

What are the effects of mobile technologies' application as an educational tool in LMICs on learning outcomes, teaching practice, education monitoring and information systems (EMIS), and empowerment?

A mixed-methods systematic review of evidence.

March 2014

Laurenz Langer, Ruth Stewart, Niall Winters

Table of Contents

| | |
|---|-----------|
| Background | 3 |
| A theory of mobile learning for development? | 3 |
| A sufficient evidence base? | 4 |
| Systematic review approach | 6 |
| Search strategy | 9 |
| Search Rationale | 11 |
| Inclusion/exclusion criteria | 19 |
| Screening and identification of studies | 21 |
| Coding and data extraction | 21 |
| Critical Appraisal | 22 |
| Methods of synthesis | 22 |
| Meta-analysis:..... | 23 |
| Thematic synthesis: | 26 |
| Appendices | 28 |
| Appendix 1: Coding Sheet..... | 29 |
| Appendix 2: Critical appraisal tool | 38 |
| Appendix 3: Effects size calculations..... | 49 |
| Appendix 4: Deductive themes for thematic synthesis | 50 |

Background

The idea to harness the potential of mobile technology in education is an emerging pedagogical approach. Mobile devices' ubiquity, appeal, and wide applicability profoundly alter the process of learning and teaching. They might also serve as tools of empowerment in the hands of educators and learners by e.g. increasing ownership of the learning/teaching process. Further, modern devices with their improved features and applications can perform sophisticated monitoring and information management tasks in an educational context.

Mobile devices might add particular value to developing-country education systems. A resource-poor educational environment, e.g. in terms of infrastructure, learning materials, or teaching staff and support, opens up different applications for mobile technologies to benefit disadvantaged learners. Firstly, the evident ubiquity of, in particular mobile phones in developing countries, justifies the appeal of mobile devices' use as a vehicle to facilitate learning and teaching. Mobile devices as a medium for teaching and learning might allow geographically or socially isolated groups to better participate in the educational process. In addition to the increase in access to education, the use of mobiles might secondly improve the quality of teaching and learning itself. Learning or teaching with mobiles can lead to more effective educational outcomes, and excite interest and motivation of both, educators and learners. Thirdly, mobile technologies might also improve monitoring and information management systems in education. They can be used to collect and store large data sets, or to enable more horizontal sharing of and access to information.

Fourthly, an often-overlooked aspect of learning and teaching with mobile devices is the impact it can have on educators' and learners' capabilities and sense of empowerment. Developing-country education systems are often organised in a top-down and hierarchical fashion with little voice and agency for both teachers and learners. Teachers in developing countries have little input in the design of curricula and learning materials; learners likewise have few opportunities to shape how learning is facilitated or how the educational process is organised. The application of mobile devices might allow both groups to take ownership of the teaching/learning process and thereby arguably experience an advent of empowerment and increase in capabilities. Mobile devices also allow for better accountability in the educational process, an aspect closely related to improved EMISs.

A theory of mobile learning for development?

Two related schools of thought have attempted to conceptualise how mobile devices, applied in developing countries, relate to education and development. 'Mobile learning for development' is the older and arguable more theoretically grounded school. Based on pedagogic theories of mobile learning, scholars have investigated how this concept applies in the context of developing countries. Mobile learning for development then places an emphasis on the process of learning, and focuses how learners can benefit from learning with mobile devices, given the unique challenges faced in resource-poor contexts.

'Mobile education for development' on the other side is a more recent term, advocated mainly by donor and corporate organisations. It is arguably an extension of mobile learning, and perceives itself as going beyond the mere processes of teaching and learning with mobiles, to include the 'full range' of opportunities, which mobile technologies offer (i.e. improving

assessment tools, and educational administration and management). Mobile education is however not informed by an underlying theoretical framework, and has rarely been investigated in a scientific manner thus far. Both terms, mobile learning and mobile education for development, unsurprisingly regard mobile technologies as a key component to improve education in developing countries and, as explained above, link this to the process of international development.

Yet, whilst the growing adoption of mobile technologies in the domain of international development is an observed fact – the majority of poor people in developing countries have access to a mobile phone – there is little empirical evidence that proves the relationship between mobile technologies, education, and international development; or how these three institutions impact on each other. A simplified view often regards educational outcomes as synonymous with development outcomes. Similarly, the provision of mobile devices or contents to learners and teachers is repeatedly thought to naturally translate into effective educational outcomes.

Such simplifications show a lack of adequate program theory as well as a neglect of underlying mechanisms (e.g. appropriate software or training for teachers). This compromises the understanding of mobile learning/mobile education for development. It appears that the '*for*' in these terms is insufficiently conceptualised, challenging the broader conception of mobile learning/education for development. Little academic inquiry has investigated the program theory that connects mobile technologies, education, and international development. Arguably, if mobile learning/education for development is to make a contribution to wider processes of development there is a need to de-reify 'mobile technologies', and, 'education', and 'international development', to establish an evidence-informed program theory of how the former relate to each other.

A sufficient evidence base?

Such a 'de-reification' and evidence-informed approach is further necessitated, as much of the widely cited research in the field is anecdotal or compromised by bias. The GSMA and McKinsey for example have published major reports in the field, which by nature cannot be separated from corporate interest. Even though of arguably more objective stature, the equally influential UNESCO mobile learning series reports mainly draw from evidence derived by project evaluations and case studies. Evidence of such nature needs to be treated with caution.

Notwithstanding, a number of impact studies have recently attempted to provide relevant and up-to-date evaluations of mobile learning projects in developing countries. These range from qualitative evaluations to RCTs, and thus vary in rigour and approach. Currently, the UNESCO is also engaged in a comparative review of mobile learning projects targeting teachers, as well as a global review of mobile learning and gender empowerment¹.

All in all, the evidence base of mobile technologies application to foster education in developing countries has not been systematically laid out. There have been a number of impact studies

¹ These reviews are likely to be in the fashion of previous UNESCO reports and reviews of initiatives, which – albeit of high quality – do not apply a systematic methodology to review and synthesis the research.

applying different designs but no publication has critically reviewed these to present synthesised findings. In contrast, systematic reviews have successfully been conducted on topics of education and international development, e.g. on the effectiveness of interventions to keep learners in school. Outcomes in these reviews focus among others on learning achievement, teacher performance, or monitoring. An ongoing review also attempts to synthesise findings on empowerment and accountability at community level. In the Information and Communication Technology (ICT) sector, a vast body of reviews investigates the application of ICTs in education in the developed world. Less work has been conducted on ICTs in the context of development, but a small number of reviews exist, including a DFID funded systematic review on ICTs for development partnerships. Similarly, mobile learning in the developed world has been subject to some review; but in the context of education in developing countries, no reviews of mobile learning interventions have been conducted.

Systematic review approach

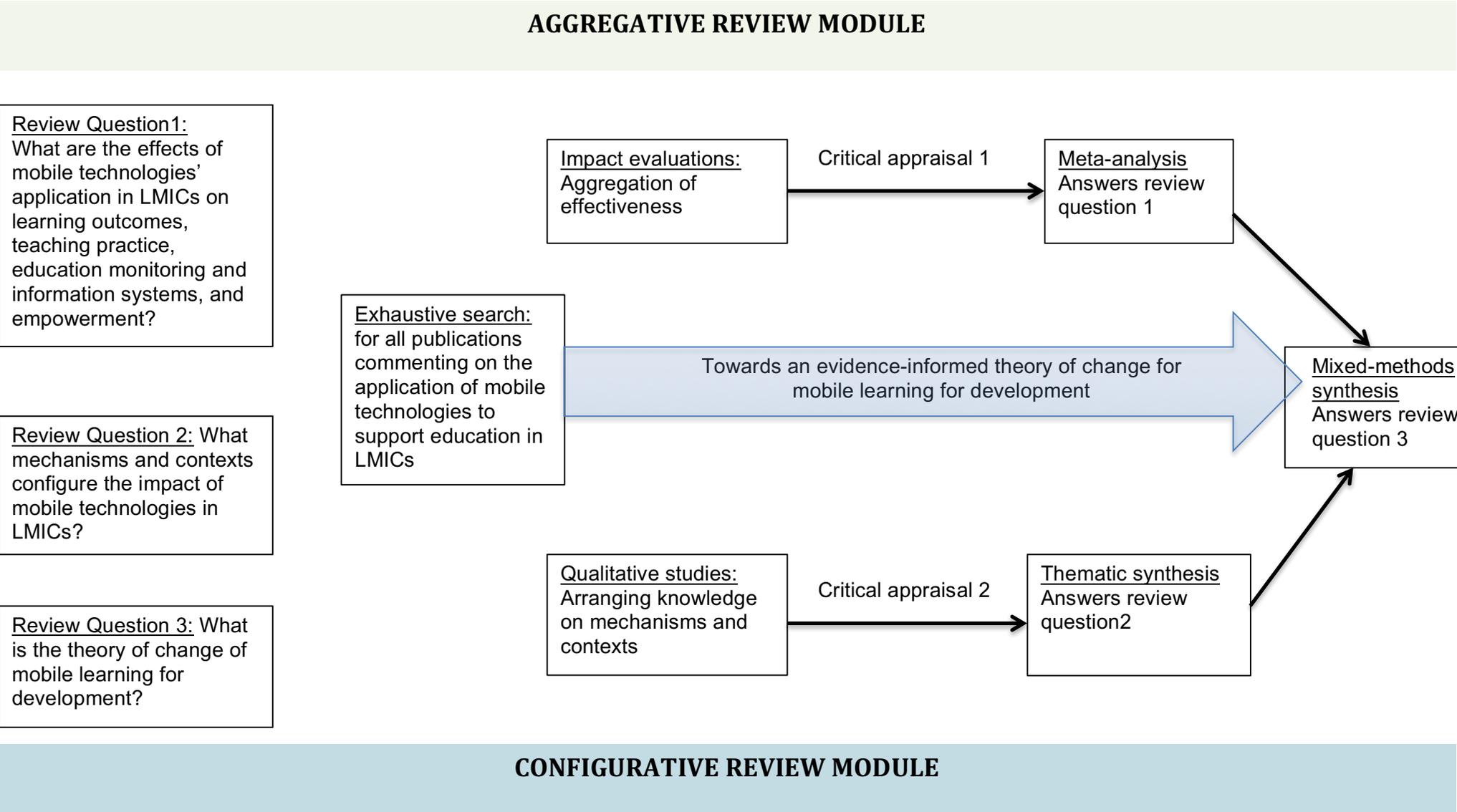
I follow a mixed-method systematic review design in this thesis. This mixed-methods design is operationalised through a two-module review approach². That is, the mixed-methods systematic review features a distinct aggregative review module and a second configurative review module. This two-module design was informed by the need to meet three research objectives in my systematic review:

- (1) Generate evidence of ML4D's impact (or lack of impact).
- (2) Generate analytical themes on mechanisms and contexts explaining this impact (or lack of impact).
- (3) Construct an evidence-informed theory of change of 'what works, how and why' in ML4D.

