groups and data collection methods of the review studies

Only eight of the 17 review studies mentioned or specified their research designs, with nine of them not mentioning any research design at all (see Table 5). Of these eight studies, two mentioned a document analysis method and the PRISMA principles as their research design. All of the review studies did not specify their sample groups and their data collection methods.

Review study	Research design	Sample group(s)	Data collection method(s)
RS 1	Not mentioned	154 articles out of initial 3,995 papers (But not stated as such).	Database employed (But not mentioned as a data collection method).
RS 2	Systematic review	177 participants = reviewed studies; 164 studies out of initial 887 results (But not stated as such).	Data sources and search strategies (But not mentioned as a data collection method).
RS 3	Not mentioned	86 articles out of initial 152 papers (But not stated as such).	Databases employed (But not mentioned as a data collection method).
RS 4	Not mentioned	37 articles out of initial 329 articles (But not stated as such).	Databases employed (But not mentioned as a data collection method).
RS 5	Systematic review	79 articles (Initial number not mentioned) (But not stated as such).	Database employed.
RS 6	Document analysis method	45 articles out of initial 87 studies (But not mentioned as a sample group).	Databases employed (But not mentioned as a data collection method).
RS 7	Not mentioned	9,154 participants = reviewed studies; 45 papers (43 articles and 2 doctoral dissertations) out of initial 1,076 publications (721 articles, 287 conference papers and 68 doctoral dissertations) (But not stated as such).	Databases employed (But not mentioned as a data collection method).
RS 8	Interpretative design	36 papers - precise initial numbers not provided (But not stated as such).	Databases employed (But not mentioned as a data collection method).
RS 9	Not mentioned	13 studies out of initial 25 papers (But not mentioned as a sample group). (30-49 = 7 studies; 50-100 = 5 studies; and 100-200 = 1 study); NB: Total sample number for the 13 studies = 813; Average sample (n = 62.5 (63)).	Databases employed (But not mentioned as a data collection method).
RS 10	Not mentioned	144 articles out of initial 162 articles (But not mentioned as a sample group). NB: higher education students (50.75%); elementary or primary school students (19.40%); elementary or primary school teachers (13.43%); high or secondary school students (7.46%); high education instructors (1.49%).	Databases employed (But not mentioned as a data collection method).
RS 11	Followed the PRISMA principles.	113 articles out of initial 321 articles (But not mentioned as a sample group).	Database employed (But not mentioned as a data collection method).
RS 12	Systematic review	233 articles out of initial 2,727 articles (But not mentioned as a sample group).	Databases employed (But not mentioned as a data collection method).
RS 13	Literature review	33 articles out of initial 1,241 articles (But not mentioned as a sample group).	Databases employed (But not mentioned as a data collection method).
RS 14	Not mentioned	16 articles out of initial 3,008 titles (But not mentioned as a sample group). NB: Sample size of reported articles not explicit enough (e.g., experiment: 1-50 (n = 15); control: 1-50 (n = 15)).	Databases employed (But not mentioned as a data collection method).

RS 15	Not mentioned	37 articles – initial number of studies not provided (But not mentioned as a sample group).	Databases employed (But not mentioned as a data collection method).
RS 16	Meta-analysis	Total number of participants not provided; 14 studies out of initial 1,044 studies (But not mentioned as a sample group).	Databases employed (But not mentioned as a data collection method).
RS 17	Not mentioned	11 articles – initial number of studies not provided (But not mentioned as a sample group).	Database employed (But not mentioned as a data collection method).