In order to construct an evidence-informed theory of change of ML4D I require a synthesis of two types of information: (1) evidence of ML4D's effectiveness, defined as aggregative data measuring the impact of mobile learning in LMICs and (2) analytical themes of mechanisms and contexts to explain this impact (or lack of impact). These analytical themes will allow for a configuration of the potential impacts of ML4D, unpacking the black box of how the interventions might have led to the established outcomes. I therefore design two distinct review modules to be able to generate both types of information and to synthesise them in a mixed-methods synthesis. Each module follows its own logic and design applying an aggregative and a configurative approach respectively. This two-module mixed-methods review approach is presented in Figure 1.

² This two-module review approach is inspired by Thomas et al (2003).

Figure 1: Systematic review approach



The contribution of both review modules is of equal weight in generating the evidence-informed theory of change. The relation between the aggregative and configurative module can best be described as a 'sequential explanatory' design (Pluye & Hong 2014), with the configurative module arranging the aggregative findings. Yet, the combination of both types of knowledge to generate an evidence-informed theory of change thereafter is more adequately characterised as a 'convergent' design (ibid). There is a danger that the sequential approach might affect the identification of inductive themes in the configurative module, but this risk seems justified keeping in mind that the module's main mandate is the arrangement and unpacking of the aggregative findings. As a result, after a combined systematic search and application of inclusion criteria, the review commenced with the conduction of the aggregative module. Both review modules used distinct critical appraisal tools and methods to synthesise research findings.

The aggregative module mirrors a traditional effectiveness review. It only includes rigorous impact evaluations of mobile learning programmes in LMICs, judges their quality thoroughly through a risk of bias assessment, and conducts a statistical meta-analysis of the included studies in order to yield a numerical value of ML4D's effectiveness. This pooled effect size is interpreted as evidence of ML4D's impact and since assumed to answer the thesis's first sub-research question. Meta-analysis as a method to establish 'what works' in ML4D was chosen because the method presents the most rigorous tool in synthesising quantitative impact measures (Borenstein et al 2009).

The configurative module resembles a systematic review of qualitative research. It is not limited to a particular study design, places a greater focus on relevance and context in its critical appraisal of research, and applies a thematic synthesis in order to establish analytical themes on contexts and mechanisms at play when teaching and learning with mobiles in LMICs. It thereby aims to arrange the aggregative knowledge with the help of the analytical themes on context and mechanism configurations—a process required to design ML4D's theory of change. Thematic synthesis (Thomas & Harden 2008) was identified as the most rigorous and transparent approach to extract codes and descriptive themes from the included studies, and further, to use these themes to identify mechanisms and contexts in ML4D.

Lastly, the findings of both review modules will be brought together in a mixed-methods synthesis to construct the evidence-informed theory of change of ML4D in graphical and narrative format. In this, the findings generated in the configurative review module allow me to unpack the links between the provision of mobiles, educational process and outcomes, and subsequent changes in socio-economic development. Having outlined these links and relations, I then plot the results of the aggregative module against this theory of change to highlight for which steps and processes in the educational use of mobiles in LMIC there is evidence of effects.

subsequently how these then link to international development and poverty reduction.

This review approach is related to and borrows aspects of (a) realist synthesis (Pawson et al, 2006), (b) theory-based (effectiveness plus) reviews (Snilstveit, 2012), and (c) mixed-methods/mixed knowledge reviews (Gough et al. 2012; Thomas, 2003).

The mixed-method approach to systematic review presents a middle ground between aggregative and configurative reviews. Mixed-methods reviews have been operationalised in different forms depending on the specific review question. Van der Kaap et al. (2009) for example advocate to first conduct an effectiveness review and subsequently apply a realist synthesis to generate different type of insights. A more fluid approach is taken by Thomas et al. (2003), who develop different review questions, review these distinctively, and bring together the findings in a combined synthesis.

In international development, mixed-methods reviews are increasingly popular. Stewart et al. (2012) for instance report findings from studies of different quality separately, and are cited by Gough et al. (2012) as an example of mixed-method international development reviews. Recently, Snilstveit (2012) has coined the term 'theory-based systematic reviews' to refer to reviews of international development interventions which apply a mixed-methods approach. The idea of theory-based reviews however has its origins in theory-based evaluations (White 2009), and is not synonymous with theory-linked reviews, a more established review approach developed by the EPPI Centre.

Theory-based systematic reviews present an attractive review approach, as they address most of the specific challenges I am faced with in reviewing mobile learning for development interventions (diverse evidence with few effectiveness studies, lack of clear program theory, complex intervention, importance of context). However, Snilstveit's conceptualisation of theory-based reviews is still linked to the idea of effectiveness, as she states, "the additional review module [effectiveness plus with a parallel review component] should be designed to answer specific questions of *relevance to the effectiveness of the intervention*. The idea of effectiveness thus continues to feature predominantly in this otherwise expanded notion of theory-based systematic reviews. The review approach that I propose would rather assume evidence of impact and evidence of context/mechanism/perception as equally beneficial and complementary to each other. I therefore propose to broaden the idea of theory-based reviews, and extend it to include studies that can be of mere relevance to the program theory of the intervention, and need not be related to the effectiveness of the intervention.

Search strategy

Systematic review by definition of the methodology entails the conduction of an exhaustive, transparent and replicable search for all available relevant material. This requires a wide range of search sources to cover published and unpublished materials, as well as explicit search terms, which identify key words and terms related to the intervention.

The search strategy for this systematic review will be designed as broadly as possible, keeping in mind the following aspects, which arise from the nature of the evidence base and the nature of the review format:

- mobile learning publications are anticipated to be found mainly in the academic literature, published as journal articles or books.
- mobile learning conferences pioneered its export into the field of international development, and valuable information can therefore be accessed in the conference proceedings and reports
- the UNESCO is a key advocate and knowledge-producer in the field of mobile learning for development
- a number of key experts have featured on most major publications on mobile learning for development
- mobile education publications are anticipated to be found mainly in the grey literature in form of reports of various formats
- the GSMA as well as USAID are key advocates of mobile education, and their websites feature the most extensive collection of mobile learning/education initiatives
- evaluations of these projects might be a valuable source
- a mixed-methods systematic review can include quantitative and qualitative evidence if studies are of sufficient rigour
- a mixed-methods systematic review tries to identify contexts and mechanisms as much as the final outcomes, which the former influence
- a mixed-methods systematic review can include findings from different interventions if the underlying contexts and mechanisms apply to the context of the review (e.g. mobile health projects in developing countries might investigate empowerment related outcomes or processes; if mechanisms can be identified that are likely to apply to the context of mobile learning/education, then these findings of the study can be incorporated in the review

I will conduct a transparent, systematic and exhaustive search for publications. The search strategy is deliberately formulated to be over-inclusive and as broad as possible. These two key features inform both the design of the search terms as well as the choice of search sources. The search terms, for example, include “mobile learning” as well as mobile AND learning; technically the latter term encompasses “mobile learning”, but both terms were applied nevertheless. The search sources were equally over-inclusive. They include a number of journals (e.g. *Journal of Educational Development*), which are included in one of the databases but were equally searched individually. Lastly, in databases that feature sufficient search engines, scoping searches were run prior to the application of the master string. For example, before applying the mobile learning terms and the development terms with the ‘AND’ boolean (which reduces the number of search hits to studies which entail both terms) a search for mobile learning only was run to see how many and what type of mobile learning publications the database has stored. If the application of, and combination with, the development terms led to an unrealistic reduction in search hits, or to an entirely different set of publications, the search strategy was reconsidered. An unanticipated nuance of some databases was the auto-correction of the term ‘mEducation’ to ‘medication’, and the subsequent inflation of search hits with studies from the medical literature. The search strategy was then accordingly adjusted to include the parameter of ‘NOT medication’ in these databases.

Search Rationale

The rationale behind this over-inclusive search strategy is three fold. Firstly, the literature on mobile learning for development is limited, and not structurally archived. This negates a conventional search strategy, which applies one master string of search terms in a small number of academic databases. Studies on mobile learning in the context of international development are predominantly published as policy papers or evaluation reports, which cannot be identified through academic database searches only. Search sources thus need to be diversified and place special emphasis on the grey literature, such as organisations' websites and libraries; International conferences on mobile learning and development have also been key to the advancement of the field (e.g. UNESCO Mobile Learning Weeks) and have contributed valuable research, which is not necessarily published in the academic literature. Forward snowballing such as citation searches using Google Scholar, or Twitter searches were also applied to more comprehensively screen the grey literature.

Mobile learning interventions further are not described applying a unique terminology. Early interventions were often formulated under the notion of ICTs for education (e.g. TESSA and DEEP), and current projects have also used different labels (mobile education for development; human-computer interaction for development). In some instances the interventions are merely described by the applied technologies (e.g. smart animations for development; tablets for development, etc.). This lack of a clear terminology complicates the design of one master string of search terms and informed the decision to extend the number of search sources in order to control for the potential neglect of studies at the margins of conventional terminology. As a result, the literature of mobile learning for development – unlike the health care literature from which systematic searching emerged – is not neatly collected and compiled structurally. The literature therefore cannot be searched using one umbrella term. The broad range of search sources was designed to address this specific characteristic of the literature.

Secondly, the bias and limitation of being a single reviewer informs the considerable overlap between search sources. I deliberately ensured that search sources are likely to feature similar results (databases/journals) to address the shortcomings of having a single reviewer screening all search hits. A rigorous systematic review approach would require two independent reviewers to each screen the same (at the least subset) of the search results. The reviewers would then compare their results and ensure that they are in agreement and apply the same decision criteria. This aims to reduce bias in the review process. Given the scope and funding of this research, such an approach was not feasible, and I am aware that being a single reviewer introduces the risk of personal bias and mistakes (e.g. failing to save or flag a study).

The considerable overlap between search sources assumes that the same studies were often identified from a number of sources, and that I as a reviewer would be faced with a decision to include or exclude them multiple times. This mechanism attempts to control for bias such as personal condition (fatigue), or context (previous inclusions/exclusions), or administrative mistakes (not saving the PDF). This is by no means an alternative to having two independent reviewers but nevertheless aims to introduce some check and balance to increase the rigour of the review.

Thirdly, I deliberately attempted to design an exhaustive search strategy – that is a search strategy that aims to identify and collect all relevant studies. Albeit, this might rarely be possible

in practice, exhaustive searching as a stated aim necessitates the design of search terms and choice of search sources, which are likely to get the search effort as closely as possible to this aim. The application of an exhaustive search strategy is a feature of effectiveness reviews, and has been challenged by reviewers using a realist or mixed-method approach to systematic reviews (Pawson et al 2006; Thomas et al 2004). These reviewers present the concept of purposeful sampling leading to theoretical saturation as an alternative to exhaustive searches. In this search approach, which has its origin in qualitative research, the search strategy does not aim to identify all relevant studies, but rather aims to identify all relevant theories and phenomena related to the research question. Once a theory is identified and explained adequately in one study, it is not necessary to search for another study describing the same theory.

The deviation from an exhaustive search in favor of a purposeful sampling however compromises the rigor, and more crucially, the replicability of the systematic review. Points of saturation might differ between reviewers and depend on prior knowledge. It is also more difficult to keep purposeful samples updated during the review process. Further, the effectiveness component of my review requires the application of an exhaustive search strategy.

Based on the above, a master search string is presented below. A master string serves as an example of how the search will be designed for the most sophisticated search engines. In most cases however search terms will have to be adapted, or key word searches will apply.

Terms:

(1) *Intervention terms:*

("mobile learn* ") OR mlearning OR m-learning OR ("mobile educat*") OR meducation OR m-education OR ("mobile teach*") OR "portable interactive learning technology" OR ("mobile phon*" AND learn*) OR ("mobile phon*" AND educat*) OR ("mobile phon*" AND teach*) OR ("smart phon*" AND learn*) OR ("smart phon*" AND educat*) OR ("smart phon*" AND teach*) OR ("mobile devic*" AND learn*) OR ("mobile devic*" AND educat*) OR ("mobile devic*" AND teach*) OR ("mobile technolog*" AND learn*) OR ("mobile technolog*" AND educat*) OR ("mobile technolog*" AND teach*) OR "mobile and contextual learning"

AND

(2) *Development terms*

development OR "international development" OR "social development" OR poverty OR inequality OR "social change" OR ("develop* countr*") OR ("develop* nation") OR ("develop* world") OR "majority world" OR ("less developed countr*") OR ("less developed nation*") OR ("less developed world*") OR ("underdeveloped countr*") OR ("underdeveloped nation") OR ("underdeveloped countr*") OR ("underdeveloped world") OR ("low income countr*") OR ("low-income countr*") OR ("low income nation") OR ("low-income nation") OR ("middle income countr*") OR ("middle-income countr*") OR ("middle income nation") OR ("middle-income nation") OR "third world"

OR (“deprived countr*”) OR (“deprived nation”) OR (“poor* nation”) OR (“poor* countr*”) OR “global south”

OR

(3) *Country terms*

A full list of countries based on the World Bank’s list LMICs will be used.

(4) *Methods filter*

Due to the nature of the review format, no methods filter will be applied.

(6) NOT medication

Combining these terms, the *default search* can be simplified as:

(1) AND (2) OR (3) OR (4)

As stated above, the default search requires a sophisticated search engine to be applicable. This is usually only the case for major academic databases. Since those present only a minority of search sources (see below), search terms will have to be adapted individually. In such instances, only key terms from (1) will be applied ensuring the search strategy remains as broad as possible and no studies are overlooked. If the identified numbers appear too large, a term of (2) will be connected with the ‘AND’ boolean operator to reduce search hits. In general, a flexible approach to search terms will be applied, carefully weighting their value against possible predefined key word categories in databases, and redefining the terms for most sensible usage in each individual database.

Search Sources:

Studies connecting the ideas of mobile technologies, education, and international development will come from a variety of sources, including the academic as well as grey literature, and most likely published and unpublished reports.

Academic sources

(1) Databases:

EBSCO all databases including among other (Academic Search Complete, ERIC, EconLit, Education FullTexts, Teacher Reference Centre); EdITLib, (Digital Library for Education and Information technology), Ingenta Connect; JSTOR; SabiNet; SAGE Journals Online; Science Direct; Taylor&Francis Online; ISI Web of Science; ELDIS, IOE OER guide and DOAJ guide and International Education Guide

(2) Journals:

Most of the journals below are included in at least one of the database above. However, to double check that the databases search parameters did not miss relevant publications, individual key word searches will be conducted for the period of 2004-2014 in the journals listed below. Individual key word searches address the redundancy of applying the search

terms rigidly; for example, it does make little sense to search for the term 'mobile learning' in a journal dedicated to the subject of mobile learning. Journals listed with an asterisk (*) are assumed to be key journals relevant to the type of publications this review seeks to identify, and will therefore be hand searched exhaustively on title and abstract for the same period.

Development related:

- Economic Development and Cultural Change
- Journal of Development Economics
- *Journal of Development Effectiveness
- *Journal of International Development
- Journal of Sustainable Development
- *World Development
- World Bank Research Observer
-

Education related:

- British Journal of Education
- Educational Researcher
- European Journal of Open and Distance Learning
- *International Journal of Educational Development
- Journal of the Learning Sciences

ICT related:

- African Journal of Information and Communication
- Computers and Education
- Computers and Human behavior
- Electronic Journal of Information Systems in Developing Countries
- Information, Communication & Society
- Information Technologies and International Development
- Information Technology for Development
- International Journal of Education and Development Using Information and Communication Technology
- International Journal of Mobile and Blended Learning
- International Journal of Education and Development Using Information and Communication Technology
- International Journal of Information and Communication Technologies for Human Development
- International Journal of Mobile learning and Organisation
- International Journal on Advances in ICT for Emerging Regions
- Journal of Information Technology in Social Change
- Journal of Computer Assisted Learning
- Learning Media and Technology
- South African Journal of Information Management
- International Journal of Learning Technology.
- International Journal of Interactive Mobile Technologies (IJIM)
- International Journal of ICT Research and Development in Africa
- Journal of Health Informatics in Developing Countries

- International Journal of Information Systems and Social Change
- International Review of Research in Open and Distance Learning

(3) Thesis/Dissertation search:

- ProQuest
- Ethos

Grey literature

(1) Google, Google Scholar

(2) Websites of key organisations

AusADI; Bill&Melinda Gates Foundation; Centre for Development Informatics; Consortium for Research on Educational Access, Transitions and Equity (CREATE);

DFID; GSMA; MIT Poverty Action Lab; ICT4D blog (Heeks); IICD; IDRC; Innovations for Poverty Center; International Association for Mobile Learning; London Knowledge Lab; Network for Policy Research, Review and Advice on Education and Training (NORRAG); Network of Networks Impact Evaluation Initiative (NONIE); OECD; ODI; Research Consortium on Educational Outcomes and Poverty (RECOUP); UNESDOC; USAID (mEducation alliance); World Bank (especially their impact evaluation section); World Bank (EduTech blog); World Bank (infoDev); 3ie.

(3) Conference proceedings:

eLearning Africa; IADIS International Conference Mobile Learning (2005-2011), mEducation symposium (2011, 2012, 2013), mLearn,(2002-2011; UNESCO mobile learning weeks (2011; 2013;2014), Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE); INTEL education summit (2011; 2012) ICTD (2006-2013).

Snowballing

(1) Expert contacts:

Mohamed Ally; John Cook; Jonathan Donner; Richard Heeks; Agnes Kukulska-Hulme; Dorothea Kleine; Tim Unwin; Mike Sharples; John Traxler; Michael Trucano; Niall Winters; Maggie Verster; Steve Vosloo

(2) Twitter search:

- # m4d
- # ict4d
- # ict4e
- # mlearning
- # Edchat

(3) Reference/Citation searches of key publications:

- **Ally** (2009) Mobile Learning: Transforming the Delivery of Education and Training (205)
- **Donner** (2008) Research Approaches to Mobile Use in the Developing World: A Review of the Literature. (381)
- **Frohberg** (2009) Mobile learning projects – a critical analysis of the state of the art (162)
- **Kinuthia & Marshal** (eds) (2013) On the move: Mobile Learning for Development (0)
- **Kukulska-Hulme & Traxler** (2005) M-learning: A Handbook for Educators and Trainers. (401)
- **Naismith et al** (2004) Literature Review in Mobile Technologies and Learning. 629
- **Pachler et al** (2010) Mobile learning: structures, agency, practices. 177
- **Sharples et al** (2007) A Theory of Learning for the Mobile Age. (407)
- **Traxler** (2006) Mobile learning in developing countries (25)
- **Traxler** (2009) Learning in a mobile age (129)
- **Worldreader** (2012) iRead Ghana study. Final evaluation report. (0)

- **IICD** Mobile Learning reports
- **GSMA** Mobile Learning Reports
- **UNESCO** Mobile Learning Reports turning on 8
- **WB** Maximizing Mobile

(4) Previous research synthesis (SRs, MAs) and their included studies:

| Key words | Author | Type of review |
|---|--------------------------------|-------------------------------|
| Combined | | |
| All interventions on enrollment; attendance; drop-outs; learning outcomes | 3ie (2013) | 3ie: Working Paper |
| All interventions on enrollment | Petrosino (2012) | 3ie: SR |
| All interventions that improve learning at primary school level | McEwans (2013) | WB: SR & meta-analysis |
| Synthesis: Evidence-based education in Africa | JPAL (2013) | JPAL: Brief Synthesis |
| Synthesis: Overview of RCTs in developing-country education | Kremer (2009) Kremer (2013) | WB: Overview ScienceDirect |
| Effect of school resources on educational outcomes | Glewwe (2011) | WB: SR |
| Teachers | | |
| Effectiveness of rise in teacher salaries | Carr (2011) | EPPI-centre: SR |
| Interventions to increase teacher attendance | Guerrero (2012) | EPPI-centre: SR |
| Strategies to improve performance of under-trained? | Orr (2013) | EPPI-centre: SR |
| Effectiveness of pedagogic approaches | Westbrook (2013) | EPPI: literature review |

| | | |
|---|----------------------|-------------------|
| | | |
| Health | | |
| Female toilets on girl's enrollment and attendance | Birdthishle (2011) | SR: (LSHTM/IOE) |
| School feeding on nutritional outcomes | Kirstjansson (2006) | EPPI-centre: SR |
| Deworming and attendance & nutrition | T.-Robinson (2012) | EPPI-centre: SR |
| Effects of assessment programs on educational policy | Best (2013) | EPPI-centre: SR |
| | | |
| ICTs | | |
| ICT4D partnerships | Geldorf (2011) | DFID: SR |
| M4D approaches | Donner (2007) | Literature review |
| ICTs and other technology (no mobile) | Waxman (2003) | MA |
| Lit review of evaluations of OLPC | Nguhro (2010) | Literature review |
| Mobile computer-supported collaborative learning. A review of experimental research | Hsu (2013) | Review |
| | | |
| Single interventions | | |
| CCT/UCTs on enrollment; attendance; test scores | Baird (2013) | Campbell: SR |
| Effectiveness of school vouchers | Morgan (2013) | EPPI-centre: SR |
| Effectiveness of School-based monitoring | BarreraOsorio (2012) | WB: review |
| | | |
| Other | | |
| Effectiveness of higher education | Clifford (2013) | EPPI-centre: SR |
| Education and economic growth | Hawkes (2012) | EPPI-centre: SR |
| Access to education for people with disability | Bakhshi (2013) | EPPI-centre: SR |
| | | |