Table 5: Review studies, research designs, sample groups and data collection methods

Coding procedure, data analysis, inclusion/exclusion criteria, Inter-rater reliability agreement, quality assessment, publication bias, and moderator variables and effect sizes (where applicable) of the review Studies

As depicted in Table 6, concerning the coding procedure, only four studies did not mention or conduct a coding procedure, while the rest did. Of the latter, three specified their coding procedures. Regarding data analysis, six studies did not mention any data analysis conducted, whereas eleven did. Fifteen of the seventeen studies mentioned and specified their inclusion/exclusion criteria. In this case, only three studies stated and specified their inter-rater reliability (IRR) agreement, with the remaining fourteen studies having not done so. Again, only five studies conducted a quality assessment of their included articles, and only two studies conducted a publication bias of their reviewed articles, while one study mentioned a publication bias of its included articles. Moreover, only two studies had moderator variables, and only four studies mentioned and calculated the effect sizes of their reviewed articles.

Review study	Coding procedure	Data analysis	Inclusion/Exclusion criteria	Inter-rater reliability agreement	Quality assessment	Publication bias	Moderator variables	Effect sizes
RS 1	X (But mention of four coders).	X	√ (e.g., only papers that identified as articles in the SSCI).	X	√ (But criteria not specified)	X	X (N/A)	X (N/A)
RS 2	√	X	√ (e.g., must involve mobile learning as a primary condition).	X	√ (Features of the quality of study research methodology)	X	X (N/A)	X (N/A)
RS 3	X (But mention of two coders)	X (But reference to classification analysis)	√ (e.g., articles published in peer-review journals and conference proceedings in the period 2005-2012).	X	√ (Used Grönlund and Andersson (2006)'s model)	X	X (N/A)	X (N/A)
RS 4	√	√ (Cross-study comparison and analysis)	√ (e.g., empirical research on mobile learning in preservice and in-service teacher education contexts across different disciplines).	X	√ (Two external reviewers)	X	X (N/A)	X (N/A)
RS 5	√ (Three coders)	√ (Content analysis)	√ (e.g., studies written in English and published as full-text in refereed academic SSCI journals).	√ (Computed by using Huberman and Miles' (2002) formula, was above 0.95 - discussion).	X	X	X (N/A)	X (N/A)
RS 6	√.	√ (Document & content analyses)	X	X	X	X	X (N/A)	X (N/A)
RS 7	√ (Three coders - consensus) .NB: Used Cooper's (2010) coding and the Comprehensive Meta-analysis software	√ (Cohen's d and Hedge's g)	√ (e.g., the applications of mobile devices were the key variables of the study).	X	X	√ (The fail-safe N, i.e., classic fail-safe N was used).	Nine moderator variables	g = 0.55
RS 8	√ (Two coders)	√ (Thematic analysis)	√ (e.g., sound methodological design).	X	√ (Studies with lower methodological quality not considered)	X	X (N/A)	X (N/A)
RS 9	X	√ (Hedge's g & Fisher's Z). NB: But not	√ (e.g., the study was published	X	X	X	X	g = 0.42

RS 10	√ (Two coders)	mentioned as data analysis. √ (Trend and content analyses)	between 2008 and 2015). X	X	X	X	X	X	
RS 11	√ (Two coders - consensus)	X	√ (e.g., students use mobile devices for learning).	√ (96% - discussion)	X	X	X	X	
RS 12	√ (Two coders - discussion)	√ (Content analysis)	√ (e.g., must involve mobile learning as a primary condition).	X	X	NB: Never surveyed the "grey literature"	X	X	
RS 13	√ (Two coders)	√ (Content analysis)	√ (Only articles focusing on mobile collaborative language learning).	X	X	X	X	X	
RS 14	√ (Followed Cooper and Hedges' (2009) coding procedure)	√ (Comprehensive Meta-Analysis (version 3 & Hedge's g)	√ (e.g., the study had an experimental or quasi-experimental design).	X	X	X	X		g = .67, p < .001
RS 15	X	X	√ (e.g., studies that were published between 2007 and 2016 (March). NB: Book reviews, commentaries, or editorial materials, were excluded from the review.	X	X	X	X	X	
RS 16	√ (Two coders - meeting). NB: Elements of What Works Clearinghouse Design and Implementation Assessment Device	√ (Cohen's d & Hedge's unbiased effect sizes)	√ (e.g., mobile device interventions were implemented in grades K-12).	√ (Ranged from 87% to 94%).	X	The fail-safe N procedure and the Egger test	Six moderators		g = .48
RS 17	√	X	√ (e.g., the article (was) related to mobile learning in higher education).	X	X	X	X	X	