(5) Website search of mobile learning for development projects:

List of projects

- 1 Road to reading
- 2 World Reader
- 3 Project ABC
- 4 Paje Nieta
- 5 EIA
- 6 Shaquodon

- 7 Yoza
- 8 MILLE (Mobile & Immersive Learning for literacy in emerging economies)
- 9 Programa Nacional de Alfabetización
- 10 Cambridge to Africa (deaf)
- 11 PSU Movable
- 12 learning assessment through mobile
- 13 Educational bridges
- 14 roots of mobile learning
- 15 SMILE
- 16 Jokko initiative
- 17 Fire & Gold
- 18 eSchool 360
- 19 Impact network
- 20 mUbuntu
- 21 eLife
- 22 Tangerine
- 23 Total Reading Approach for Children
- 24 Cocoa Link
- 25 Global Literacy Professional Development Network
- 26 Ustad Mobile
- 27 FATIH project turkey
- 28 LISTA
- 29 1001
- 30 across radio
- 31 AlfabeTIC
- 32 Bangladesh Virtual Interactive Classrooms
- 33 BridgeIT
- 34 Bridges to the future
- 35 Broad Class
- 36 Bunyad Mobile
- 37 Dr Math
- 38 eEGRA
- 39 FunDza
- 40 Interactive Radio Instruction
- 41 International Children's Digital Library
- 42 Women Mobile Literacy Afghanistan
- 43 SIRIP Somali Interactive Radio Instruction Program
- 44 Shellbooks

- 45 Mobiliteracy Uganda
- 46 MoMath
- 47 Senmobil
- 48 Nokia life tools
- 49 Earth Institute
- 50 Sesam
- 51 Teacher mate
- 52 Pratham
- 53 PIEQ
- 54 PAJEF
- 55 OLPC
- 56 Twaweza
- 57 Tessa
- 58 IFADEM

This search strategy aims to identify journal articles, books, book chapters, working papers, conference proceedings, final project reports, evaluation reports, pilot studies and blog posts in English. Only in the case of Google Scholar, The World Bank Documents, and Reports and UNESDOC, the search will be repeated with key word terms in Spanish and French. As regards the Google and Google Scholar searches, the first 500 search hits will be screened respectively.

Inclusion/exclusion criteria

Inclusions/exclusion criteria are not as neatly defined and narrowly applied in a mixed-methods systematic review, as in conventional review approaches. The final aim of reporting and synthesising data along the intervention's program theory and ambition to refine and improve this theory means that a broader range of information is required. This changes in particular how outcome and study design criteria are designed.

Region: All studies conducted in LMICs will be included in the review. Studies, which combine data from LICs/MICs and HICs, will be eligible for inclusion, given they report data and results separately for each region.

Population: No criteria related to the study population were applied. That is, study populations could relate to individuals (e.g. teachers or learners), any other form of human organisation (e.g. communities), and administrative groupings (e.g. schools).

Intervention: Neither the term 'mobile learning for development' nor the term 'mobile education for development' seems to be adequately conceptualised in order to derive an operating definition. The intervention will therefore broadly be defined as 'mobile

technologies'³, which refers to any ICT device capable of mobile usage. The review will thus include any mobile technology intervention targeting education in developing countries.

Outcomes: The review will focus on final and intermediate outcomes, as well as contexts and mechanisms affecting each of these. As it is impossible to identify all mechanism, contexts, and intermediate outcomes before hand (and that it is part of a theory-based reviews' purpose to discover these), only a list of final outcomes related to education will be established in this protocol.

The review will focus on four primary outcomes of interest:

- (1) *Learning outcomes:* are outcomes, which relate to learners' attendance, performance, or motivation. Any outcome that indicates changes in the process or perception of learning is treated as a learning outcome. These thus refer to the process of learning, and teachers undergoing a training course facilitated with the help of mobile technologies can equally experience learning outcomes.
- (2) *Teaching practice:* are outcomes that relate to educators' use of mobile technologies as a teaching tool. Any outcome that indicates changes in lesson design (e.g. innovation; inclusion; effectiveness), or changes in the teaching approach (e.g. collaboration among teachers), or changes in educators' professional behaviour (e.g. absenteeism rates) will be applicable.
- (3) *Education monitoring and information systems (EMIS):* this refers to changes in the way information is collected, administered, and analysed in education systems, caused by the use of mobiles technologies. It also includes the impact mobile technology can have on monitoring and evaluating actors in the education system.
- (4) *Empowerment:* might either be an outcome or a process. As an outcome it will broadly be defined to embrace various themes such as critical thinking, agency, opportunity, confidence, etc. Empowerment outcomes can be observed at any level (e.g. individual, school, community). However, on higher levels (national) it is more difficult to attribute a possible change in empowerment to the mobile learning/education intervention.

Study Design: As stated above, the criteria of a rigorous study design does not apply in a theory-based systematic review to the same extent as it does apply in an effectiveness review. With regards to the search for studies to be included in the systematic map of the evidence base, two separate criteria for qualitative and for quantitative studies will be applied. Both will be embedded in the guiding principle of relevance and rigour: relevance (does the research address the program theory under test; and rigour (does the research support the conclusions drawn from it by the authors).

In the light of this principle, qualitative and quantitative studies need to at least report on the following: methodology/methods, sampling strategy, data collection and methods of analysis. Quantitative studies need to be at least of quasi-experimental design.

³ there is disagreement whether technology can be regarded as an intervention in the context of international development at all; opponents regard it as a mere tool rather.

Language: Studies will not be excluded on grounds of language. If non-English studies are identified, Google Translate will be used to review title and abstract. Non-English studies deemed eligible for full-text inclusion will be translated in full by qualified colleagues. It should be noted however that search terms are pre-dominantly in English.

Date: A cut-off date of 1995 for publications will be applied, linked to the emerge of mobile technologies.

Screening and identification of studies

This section will lay out in more detail the process of how studies are searched for and screened in order to identify the final set of includes.

- (1) The above search terms (or key words where applicable) will be applied in the identified sources to generate a list of search hits for each source. To ensure replicability and transparency, records will be kept on what terms were applied, when, how many hits they generated, and possible limitations of the search strategy. The term 'mobile learning for development' will equally be run as a first scoping search in each source.
- (2) The search hits will then be screened on title and abstracts regarding the above stated inclusion criteria. No record will be taken at this stage of the reason for exclusion. Studies of which title and abstracts do not yield sufficient information to make a final decision will be temporarily included. Studies that do not seem to fit the criteria but appear of relevance in the wider context (e.g. mentioning an ongoing project) will be exported to endnote and flagged for follow up.
- (3) The first list of preliminary includes will be exported to EPPI-Reviewer and coded on basic information (tbc, but likely along: region, design, technology used, outcomes) to create a systematic map of the evidence base.
- (4) Based on this systematic map, weaknesses in the review protocol will be assessed and where necessary the review approach will be amended; additional searches might have to be run. The information from the systematic map will allow for conclusion whether there is sufficient evidence to answer the review question; should this be the case, the review will resume to engage in an in-depth coding, appraisal and synthesis of the evidence.
- (5) A final search might have to be run once the synthesis is almost complete. This serves to seek out additional studies that might further refine the program theories that have formed the focus of analysis. For example, one might identify that youth report an increase in capabilities just through the mere use of mobiles (not necessarily through using a mobile learning application); this mechanism might similarly be reported in studies of mHealth projects, and such studies would be targeted by this final search.

Coding and data extraction

I developed a detailed coding tool to extract relevant data from the included studies (Appendix 1). EPPI-reviewer software was used to generate coding sets and to facilitate data management. The coding strategy can best be described as 'mixed coding' (Oliver & Sutcliffe 2012). Codes for the aggregative review module were pre-defined, whereas codes for the configurative module

were open codes. However, for the configurative review module I also predefined a list of descriptive themes, which the literature suggested to be of relevance to the construction of a theory of change for ML4D. It was required to define these deductive codes in order to assess an absence of evidence on these themes.

Critical Appraisal

Critical appraisal refers to the process of assessing the trustworthiness and relevance of the studies included in a systematic review. It is a required review step in order to ensure that the conducted synthesis is based on reliable research results. Due to the diversity of configurative and aggregative data a critical appraisal tool that can cater for both qualitative and quantitative studies was required. I developed a critical appraisal tool drawing on Pluye and colleagues' (2011) Mixed-methods assessment tool (MMAT) as well as Sterne and colleagues' (2013) risk of bias tool for non-randomised studies. My critical appraisal tool uses the MMAT's basic structure but extends its criteria to assess the quality of the qualitative and quantitative studies. To ensure the comparability of the appraisal of both types of studies, six domains of appraisal judgments were developed for each study type. The full tool is provided in the Appendix 2.

Critical appraisal of quantitative studies: For quantitative studies, the developed critical appraisal tool assessed the rigour of the impact evaluation design to establish the reliability of the reported aggregative effect. The tool needed to be able to assess both randomised and non-randomised impact evaluations. In this remit, a Cochrane risk of bias tool for non-randomised studies was adapted (Sterne et al 2013). The tool assessed six domains of bias: (1) selection bias; (2) bias due to baseline confounding; (i) bias due to ineffective randomisation⁴; (3) bias due to departures from intended interventions; (4) bias due to missing data; (5) outcome reporting bias; and (6) bias in selection of reported results. Studies were judged on a scale from critical to low risk of bias, and studies of critical risk of bias were excluded from the synthesis.

Critical appraisal of qualitative studies: For qualitative studies, the developed critical appraisal tool was based on the underlying principles of rigour (in the research conduct) and relevance (contribution to the research question). These were broken down into six domains of: (1) research is defensible in design; (2) research features an appropriate sample; (3) research is rigorous in conduct; (4) research findings are credible in claim; (5) research attends to contexts; (6) research is reflexive (CASP 2006; Dixon-Woods 2006). Diverting from the tool for quantitative studies, no scaled appraisal scale was applied to rate the qualitative studies. A study was either included as making a reliable contribution to the research question or excluded as not rigorous or not relevant. No distinction was then made between the quality of the respective contributions.

Methods of synthesis

I applied two different methods of synthesis for the aggregative and the configurative review module. Statistical meta-analysis was used in the aggregative review module and thematic synthesis was used in the configurative module.

⁴ This domain of bias was only applicable to RCTs, which by design can account for domain (1) and (2).

Meta-analysis:

In the aggregative review module, I conducted a statistical meta-analysis to investigate the impact of mobile technologies on education in LMICs. Meta-analysis is the most rigorous method to synthesise quantitative research studies (Lipsey & Wilson 2001; Borenstein et al 2009). As a statistical approach, it aggregates the numerical effect sizes of research results to report a pooled overall numerical value. This numerical value—the pooled effect size—expresses the overall finding derived from the combined primary research results. The pooled effect size reflects the direction and magnitude of the observed primary effect sizes, which are allocated different weight in the analysis depending on sample sizes and variance.

I report calculated effect sizes in tabular format as well as using forest plots. Where sufficient contextual homogeneity prevailed, effect sizes were averaged across studies by using an inverse variance weighting of the individual effect size. This weighting results in the individual effect sizes of studies with larger study samples being given more weight in the combined, pooled effect size. The meta-analysis was carried out using random effects statistical models.

Statistical synthesis of educational outcomes needs to take into consideration the diversity of study designs and outcome measures. I envisaged that learning outcomes would be the main outcome reported, and likely be reported in form of test scores or grades. This presents a continuous outcome, and outcomes measures and scales were expected to differ between studies. As a result, the meta-analysis calculated standardised mean differences (SMD) in order to yield comparable effect sizes. Cohen's d as well as Hedge's g were calculated for each included study, with g being the effect size used in the meta-analysis due to its ability to adjust for small sample bias prevailing in d (Deeks et al 2001). SMDs express the measure of effect in a change of standard deviations making the effect size difficult to interpret. For ease of interpretation, I report SMDs in this thesis alongside the corresponding percentage change in the intervention group over the control group. All formulae for effect size calculations are reported in Appendix 3.

Assessment of heterogeneity: Given the assumed heterogeneity in the true effects of studies across, for example, geographical reach, educational systems, and socio-economic contexts, a random effect model was applied. In order to visibly examine variability in the effect-size estimates, we will use forest plots to display the estimated effect sizes from each study along with their 95% confidence intervals. Subsequently, and acknowledging the limitations of a quantification of heterogeneity and the different strengths of statistical approaches, the following test for heterogeneity will be conducted: calculation of the Q statistic as a statistical test of heterogeneity (Hedges & Olkin 1985); calculation of the i^2 and Tau^2 statistic to provide estimates of the magnitude of the variability across study findings caused by heterogeneity (Higgins 2002; Higgins 2003).

Sensitivity analysis: To test the robustness of the results of the meta-analysis, a number of sensitivity analyses will be conducted. Broadly, this will involve collecting data on, and assessing sensitivity of, findings to (i) the methods of the primary studies and (ii) the methods of the review.

The included studies are likely to vary methodologically. If sufficient studies are identified, I will conduct sensitivity analyses to examine the influence of these variations on the summary

measures, in order to offer possible explanations for the differences between studies when interpreting the results. I will aim to examine whether the results are sensitive to study design, the risk of bias associated with the study, the degree of missing/incomplete data, the way outcomes were measured and the timing at which they were measured, the treatment effect estimator and comparator, and the use of adjusted analysis.

A series of sensitivity analyses will also be conducted to examine whether any decisions made during our own analyses substantially alter the review findings: for example, the specific statistical procedures and methods I selected to compute each effect size, decisions relating to transformation between effect size metrics, the way outlier effect sizes and sample sizes were handled, and whether or not I replaced missing data with substituted values. As explained above, the main objective of the sensitivity analysis will be to serve as visual tool that allows for informal comparisons whether the results of our meta-analyses are sensitive to methodological decisions of our review team. However due to the controversy of pooling studies of random and non-random design, as well as of different risks of bias, I will follow up the sensitivity analyses of these two variables with a one-way random effects ANOVA model calculated in EPPI-reviewer. That is, the mean effect size and standard error for each group of studies is calculated to test whether these means are statistically significant from one another (Lipsey & Wilson 2001).

Moderator analysis: If there are sufficient data, I will conduct moderator analyses to try to explain variation in effect sizes. Moderator analyses will be reported in tabular format below each meta-analysis. Analyses will be calculated using the same one-way random effects ANOVA model as applied for sensitivity analyses.

Below, I pre-specify a list of *a priori* moderator variables assumed to moderate the true effect of the interventions:

- **Applied technology**, e.g. Basic/feature phones; Smart phones/PDAs; Tablets
- **Model of device ownership**, e.g. 1:1 model; Shared ownership
- **Socio-economic contexts**, e.g. Urban/Rural; LICs/LMICs/UMICs
- **Education system**, e.g. Formal education/Informal education; type of schooling
- **Population characteristics**, e.g. Gender; Prior level of education
- **Mobile learning intervention category**, e.g. learners-centred/teacher centred
- **Pedagogy**, e.g. game-based

Criteria for determination of independent findings: Sometimes one evaluation leads to several study reports (e.g. working papers and journal articles). Efforts will be made to identify all affiliations between study reports before coding commences, using information on study sample sizes, intervention details, grant numbers, and so on. In cases where multiple reports are found to relate to a single study, I will choose one as the main report (e.g. the publication containing the most complete dataset). When extracting data, the full set of available reports will be used to code each study.

It may also be the case that a single report describes more than one study (e.g. a single publication could describe a series of evaluations conducted in different countries, using

different datasets). In this event, each study will be coded individually as if they had come from separate reports.

Sometimes an intervention will have been evaluated several times. If I find that several evaluations are based on the same data, these will be treated as part of the same study, even if the reports are written by different authors. If the intervention has been evaluated on multiple occasions using different datasets, then I will treat the different reports as separate studies whilst noting their relationship. Should it be unclear whether multiple reports provide independent findings, authors of primary studies will be contacted for clarification.

I intend to use meta-analysis to synthesise results across primary studies. In a single meta-analysis, it is important to include only one effect size measure per study. Estimated treatment effects cannot be regarded as independent of each other if the underlying data are derived from the same sample populations (i.e. some participants contribute information to more than one effect estimate).

Individual effect size estimates may be correlated if, for example, the study analyses:

- Different sub-groups of the treatment population (e.g. young women; highly educated women);
- Outcomes at different times but on the same units (e.g. midline and endline findings);
- Multiple outcome constructs for the same type of outcome (e.g. effects on hourly wages and earnings);
- Multiple interventions with the same sample of participants;
- Multiple treatment groups and the same control group;
- Effects using different methods;
- Several types of treatment effect estimates (e.g. Average Treatment Effect (ATE), Intention to Treat Analysis (ITT));
- Multiple time-points for the same individual (e.g. repeated observations for several follow-up periods); and
- Variations of the above.

In such cases, one of the following approaches will be pursued, as appropriate. Where studies report multiple effect sizes by subgroup I will report data in separate analyses. In the event that I have more than one effect size per outcome construct and study, I will combine different estimates within each study into one effect size per subgroup. Estimating a single (within-study) composite effect size will involve computing a sample-weighted average effect size for each study that accounts for differences in sample size, using appropriate formulae (Borenstein et al 2009). If a study included in the review has used several different techniques to estimate treatment effects for the same outcome (e.g. both statistical matching and regression analysis), the estimate with the lowest risk of bias will be used in the meta-analysis.⁵ In the event that risk of bias assessments are similar, I will choose between different estimation approaches using the approach outlined by Tripney and colleagues (2013) (e.g. comparing the efficiency of the estimator in studies that use both matching and covariate adjustment

⁵ A study may have used the same set of data but different estimation methods and published the results in a single report or in separate multiple reports (e.g. a different report on each of the estimation methods used).

Thematic synthesis:

In the configurative review module, I conducted a thematic synthesis to configure the impact of mobile technologies on education and development in LMICs. The findings of qualitative research studies were synthesised in form of analytical themes on intervention mechanisms and contexts to unpack how and why learning and teaching with mobiles might (or might not) have an impact. I followed Thomas and Harden's (2008) approach to thematic synthesis. They suggest three key stages in thematic synthesis based on thematic analysis in primary research: coding text; developing descriptive themes; and generating analytical themes.

In **stage one**, the reported research findings of the included qualitative studies were subject to line-by-line coding. Findings would ideally have referred to the primary data of each included study (e.g. interview excerpts), but due to limited reporting of this information, authors' analyses and conclusions represented study findings and the unit of analysis in my thematic synthesis. The line-by-line coding feature in EPPI-reviewer was applied to guide and manage the coding of the reported analyses and conclusions. Guidelines to thematic analysis, as applied in qualitative primary research, informed this process of creating thematic codes from the included studies.

In **stage two**, the identified codes were then grouped into descriptive themes. In addition to the inductive creation of descriptive themes from studies' codes, a number of pre-defined (deductive) descriptive themes were introduced in the synthesis and controlled for during line-by-line coding (Appendix 4). These themes relate to common claims and statements in the literature on ML4D, for example that females are discriminated against in access to mobiles. I needed to introduce these deductive descriptive themes in order to identify a possible absence of evidence on these themes, which would have not emerged in a purely inductive thematic synthesis.

In **stage three** of the thematic synthesis, I translated the descriptive themes into analytical themes. This translation is the key process in generating new data in the thematic synthesis. In the context of the thesis' research questions, analytical themes were formulated exclusively around mechanisms and contexts that can configure ML4D's impact or lack of impact. I therefore used a mechanism-context framework to guide my translation of descriptive themes into analytical themes. In this, I adapted Pawson's (2006) definitions of mechanisms and contexts in realist synthesis. I define mechanisms as 'changes caused by the use of mobile technologies that influence its impact'. For example, using mobile phones might change learner's ability to access information, which then might lead to better learning outcomes. Access to information is thus not part of the actual intervention, but a change induced by it that supports its effect. Subsequently, I define contexts as 'variables exogenous to the ML4D intervention that influence its impact'. For example, using mobile phones might only change learners' abilities to access information if a reliable internet connection exists. The contextual factor of a reliable network connection, which is not part of the actual intervention, thus also determines its effect.