Table 6: Coding procedure, data analysis, inclusion/exclusion criteria, inter-rater reliability agreement, quality assessment, publication bias, moderator variables and effect sizes

Main Findings of the Review Studies

As illustrated in Table 7, the findings of the 17 review studies point to multiple focal areas that these studies had concerning MLL, even though in certain cases, there were overlaps. For example, in relation to these overlaps, most mobile learning studies took place in higher education, with mobile phones, smartphones and PDAs as the most used mobile devices. In addition, most mobile learning studies occurred in language arts and science. With regard to language, English as a foreign language (EFL) was the main target language, with vocabulary learning as the key focus of MLL.

Review study	Main finding(s)
RS 1	Most mobile and ubiquitous learning (MUL) studies took place in higher education. Science had most MUL studies, followed by language and art.
RS 2	The research purpose of most mobile learning studies focused on effectiveness, followed by mobile learning system design. Mobile phones and PDAs were the most commonly used devices for mobile learning.
RS 3	MALL is mainly unified by the focus on mobile technologies; it is a theoretically immature field; its borders to CALL or e-learning, are conceptually unclear.
RS 4	There was a scant report of theoretical and conceptual perspectives in mobile learning.
RS 5	Teaching vocabulary with the use of cell phones and PDAs has remained popular in MALL over the reviewed period (2000-2012).
RS 6	Mobile learning studies in Turkey were mostly descriptive studies intended to determine students' or faculty members' opinions.
RS 7	The study found that learning/teaching with mobile devices yielded a moderate mean effect size of 0.531 and 0.550, for the achievement and affective effects, apiece.
RS 8	The majority of mobile and ubiquitous learning studies in higher education indicated positive effects. However, more empirical studies are needed.

RS 9	MALL was an important tool for EFL vocabulary instruction, suggesting that it fostered the EFL instruction. The overall effect size was ($d = 0.8$).
RS 10	Most mobile learning was conducted in the higher education sector (followed by the primary school sector); and most mobile learning supported learning occurred in language and arts, followed by science; smartphones the most used devices.
RS 11	Many studies occurred in formal educational contexts at 50%, and mobile phones had the largest types of devices used at 34%.
RS 12	Mobile learning in higher education is a growing field as shown by various research topics; and mobile phones continued to be the most widely used devices in mobile learning studies.
RS 13	Affordances like flexible use, continuity of use, timely feedback, personalisation, socialisation, self-evaluation, and active participation were emphasised through mobile learning.
RS 14	Overall, with the effect size of 0.67, there is evidence suggesting that mobile devices could be effective in vocabulary learning.
RS 15	Smartphones, mobile phones, and PDAs were the frequently used mobile devices, with English as a foreign language as the most common target language.
RS 16	The meta-analysis found that the achievement scores in science, mathematics and reading subject areas differed depending on the use of mobile devices or traditional teaching methods in the classrooms.
RS 17	The most popular topics were learners' perceptions of mobile technologies and m-learning and students' acceptance of m-learning.

Table 7: Main finding(s) of review studies

Discussion

As stated earlier, the current overview had a dual purpose: to compare twenty-three different aspects investigated by 17 review studies analyzed in this overview and to synthesise the main findings of these 17 review studies. Four typologies were identified by the overview. These were: reviews ($n = 4$), literature reviews ($n = 2$), systematic reviews ($n = 5$), and meta-analyses ($n = 6$). These typologies are consistent with some of those identified by scholars such as Evans and Popova (2015), Grant and Booth (2009), Kim et al. (2018), Pieper et al (2012), Polanin et al. (2016), and Uddin and Arafat (2016). Except for four meta-analyses that provided statistically quantitatively results, the rest of the reviews (including two meta-analyses) provided qualitative results. This finding is consistent with Kim et al.'s (2018) overview study of 171 review studies in which there were more qualitative reviews than quantitative reviews, even though this study focused on hospitality and tourism and not on MLL.

Of the 17 review studies, only nine had clear purposes, of which five had clear and specific purposes. As is the case with all research types, the importance of a clearly and precisely formulated purpose need not be over-emphasised for review studies. Notwithstanding that these 17 reviews had multiple purposes, ten of them had overlapping purposes. One prime example is reviewing research trends and mobile devices in MLL. It is not uncommon for reviews to have diverse purposes or foci. For example, Polanin et al. (2016) report that the education research overviews included in their study focused on a variety of topics. Concerning theoretical frameworks, only two of the 17 reviews employed such frameworks. Elsewhere, highlighting a lack of theoretical frameworks in MLL, especially, Viberg and Grönlund (2013) point out that MLL is a theoretically immature field when compared to fields such as e-learning and computer-assisted language learning (see Yükselir, 2017), despite the fact that MLL has grown exponentially in the last two decades (Arvanitis & Krystalli, 2020; Cho et al., 2018; Lin & Lin, 2019).

Barring two review studies that failed to specify their coverage periods, the other reviews (except one) had varying coverage periods that started from or after 2000. Altogether, these fifteen review studies had a combined coverage period of 24 years, with one review's coverage period spanning 20 years. In Kim et al.'s (2018) overview, the longest review time-frame was 52 years (1960-2012). Four reviews employed one database each, and four reviews utilised two databases. The other reviews employed more than two databases, with one review having used the most databases ($n = 9$). Regarding included articles, one review had the least included articles ($n = 11$), whereas the other review had the most included articles ($n = 233$). In all, the included articles for the 17 review studies were 1,260 (see Figure 3). Of this total number, systematic reviews had the most included articles ($n = 547$), and were followed by meta-analyses ($n = 396$) and reviews ($n = 261$). Similarly, in Kim et al.'s (2018) overview, systematic reviews had the highest sample

sizes (n = 731). However, they were followed by thematic reviews (n = 433) and narrative reviews (n = 124).

There were 34 different mobile device types cited by eleven reviews that reported mobile device types used by their reviewed publications, even though the precise and specific mobile device typologies identified by this overview were 25. One review reported the highest number of mobile device types (n = 16) used by its publications. As depicted in Figure 4, PDAs were mobile devices reported by most reviews (n = 10) followed by tablets (n = 8) and mobile phones (n = 7). Nevertheless, when both mobile phones and cell phones are subsumed under one typology, mobile phones emerge as the second most reported mobile devices (n = 9) after PDAs. In Wu et al.'s (2012) review, mobile phones and PDAs were the most used mobile devices, whereas in Ok and Ratliffe's (2018) review of eleven publications in which the main reported mobile device types were iPods, iPads, e-readers, and smart pens, both iPods and iPads were the most used mobile devices. In line with the theoretical framework adopted in this paper, the different mobile device types reported by eleven reviews point to the mobile-dependence and mobile-enhancement of the learning activities offered through mobile devices. They also reflect mobile presence learning often afforded by mobile devices.

Notwithstanding that three reviews did not report the educational levels of their included publications, the three most reported educational levels were seven (RS 7), six (RS 6) and five (RS 11) educational levels. School levels were the most reported educational levels with the elementary school level cited by most review studies (n = 9) in this sector. The second most reported educational level after this sector was the higher education sector. This finding tallies with Nikou and Economides (2018) review of 43 articles on mobile-based assessment in which the elementary school level resulted as the most assessed educational level followed by the university level. However, this contrasts with Neira et al.'s (2017) systematic review of 288 studies on emerging technologies in which higher education was the prominent educational level. As regards educational contexts, while five reviews failed to report any educational level, one review reported the most educational contexts (n = 7) (see Figure 3). The educational context reported by most reviews was the formal context (n = 12) followed by the informal context (n = 7). In terms of disciplines, natural sciences was the most reported discipline followed by social sciences, while language was the most reported subject area followed by mathematics. In a different but related instance, science and mathematics topped the list of a mobile-based assessment in Nikou and Economides' (2018) review.