I report the findings of the thematic synthesis in narrative tables of all identified analytical themes, divided into mechanisms and contexts, and illustrate the underling descriptive themes.

Mixed-methods synthesis: In a last step of my mixed-methods review, the findings of the aggregative meta-analysis and the configurative thematic synthesis are then brought together in a mixed-methods synthesis to construct the evidence-informed theory of change of ML4D. This construction involved three stages. First, it entailed the plotting of the intervention-to-outcomes pathways in the included ML4D interventions. In this, I used standard templates for developing theories of change in international development (Vogel 2012; Valters 2014). These suggest to break down intervention-to-outcome pathways into: inputs, immediate changes (or outputs), outcomes (on a spectrum from intermediate to final, or short-term to long-term), and impact. Second, I plotted the results of the meta-analysis against the outcomes represented on the theory of change. This provided a visual breakdown for what steps in the theory of change of ML4D there is reliable empirical evidence of effects. Third, I also plotted the mechanism and context themes against the outlined intervention-to-outcome pathways. This aimed to indicate conditions that are associated with the observed effects. Taken together, the theory of change thus presents a visual representation of (i) the assumed links between the provision of mobile devices, their impact on educational outcomes, and subsequently their link to development outcomes; (ii) the evidence of effects for each step in this process of an assumed link between technologies, education, and development; and (iii) the underlying mechanisms and contexts that might explain these links and the observed effects. I then expanded in narrative on the theory of change highlighting what implications the combined systematic review findings have on the conception and positioning of ML4D.

Appendices

Appendix 1: Coding Sheet

NB: Coding was done using EPPI reviewer 4, which allows for a better representation of the dimension of the coding set.

| Code | Answer | Comments |
|---|--|----------|
| ADMIN CODES | | |
| <i>Include/Exclude:</i> | Aggregative Configurative Commentary Exclude | |
| <i>Nature of the study:</i> | Mobile Learning in LICs Mobile learning & Development Mobile learning & Pedagogy & Development | |
| <i>Family of intervention?</i> | Embedded in formal education programme Embedded in informal education programme Embedded in other programme Independent programme | |
| <i>Citation of the study:</i> | | |
| <i>Region:</i> | | |
| <i>Research question:</i> | | |
| <i>Domain of the study:</i> | Academia Grey literature | |
| <i>Type of study:</i> | Journal article Research report Evaluation report Conference paper Book/chapter Thesis/Dissertation | |
| <i>Linked studies</i> | | |
| INTERVENTION CODES | | |
| <i>What intervention is applied (face value)?</i> | | |
| <i>What ML4D approach?</i> | Learner focused teacher focused | |
| <i>Describe the intervention:</i> | | |
| <i>Intervention category:</i> | Literacy Numeracy Teacher Development Extension service Feasibility mHealth Distance education | |

| | | |
|---|--|--|
| | Other | |
| <i>Family of intervention:</i> | Embedded in formal education Embedded in informal education Embedded in other programme independent programme | |
| <i>Identify components if possible:</i> | Provide mobile tech Provide new curriculum/content Provide software/infrastructure Training on tech use Content development Social connections Co-intervention facilitated? (record & specify) | |
| <i>Identify educational components if possible:</i> | 'text' books lesson plans/video lessons quizzes/exercises extra/supplementary information forum reminders educational software | |
| <i>Describe what type of educational content is used?</i> | Local content External content | |
| <i>Who developed the content?</i> | Local Foreign Joint | |
| <i>Was the content used before?</i> | Yes No | |
| <i>What technology is used?</i> | Basic/Feature phone Smart phone Tablets/e-readers Radio PDAs Laptop/netbooks Gaming device MP3 | |
| <i>What application of it?</i> | Voice (interactive) Voice Notes SMS Internet access: WIFI/WLAN Internet access: 3G e-books Social networks Camera/video GPS Games Speech recognition | |
| <i>What makes the intervention mobile?</i> | The technology is potentially mobile The learning/teaching experience is potentially mobile | |
| <i>Where is the mobile context?</i> | Literally learning anytime, anywhere Home use of the device Using the device in a work setting Using the device in a semi-formal setting (eg field trips) Classroom use | |
| <i>How is the tech used?</i> | To teach to learn to administer to connect/communicate | |

| | | |
|---|--|--|
| | To access information Other | |
| <i>Pedagogic approach (if stated)?</i> | Learner-centered Context-based/aware learning Game-based Inquiry-based/experiential Interactivity Collaborative Social transformation Lifelong learning Activity-based Blended learning not stated other | |
| <i>How is learning affected (assumed)?</i> | Additional learning (eg by extending learning to informal contexts) Change in learning pattern (way to access/process information) More fun/increased motivation Connection/Collaboration Personalisation Relevance Organisation Other | |
| <i>How is teaching affected (assumed)?</i> | Additional teaching Change in teaching pattern More fun/increased motivation Connection/Collaboration Efficiency Subject knowledge Other | |
| <i>Subject (if applicable)</i> | Literacy (what language) Numeracy Math English Natural science Medical Other | |
| <i>Interval of learning event?</i> | Anytime Daily (4x plus / week) Weekly Monthly Other | |
| <i>Access (ownership):</i> | Self-owned provided for the duration of project Donation | |
| <i>Access (ratio)</i> | 1:1 shared by household shared by teaching/learning group shared by village/community | |
| <i>Programme theory:</i> | See separate diagram | |
| <i>Identify mechanism if possible:</i> | Contextualisation Affect resource ratio Connectivity Adaptability Better retention Economic returns of tech use Peer support/role model Personalisation Organisation (Reminders) Self-direction Critical thinking Social learning Inter-activity Other | |
| <i>Identify context/assumptions:</i> | Affordability Usability Convenience Ubiquity Ownership Teacher training Local language Perception of tech Lack of access to information Current barriers to education Not stated Other | |
| <i>Addresses current barriers to education:</i> | No education services provided | |

| | | | |
|--|---|--|---|
| | Not enough/sufficient education materials Not enough/insufficient human resources Insufficient educational approach | | |
| <i>How is development assumed to be affected?</i> | | | |
| Outcome codes | | | |
| <i>Which outcomes are targeted?</i> | | (1) Learning outcomes (2) Teaching practice (3) EMIS (4) Empowerment | Subject: |
| <i>Which intermediate outcomes are targeted?</i> | | (A) | |
| <i>What outcome indicators are used?</i> | <i>And how are they measured ?</i> | (1) test scores (A) test scores | (1) pre-defined assessments (2) pre-defined likert scale |
| <i>When were the outcome measured?</i> | | Pre-test: Post-test: | |
| <i>Cost analysis?</i> | | | |
| <i>Do they describe the learning event?</i> | | | Note configurative |
| <i>Reference to development related outcomes?</i> | | | |
| Context codes | | | |
| <i>Who initiated the ML4D programme?</i> | | Intrinsic (poor people) research interest government IGOs NGOs Corporate Other | |
| <i>Where is the intervention conducted?</i> | | LICs LMIC UMICs State country: | |
| <i>How many participants are exposed to the intervention?</i> | | | |
| <i>How many sites?</i> <i>Intervention sites?</i> <i>Experimental sites?</i> | | | |
| <i>Socio-economic?</i> | | Rural | Peri-urban Urban Mixed |
| <i>Informal or formal context?</i> | | Informal | Formal Both |
| <i>Age:</i> | | Children (3-12) Youth (12-25) Adults (25+) Mixed | |

| | | | | |
|---|---|-------------------------------|------------|---|
| <i>Gender:</i> | Male | Female | Both | |
| <i>Literate:</i> | Literate | Semi-literate | Illiterate | |
| <i>If applicable, type of schooling:</i> | Early childhood Special Needs | Primary Extension services | Highschool | Tertiary Private Distance |
| <i>Describe existing educational situation:</i> | | | | |
| <i>Teacher/pupil ratio:</i> | < 30:1 > 30:1 > 50:1 | | | |
| <i>Prevailing educational approach:</i> | Frontal/teacher-centered constructivist Other no info | | | |
| <i>Describe pattern of teacher/learner interaction?</i> | | | | |
| <i>Information on teacher qualification?</i> | Low skilled Skilled Highly skilled | | | |
| <i>Information on educational materials?</i> | | | | |
| <i>Describe supporting infrastructure? Consider</i> | Electricity Safety/storage Network coverage Other | | | |
| <i>Previous exposure to technology? (who/what)?</i> | Yes No | | | |
| <i>Previous perception of tech (if reported?)</i> | Positive Negative Other | | | |
| Findings | | | | |
| <u>Aggregative</u> | | | | |
| <i>Possible include in MA, features</i> | Means SD Size effect Confidence interval sample numbers Other | | | |
| <i>What findings do they report?</i> | | | | |
| <i>ML4D is effective</i> | <u>Impact on learning:</u> _____ Intermediate: Impact on attendance Impact on access Impact on motivation Other <u>Impact on teaching practice:</u> _____ Change in practice Change in behaviour Change in resources Other <u>Impact on EMIS:</u> _____ | | | These are a bit more detailed on EPPI to allow cross-tabs |

| | | |
|---|---|--|
| | <u>Impact on Empowerment:</u> _____ | |
| <i>ML4D has no impact:</i> | <u>Failure to impact on learning :</u> <u>Failure to impact on teaching:</u> <u>Failure to impact on EMIS:</u> <u>Failure to impact on empowerment:</u> | |
| <i>ML4D is harmful (describe):</i> | | |
| <i>ML4D has mixed impact:</i> | Use of same categories as above for effective/ineffective aspects | |
| <i>Checked for retention:</i> | | |
| <i>Cost-benefit:</i> | | |
| <i>Configurative</i> | | |
| <i>Was the tech feasible? (hardware/software)</i> | Yes ('but' if applicable) No ('because') Theft Breakage System crashes/content loss Interface Battery Screen size Memory capacity Need for local language Need for teacher training Need for user training Need for ownership | |
| <i>How has the tech been received?</i> | For Learners: Positive Negative Mixed No information For Teachers: Positive Negative Mixed No information | |
| <i>Is there evidence of teachers/learners appropriating the tech to their contexts?</i> | | |
| <i>Describe the new educational situation?</i> | | |
| <i>Has the teacher/pupil ratio been affected?</i> | | |
| <i>Has the material/pupil ratio been affected?</i> | | |
| <i>Has the educational approach been affected?</i> | Contextualised Game-based Inquiry / experimental Interactive learner-centered critical thinking Social transformation | |