While more than 17 of the reviews mentioned their data analysis (n = 11) and specified their inclusion/exclusion criteria (n = 15), fewer reviews did so in terms of their research designs (n = 8), coding procedure (n = 4), IRR agreement (n = 3), quality assessment (n = 5), publication bias (n = 3), moderator variables (n = 2), and effect sizes (n = 4). Conversely, all the reviews failed to mention anything about their sample sizes and data collection methods. Nonetheless, these breakdowns of the review characteristics should be seen against the fact that of the 17 reviews, only six were meta-analyses. The latter, and to some extent systematic reviews, traditionally report on or carry out most of the review factors mentioned here, whereas most classical reviews (ordinary reviews and literature reviews) do not. Of the six meta-analyses, five did not calculate an IRR agreement and did not conduct a quality assessment of their included articles. In contrast, three systematic reviews did not calculate an IRR agreement nor did they conduct a quality assessment of their included articles. Additionally, four meta-analyses did not conduct a publication bias and did not focus on moderator variables. Moreover, all systematic reviews specified their inclusion/exclusion criteria, while one meta-analysis did not.

Finally, the findings of the 17 review studies had overlaps despite the multiple MLL focal areas. The major overlap is that most mobile learning occurred in higher education, and mobile phones, smartphones and PDAs were the most used mobile devices. Additionally, most mobile learning took place in language arts and science, with EFL as the main target language and vocabulary learning as the primary focus of MLL.

Conclusions, Limitations and Recommendations

Having reviewed the twenty-three different characteristics of the 17 review studies, there are salient conclusions to be drawn from the findings discussed above. First, most reviews were qualitative in nature and few of them had clearly and precisely formulated purposes. The last point highlights a flaw in the formulation of the purposes of most of these review studies. These purposes were many but overlapped in terms of their foci on research trends and mobile devices used in MLL. Second, only two reviews had theoretical frameworks. This reveals a lack of theoretical frameworks employed by the majority of these review studies. Third, there were 25 precise and specific mobile device typologies reported by eleven

reviews, with PDAs and mobile phones as the two most reported mobile devices, respectively. Fourth, the elementary school level was the most cited level within the school levels and was followed by the higher education sector. To this end, the formal context was the most reported MLL context and natural sciences was the most reported discipline. Nonetheless, language was the most reported subject area followed by mathematics.

Fifth, more reviews mentioned their data analysis and specified their inclusion/exclusion criteria, while fewer reviews did so in relation to aspects such as research designs, coding procedure, IRR agreement, and quality assessment. Sixth, all the reviews did not mention their sample sizes and their data collection methods. Seventh, five meta-analyses did not calculate an IRR agreement and did not conduct a quality assessment of their included articles, and four meta-analyses did not conduct a publication bias and did not focus on moderator variables. All of these gaps point to the design and methodological weaknesses, especially a lack of design and methodological rigour for most of these review studies. Eighth, the main overlaps from the main findings of the 17 review studies are: most MLL occurred in higher education, mobile phones, smartphones and PDAs were the most utilised mobile devices and most mobile learning occurred in language art and science, with EFL as the prominent target language and vocabulary learning as the main focus of MLL.

Pertaining to the limitations, the number of the articles included in this overview was small, even though the review studies themselves comprised varied typologies. The current overview also has a publication bias. It is skewed towards those publications that it was able to garner from the online search platforms and from the online databases it employed. Additionally, it is slanted towards peer-reviewed journal articles as books, conference papers, dissertations/theses and unpublished articles (grey literature) were not considered. Again, the overview did not conduct a hand search. However, a decision not to do all this was caused by the ongoing COVID-19 pandemic that rendered a hand search an unsafe practice. Notwithstanding the aforesaid limitations, some of the findings of this overview are likely to be applicable and reproducible in future overviews of reviews on MLL. Importantly, this overview is likely to serve as a reference point for future MLL overviews and for subject areas in which overviews of reviews have not yet been conducted. To this end, a major recommendation in this regard is that future MLL overviews of reviews should focus on a higher number of review typologies.

References (* = Reviewed studies)

- Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24, 567–598. <https://doi.org/10.1007/s10758-019-09408-7>
- Arvanitis, P., & Krystalli, P. (2020). Mobile assisted language learning (MALL): Trends from 2010 to 2020 using text analysis techniques. *European Journal of Education*, 3(3), 84–93. <https://journals.euser.org/index.php/ejed/article/view/4881/4738>
- Baillie, L. (2015). Promoting and evaluating scientific rigour in qualitative research. *Nursing Standard*, 29(46), 36–42. <https://doi.org/10.7748/ns.29.46.36.e8830>
- *Baran, E. (2014). A review of research on mobile learning in teacher education. *Journal of Educational Technology & Society*, 17(4), 17–32. <https://www.learnedtechlib.org/p/156112>
- Belur, J., Tompson, L., Thornton, A., & Simon, M. (2018). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*. <https://doi.org/10.1177%2F0049124118799372>
- Bobrovitz, N., Onakpoya, I., Roberts, N., Heneghan, C., & Mahtani, K. R. (2015). Protocol for an overview of systematic reviews of interventions to reduce unscheduled hospital admissions among adults. *BMJ Open*, 5(8). <http://dx.doi.org/10.1136/bmjopen-2015-008269>
- Chaka, C. (2015). Digital identity, social presence technologies, and presence learning. In R. D. Wright (Ed.), *Student teacher interaction in online learning environments* (pp. 183-203). IGI Global.
- Chaka, C., Nkhobo, T., & Lephalala, M. (2020). Leveraging Moya^{MA}, WhatsApp and online discussion forum to support students at an open and distance e-learning university. *Electronic Journal of e-Learning*, 18(6), pp. 494-515. <https://doi.org/10.34190/JEL.18.6.003>
- *Chee, K. N., Yahaya, N., Ibrahim, N. H., & Hasan, M. N. (2017). Review of mobile learning trends 2010-2015: A meta-analysis. *Educational Technology & Society*, 20(2), 112–126. <https://www.jstor.org/stable/90002168>
- Cho, K., Lee, S., Joo, M-H, & Becker, B. J. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences*, 8(105), <https://doi.org/10.3390/educsci8030105>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. <https://doi.org/10.1177%2F001316446002000104>
- Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., Garside, R. (2018). Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies. *BMC Medical Research Methodology*, 18. <https://doi.org/10.1186/s12874-018-0545-3>
- Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, 67(6), 446–462. <https://doi.org/10.1037/a0027119>
- *Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, 110, 51–63. <https://doi.org/10.1016/j.compedu.2017.03.013>