| | | |
|---|---|--|
| | Collaboration Constructivist | |
| <i>Has the teacher/learner interaction been affected?</i> | | |
| <i>Has the learning experience/practice been altered?</i> | Ownership/independence of learning process More motivation/fun/passion etc more efficient Deeper learning (retention/understanding of concepts) More collaborative Connected Relevance Cognitive load Ad hoc access to information Access to more information more self-paced | |
| <i>Has the teaching experience/practice been altered?</i> | Ownership/independence of teaching process More motivation/fun/passion etc more efficient collaboration Better instructions Better preparations Use of local content More subject knowledge Sense of professionalism Easier to handle large groups | |
| <i>Special interest</i> | | |
| <i>Does tech affect groups differently?</i> | Consider age, gender, socio-economic, performance, urban, level of schooling | |
| <i>Reference to 'matching' enabled by mobiles?</i> | | |
| <i>Did participants spend as much time with the devices as expected?</i> | | |
| <i>Is the programme integrated with the national curriculum?</i> | | |
| <i>Is there reference to multipliers/spillovers?</i> | | |
| <i>Dependence on tech?</i> | | |
| <i>Is there reference to how tech affects the role of teachers in the ML4D programme?</i> | Programme bypasses teachers Teachers implement programme Teachers supplement/support tech Tech supplements/supports teachers | |
| <i>Teachers perceive tech as unhelpful:</i> | | |

| | | |
|---|--|--|
| <i>Students encouraged to ask questions:</i> | | |
| <i>Limitations to informal contexts:</i> | | |
| <i>Sense of professionalism:</i> | | |
| <i>Theme of creativity:</i> | | |
| <i>Does excitement for tech wear off?</i> | | |
| | | |
| <i>Development</i> | | |
| <i>Development related findings (describe)</i> | | |
| NOTE: | | |
| | | |
| RATIONALE: | | |
| <i>Describe the underlying case for why mobile learning was needed?</i> | | |

Personal reflections:

Appendix 2: Critical appraisal tool

Mixed-methods critical appraisal tool

| Study type | Methodological appraisal criteria | Response | | |
|--|---|----------|----|-------------------------------|
| | | Yes | No | Comment |
| <p><i>Screening questions: assessing 'fatal flaws' (Dixon-Woods 2005)</i></p> | <p><u>Aggregative assessment:</u></p> <ul style="list-style-type: none"> ✓ Study reports primary data and applied methods ✓ Study reports before and after data¹ ✓ Study features an intervention and control group | | | |
| <p><i>Aggregative 'fatal flaws' based on Stewart et al (2014)</i></p> <p><i>Configurative 'fatal flaws' based on Pawson (2003) TAPUS framework</i></p> | <p><u>Configurative assessment:</u></p> <ul style="list-style-type: none"> ✓ Study reports primary data and applied methods ✓ Study states clear research questions and objectives ✓ Study states clear research design, which is appropriate to address the stated research question and objectives (<i>Purposivity</i>) ✓ The findings of the study are based on collected data, which justify the knowledge claims (<i>Accuracy</i>) | | | |
| <p><i>Screening question based on abstract and/or superficial reading of full-text: Further appraisal is not feasible or appropriate when the answer is 'No' to any of the above screening questions!</i></p> | | | | |
| Study type | Methodological appraisal criteria | Response | | |
| | | Yes | No | Comment / Confidence judgment |
| <p><i>1. Qualitative</i></p> <p><i>e.g.</i></p> <p><i>(A) Ethnography</i></p> <p><i>(B) Phenomenology</i></p> <p><i>(C) Narrative</i></p> | <p>I. RESEARCH IS DEFENSIBLE IN DESIGN (providing a research strategy that addresses the question)</p> <p><u>Appraisal indicators:</u></p> <ul style="list-style-type: none"> ✓ <i>Is the research design clearly specified and appropriate for aims and objectives of the research?</i> | | | |

| | | | | | | | |
|--|--|------------------------|----------------------|---------------------------|---------------------------|--|--|
| <i>(D) Grounded theory (E) Case study</i> | Consider whether | | | | | | |
| | <i>i. there is a discussion of the rationale for the study design</i> | | | | | | |
| | <i>ii. the research question is clear, and suited to qualitative inquiry</i> | | | | | | |
| | <i>iii. there are convincing arguments for different features of the study design</i> | | | | | | |
| | <i>iv. limitations of the research design and implications for the research evidence are discussed</i> | | | | | | |
| | Defensible | Arguable | Critical | Not defensible | <i>Worth to continue:</i> | | |
| | II. RESEARCH FEATURES AN APPROPRIATE SAMPLE (following an adequate strategy for selection of participants) | | | | | | |
| | <u>Appraisal indicators:</u> | | | | | | |
| | Consider whether | | | | | | |
| | <i>i. there is a description of study location and how/why it was chosen</i> | | | | | | |
| <i>ii. the researcher has explained how the participants were selected</i> | | | | | | | |
| <i>iii. the selected participants were appropriate to collect rich and relevant data</i> | | | | | | | |
| <i>iv. reasons are given why potential participants chose not take part in study</i> | | | | | | | |
| Appropriate sample | Functional sample | Critical sample | Flawed sample | <i>Worth to continue:</i> | | | |
| III. RESEARCH IS RIGOROUS IN CONDUCT (providing a systematic and transparent account of the research process) | | | | | | | |
| <u>Appraisal indicators:</u> | | | | | | | |
| Consider whether | | | | | | | |
| <i>i. researchers provide a clear account/description of the process by which data was collected (e.g. for interview method, is there an indication of how</i> | | | | | | | |

| | | | | | | |
|--|---|----------------------------|-------------------------|-----------------------|--|--|
| | <i>interviews were conducted?/procedures for collection or recording of data?)</i> | | | | | |
| | <i>ii. researchers demonstrate that data collection targeted depth, detail and richness of information (e.g. interview/observation schedule)</i> | | | | | |
| | <i>iii. there is evidence of how descriptive analytical categories, classes, labels, etc. have been generated and used</i> | | | | | |
| | <i>iv. presentation of data distinguishes clearly between the data, the analytical frame used, and the interpretation</i> | | | | | |
| | <i>v. methods were modified during the study; and if so, has the researcher explained how and why?</i> | | | | | |
| | Rigorous conduct | Considerate conduct | Critical conduct | Flawed conduct | <i>Worth to continue:</i> | |
| | | | | | | |
| IV. RESEARCH FINDINGS ARE CREDIBLE IN CLAIM/BASED ON DATA (providing well-founded and plausible arguments based on the evidence generated) | | | | | | |
| <u>Appraisal indicators:</u> | | | | | | |
| Consider whether | | | | | | |
| | <i>i. there is a clear description of the form of the original data</i> | | | | | |
| | <i>ii. sufficient amount of data are presented to support interpretations and findings/conclusions</i> | | | | | |
| | <i>iii. the researchers explain how the data presented were selected from the original sample to feed into the analysis process (i.e. commentary and cited data relate; there is an analytical context to cited data, not simply repeated description; is there an account of frequency of presented data?)</i> | | | | | |
| | <i>iv. there is a clear and transparent link between data, interpretation, and findings/conclusion</i> | | | | | |
| | <i>v. there is evidence (of attempts) to give attention to negative cases/outliers etc.</i> | | | | | |
| | Credible claims | Arguable claims | Doubtful claims | Not credible | <i>If findings not credible, can data still be used?</i> | |
| | | | | | | |

| | | | | | | |
|---|---------------------------|--------------------------|------------------------------|---|--|--|
| V. RESEARCH ATTENDS TO CONTEXTS (describing the contexts and particulars of the study) | | | | | | |
| <u>Appraisal indicators:</u> Consider whether | | | | | | |
| <i>i. there is an adequate description of the contexts of data sources and how they are retained and portrayed?</i> | | | | | | |
| <i>ii. participants' perspectives/observations are placed in personal contexts</i> | | | | | | |
| <i>iii. appropriate consideration is given to how findings relate to the contexts (how findings are influenced by or influence the context)</i> | | | | | | |
| <i>iv. the study makes any claims (implicit or explicit) that infer generalisation (if yes, comment on appropriateness)</i> | | | | | | |
| Context central | Context considered | Context mentioned | No context attention | | | |
| | | | | | | |
| VI. RESEARCH IS REFLECTIVE (assessing what factors might have shaped the form and output of research) | | | | | | |
| <u>Appraisal indicators:</u> Consider whether | | | | | | |
| <i>i. appropriate consideration is given to how findings relate to researchers' influence/own role during analysis and selection of data for presentation</i> | | | | | | |
| <i>ii. researchers have attempted to validate the credibility of findings (e.g. triangulation, respondent validation, more than one analyst)</i> | | | | | | |
| <i>iii. researchers explain their reaction to critical events that occurred during the study</i> | | | | | | |
| <i>iv. researchers discuss ideological perspectives/values/philosophies and their impact on the methodological or other substantive content of the research (implicit/explicit)</i> | | | | | | |
| Reflection | Consideration | Acknowledgement | Unreflective research | <i>NB: Can override previous exclusion!</i> | | |

OVERALL DECISCON - EXCLUDE / INCLUDE
 (study generates new knowledge relevant to the review question and complies with minimum criteria to ensure reliability and empirical grounding of knowledge)

Sources used in this section (in alphabetical order); Campbell et al (2003); CASP (2006); CRD (2009); Dixon-Woods et al (2004); Dixon-Woods et al (2006)^{cited in Gough 2012}; Greenhalgh & Brown (2014); Harden et al (2004)^{cited in SCIE & Gough 2012}; Harden et al (2009); Harden & Gough (2012); Mays & Pope (1995); Pluye et al (2011); Spencer et al 2006; Thomas et al (2003); SCIE (2010).