- Dennen, V. P., & Hao, S. (2014). Intentionally mobile pedagogy: The M-COPE framework for mobile learning in higher education. *Technology, Pedagogy and Education*, 23(3), 397–419. <https://doi.org/10.1080/1475939X.2014.943278>
- *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>
- Evans, D. K., & Popova, A. (2015). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. Policy Research Working Paper 7203. World Bank Group. <https://openknowledge.worldbank.org/bitstream/handle/10986/21642/WPS7203.pdf?sequence=1>
- Frohberg, D., Göth, C., & Schwabe, G. (2009). Mobile learning projects—A critical analysis of the state of the art. *Journal of Computer Assisted Learning*, 25(4), 307–331. <https://doi.org/10.1111/j.1365-2729.2009.00315.x>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Haßler, B., Major, L. & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning*, 32(2), 139–156. <https://doi.org/10.1111/jcal.12123>
- Hunt, H., Pollock, A., Campbell, P., Estcourt, L., & Brunton, G. (2018). An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. *Systematic Reviews*, 7. <https://doi.org/10.1186/s13643-018-0695-8>
- *Hwang, G.-J., & Tsai, C.-C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 42(4), E65–E70. <https://doi.org/10.1111/j.1467-8535.2011.01183.x>
- Jalali, S., & Wohlin, C. (2012). Systematic literature studies: Database searches vs. backward snowballing. *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, September, 2012*. (pp.29-38). <https://doi.org/10.1145/2372251.2372257>
- Kim, C. S., Bai, B. H., Kim, P. B., & Chon, K. (2018). Review of reviews: A systematic analysis of review papers in the hospitality and tourism literature. *International Journal of Hospitality Management*, 70, 49–58. <https://doi.org/10.1016/j.ijhm.2017.10.023>
- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* [CiteSeer] (pp. 1–57). Keele University. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.471>
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M. R., Bailey, J., & Linkman, S.G. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*. 51(1), 7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- *Korkmaz, Ö. (2015). New trends on mobile learning in the light of recent studies. *Participatory Educational Research*, 2(1), 1–10. <http://dx.doi.org/10.17275/per.14.16.2.1>
- *Krull, G., & Duarte, J. M. (2017). Research trends in mobile learning in higher education: A systematic review of articles (2011–2015). *International Review of Research in Open and Distributed Learning*, 18(7), 1–23. <https://doi.org/10.19173/irrodl.v18i7.2893>
- *Kukulska-Hulme, A. & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. <https://doi.org/10.1111/bjet.12580>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–74.
- 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>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: assessment and reporting of intercoder reliability. *Human Communications Research*, 28(4), 587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- *Mahdi, H. S. (2017). 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>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052>
- McKenzie, J. E., & Brennan, S. E. (2017). Overviews of systematic reviews: Great promise, greater challenge. *Systematic Reviews*, 6, 1–4. <https://doi.org/10.1186/s13643-017-0582-8>
- Mengist, W., Soromessa, T., Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, 7. <https://doi.org/10.1016/j.mex.2019.100777>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *The BMJ*, 339. <https://doi.org/10.1136/bmj.b2535>
- Neira, E. A. S., Salinas, J., & de Benito Crosetti. (2017). Emerging technologies (ETs) in education: A systematic review of the literature published between 2006 and 2016. *International Journal of Emerging Technologies in Learning*, 12(5), 128–149. <http://dx.doi.org/10.3991/ijet.v12i05.6939>
- Nikou, S. A., & Economides, A. A. (2018). Mobile-based assessment: A literature review of publications in major referred journals from 2009 to 2018. *Computers & Education*, 125, 101–119. <https://doi.org/10.1016/j.compedu.2018.06.006>
- Ok, M. W., & Ratliffe, K. T. (2018). Use of mobile devices for English language learner students in the United States: A research synthesis. *Journal of Educational Computing*, 56(4) 538–562. <https://doi.org/10.1177%2F0735633117715748>
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37. <https://doi.org/10.17705/1CAIS.03743>
- Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, 52, 183–199. <http://dx.doi.org/10.1016/j.im.2014.08.008>
- Persson, V., & Nouri, J. (2018). A systematic review of second language learning with mobile technologies. *International Journal of Emerging Technologies in Learning*, 13(2), 188–210. <https://doi.org/10.3991/ijet.v13i02.8094>
- Pieper, D., Buechter, R., Jerinic, P., & Eikermann, M. (2012). Overviews of reviews often have limited rigor: A systematic review. *Journal of Clinical Epidemiology*, 65(12), 1267–1273. <http://dx.doi.org/10.1016/j.jclinepi.2012.06.015>
- *Pimmer, C., Mateescu, M., & Gröbriel, U. (2016). Mobile and ubiquitous learning in higher education settings. A systematic review of empirical studies. *Computers in Human Behavior*, 63, 490–501. <https://doi.org/10.1016/j.chb.2016.05.057>
- Plonsky, L., & Ziegler, N. (2016). The CALL–SLA interface: Insights from a second-order synthesis. *Language Learning & Technology*, 20(2), 17–37. <http://dx.doi.org/10125/44459>
- Polanin, J. R., Maynard, B. R., & Dell, N. A. (2016). Overviews in education research: A systematic review and analysis. *Review of Educational Research*, 87(1), 172–203. <https://doi.org/10.3102/O034654316631117>

- Ramírez-Montoya, M. S. (2020). Challenges for open education with educational innovation: A systematic literature review. *Sustainability*, 12(17). <http://dx.doi.org/10.3390/su12177053>
- Ramírez-Montoya, M.S., & Lugo-Ocando, J. (2020). Revisión sistemática de métodos mixtos en el marco de la innovación educativa [Systematic review of mixed methods in the framework of educational innovation]. *Comunicar*, 65, 9–20. <https://doi.org/10.3916/C65-2020-01>
- Rikala, J. (2013). *Mobile learning – A review of current research*. Retrieved 13 March 2015 from http://users.jyu.fi/~jeparika/report_E2_2013ver2.pdf
- Schlosser, R. W., Wendt, O., & Sigafoos, J. (2007). Not all systematic reviews are created equal: Considerations for appraisal. *Evidence Based Communication Assessment and Intervention*, 1(3), 138–150. <https://doi.org/10.1080/17489530701560831>
- *Shadiev, R., Hwang, W.-Y., & Huang, Y.-M. (2017) Review of research on mobile language learning in authentic environments. *Computer Assisted Language Learning*, 30(3–4), 284–303. <https://doi.org/10.1080/09588221.2017.1308383>
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., Henry, D. A., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Methods of Systematic Reviews and Meta-analysis*, 62(10), 1013–1020. <https://doi.org/10.1016/j.jclinepi.2008.10.009>
- *Sönmez, A., Göçmez, L., Uygun, D., & Ataizi, M. (2018). A review of current studies of mobile learning. *Journal of Educational Technology & Online Learning*, 1(1), 133–27.
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/ptj/85.3.249>
- *Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (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>
- *Taj, I. H., Sulan, N. B., Sipra, M. A., & Ahmad, W. (2016). Impact of mobile assisted language learning (MALL) on EFL: A meta-analysis. *Advances in Language and Literary Studies*, 7(2), 76–83. <https://doi.org/10.7575/aiac.all.v.7n.2p.76>
- 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>
- Tight, M. (2018). Systematic reviews and meta-analyses of higher education research. *European Journal of Higher Education*, 9(2), 133–152. <https://doi.org/10.1080/21568235.2018.1541752>
- *Tingir, S., Cavlazoglu, B., Caliskan, O., Koklu, O., & Intepe-Tingir, S. (2017). Effects of mobile devices on K–12 students' achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 1–15. <https://doi.org/10.1111/jcal.12184>
- Torgerson, C. J. (2007). The quality of systematic reviews of effectiveness in literacy learning in English: A “tertiary” review. *Journal of Research in Reading*, 30(3), 287–315. <https://doi.org/10.1111/j.1467-9817.2006.00318.x>
- Uddin, Md, S., & Arafat, S. M. Y. (2016). A review of reviews. *International Journal of Perceptions in Public Health*, 1(1), 14–24.
- *Viberg, O., & Grönlund, Å. (2013). Systematising the field of mobile assisted language learning. *International Journal of Mobile and Blended Learning*, 5(4), 72–90. <https://doi.org/10.4018/ijmbl.201310010>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *EASE '14: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 1-10. <https://doi.org/10.1145/2601248.2601268>
- *Wu, W.-H., Wu, Y.-C. J., Chen, C.-Y., Kao, H.-Y., Lin, C.-H., & Huang, S.-H. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education*, 59(2), 817–827. <https://doi.org/10.1016/j.compedu.2012.03.016>
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Yükselir, C. (2017). Meta-synthesis of qualitative research about mobile assisted language learning (MALL) in foreign language teaching. *Arab World English Journal*, 8(3), 302–318. <https://dx.doi.org/10.24093/awej/vol8no3.20>