| Study type | Methodological appraisal criteria | | | | Response | | | |
|---|--|--|----------------------------|--|---------------------------------|----|-------------------------------------|---------------------------|
| | | | | | Yes | No | Comment / risk of bias judgment | |
| <p>2. Quantitative (non-randomised; Randomised- Controlled)</p> <p>Common non-random design include:</p> <p>(A) Non-randomised CT (B) Cohort studies (C) Case-control (D) Cross-sectional analytical studies</p> <p>Most common ways of controlling for bias due to baseline confounding:</p> <ul style="list-style-type: none"> • Matching attempts to emulate randomization • Propensity score matching and methods • Stratification where | <p>I. Selection bias: (Are participants recruited in a way that minimizes selection bias?)</p> <p><u>Appraisal indicators:</u></p> <p>Consider whether</p> | | | | | | | |
| | <p>i. there is a clear description of how and why sample was chosen</p> | | | | | | | |
| | <p>ii. there is adequate sample size to allow for representative and/or statistically significant conclusions</p> | | | | | | | |
| | <p>iii. participants recruited in the control group were sampled from the same population as that of the treatment</p> | | | | | | | |
| | <p>iv. group allocation process attempted to control for potential risk of bias</p> | | | | | | | |
| | <p>Low risk of bias</p> | | <p>Risk of bias</p> | | <p>High risk of bias</p> | | <p>Critical risk of bias</p> | <p>Worth to continue:</p> |
| | <p>II. Bias due to baseline confounding: (Is confounding potentially controllable in the context of this study?)</p> <p><u>Appraisal indicators:</u></p> <p>Consider whether</p> | | | | | | | |
| <p>i. the treatment and control group are comparable at baseline</p> | | | | | | | | |

| | | | | | | | |
|--|--|--------------------------|------------------------------|--|---------------------------|------------------------------------|--|
| <i>sub-groups have been compared</i> • Regression analysis where covariates are adjusted for | <i>ii. matching was applied, and in case, featured sufficient criteria</i> | | | | | | |
| | <i>iii. the authors conducted an appropriate analysis that controlled for all potential critical confounding domains</i> | | | | | | |
| | <i>iv. the authors avoided to adjust for post-intervention variables</i> | | | | | | |
| | Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias | <i>Worth to continue:</i> | | |
| IF RANDOMISED CONTROL TRIAL, SKIP I + II AND START HERE! | | | | | | | |
| Bias due to ineffective randomisation: (Is allocation of treatment status truly random?) | | | | | | | |
| <u>Appraisal indicators:</u> Consider whether | | | | | | | |
| <i>i. there is a clear description of the randomisation process</i> | | | | | | | |
| <i>ii. the unit of randomisation and number of participants is clearly stated (pay special attention to treatment and control locations/ balance)</i> | | | | | | | |
| <i>iii. eligibility criteria for study entry are specified</i> | | | | | | | |
| <i>iv. characteristics of baseline and endline sample are provided¹</i> | | | | | | <i>Preferable condition, see 1</i> | |
| Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias | <i>If critical risk of bias, treat as non-random study</i> | | | |
| III. Bias due to departures from intended interventions (Was the intervention implemented as laid out in the study protocol?) | | | | | | | |
| <u>Appraisal indicators:</u> Consider whether | | | | | | | |
| <i>i. the critical co-interventions were balanced across intervention groups</i> | | | | | | | |
| <i>ii. treatment switches were low enough to not threaten the validity of the</i> | | | | | | | |

| | | | | |
|--|--|---------------------|--------------------------|---|
| | <i>estimated effect of intervention</i> | | | |
| | <i>iii. implementation failure was minor and unlikely to threaten the validity of the outcome estimate</i> | | | |
| | <i>iv. it is possible that intervention was taken by the controls (contamination and possible crossing-over)*</i> | | | <i>*whilst challenging in terms of estimating impact, spill-overs might be an important finding in itself (eg teachers read to pupils/village/family members)</i> |
| | <i>v. it is possible that knowledge of the intervention group affects how the two study groups are treated in course of follow-up by investigators?***</i> | | | <i>***consider only in extreme cases in which preferential treatment is clearly evident; blinding in general not expected in social interventions</i> |
| | Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias |
| | <i>Worth to continue:</i> | | | |
| | | | | |
| | IV. Bias due to missing data (attrition) (Are the intervention groups free of critical differences in participants with missing data?) | | | |
| | <u>Appraisal indicators:</u> | | | |
| | Consider whether | | | |
| | <i>i. outcome data are reasonably complete (80% or above)</i> | | | |
| | <i>ii. If 'no', are missing data reported?</i> | | | |
| | <i>iii. If missing data: are proportion of participants and reasons for missing data similar across groups?</i> | | | |
| | <i>iv. If missing data: Were appropriate statistical methods used to account for missing data? (e.g. sensitivity analysis)</i> | | | |
| | <i>v. If not possible to control for missing data, are outcomes with missing data excluded from analysis?</i> | | | |
| | Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias |
| | <i>Worth to continue:</i> | | | |
| | | | | |
| | V. Outcome reporting bias (Are measurements appropriate, e.g. clear origin, or validity known?) | | | |

| | | | | | | |
|--|---------------------|--------------------------|------------------------------|---------------------------|--|--|
| <u>Appraisal indicators:</u> | | | | | | |
| Consider whether | | | | | | |
| i. <i>there was an adequate period for follow up***</i> | | | | | | <i>***in many social science interventions, follow-up is not required to coincide with the start of the treatment; further, longer period of follow up are often required to measure changes. In the context of education, the question of retention – in particular when dealing with short intervention periods (< 1 month) is of major interest.</i> |
| ii. <i>the outcome measure was clearly defined and objective</i> | | | | | | |
| iii. <i>outcomes were assessed using standardised instruments and indicators</i> | | | | | | |
| iv. <i>outcome measurements reflect what the experiment set out to measure</i> | | | | | | |
| v. <i>the methods of outcome assessment were comparable across experiential groups</i> | | | | | | |
| Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias | <i>Worth to continue:</i> | | |
| | | | | | | |
| VI. Bias in selection of results reported (Are the reported outcomes consistent with the proposed outcomes at the protocol stage?) | | | | | | |
| <u>Appraisal indicators:</u> | | | | | | |
| Consider whether | | | | | | |
| i. <i>it is unlikely that the reported effect estimate is available primarily because it was a notable finding among numerous exploratory analyses</i> | | | | | | |
| ii. <i>it is unlikely that the reported effect estimate is prone to selective reporting from among multiple outcome measurements within the outcome domain</i> | | | | | | |
| iii. <i>it is unlikely that the reported effect estimate is prone to selective reporting from among multiple analyses of the outcome measurements</i> | | | | | | |

| | | | | | |
|--|--|--------------|-------------------|-----------------------|--|
| | iv. the analysis includes an intention to treat analysis? (If so, was this appropriate and were appropriate methods used to account for missing data?)**** | | | | ****usually in clinical RCTs, rare in social science: only rate if conducted |
| | Low risk of bias | Risk of bias | High risk of bias | Critical risk of bias | |

OVERALL RISK OF BIAS:

Sources used in this section (in weighted order): Cochrane (2014); Stewart et al (2014); Stewart et al (2012); Higgins et al (2011); Greenhalgh & Brown (2014); Pluye et al (2011); Gough et al (2007)^{Weight of evidence thingi}

| Study type | Methodological appraisal criteria | Response | | |
|---|---|----------|----|------------------------------|
| | | Yes | No | Comment /confidence judgment |
| <p>3. Mixed-methods²</p> <p><u>Sequential explanatory design</u> The quantitative component is followed by the qualitative. The purpose is to explain quantitative results using qualitative findings. E.g., the quantitative results guide the selection of qualitative data sources and data collection, and the qualitative findings contribute to the interpretation of quantitative results.</p> <p><u>Sequential exploratory design</u> The qualitative component is followed by the quantitative. The purpose is to explore, develop and test an instrument (or taxonomy), or a conceptual framework (or theoretical model). E.g., the qualitative findings inform the quantitative data collection, and the quantitative results allow a generalization of the qualitative findings.</p> <p><u>Triangulation designs</u> The qualitative and quantitative components are concomitant. The purpose is to examine the same phenomenon by interpreting qualitative and quantitative results (bringing data analysis together at the interpretation stage), or by integrating qualitative and quantitative datasets (e.g., data on same cases), or by</p> | <p>I. RESEARCH INTEGRATION/SYNTHESIS OF METHODS (assessing the value-added of the mixed-methods approach)</p> <p>Applied mixed-methods design:</p> <ul style="list-style-type: none"> ○ Sequential explanatory design ○ Sequential explorative design ○ Triangulation design ○ Embedded design <p><u>Appraisal indicators:</u></p> <p>Consider whether</p> | | | |
| | <p>i. the rationale for integrating qualitative and quantitative methods to answer the research question is explained [DEFENSIBLE]</p> | | | |
| | <p>ii. the mixed-methods research design is relevant to address the qualitative and quantitative research questions, or the qualitative and quantitative aspects of the mixed methods research question</p> | | | |

| | | | | |
|---|--|--|--|--|
| <p><i>transforming data (e.g., quantization of qualitative data).</i></p> <p><u>Embedded/convergent design</u> <i>The qualitative and quantitative components are concomitant. The purpose is to support a qualitative study with a quantitative sub-study (measures), or to better understand a specific issue of a quantitative study using a qualitative sub-study, e.g., the efficacy or the implementation of an intervention based on the views of participants.</i></p> | <p><i>[DEFENSIBLE]</i></p> | | | |
| | <p><i>iii. there is evidence that data gathered by both research methods was brought together to inform new findings to answer the mixed-methods research question (e.g. form a complete picture, synthesise findings, configuration)</i></p> <p><i>[CREDIBLE]</i></p> | | | |
| | <p><i>iv. the approach to data integration is transparent and rigorous in considering all findings from both the qualitative and quantitative module (danger of cherry-picking)</i></p> <p><i>[RIGOROUS]</i></p> | | | |
| | <p><i>v. appropriate consideration is given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results)?</i></p> <p><i>[REFLEXIVE]</i></p> | | | |

For mixed-methods research studies, each component undergoes its individual critical appraisal first. Since qualitative studies are either included or excluded, no combined risk of bias assessment is facilitated, and the assigned risk of bias from the quantitative component similarly holds for the mixed-methods research.

The above appraisal indicators only refer to the applied mixed-methods design. If this design is not found to comply with each of the four mixed-methods appraisal criteria below, then the quantitative/qualitative components will individually be included in the review:

| | | |
|--|--|---|
| <p><u>Mixed-methods critical appraisal:</u></p> <ol style="list-style-type: none"> 1. Research is defensible in design 2. Research is rigorous in conduct 3. Research is credible in claim 4. Research is reflective | <p><u>Qualitative critical appraisal:</u></p> <p>Include / Exclude</p> | <p><u>Quantitative critical appraisal:</u></p> <ol style="list-style-type: none"> 1. Low risk of bias 2. Risk of bias 3. High risk of bias 4. Critical risk of bias |
|--|--|---|

Combined appraisal:
 Include / Exclude mixed-methods findings judged with _____ risk of bias

Section based on Pluye et al (2011). Further sources consulted (in alphabetical order): Creswell & Clark (2007); Crow (2013); Long (2005); O’Cathain et al (2008); O’Cathain (2010); Pluye & Hong (2014); Sirriyeh et al (2011).

¹Two theoretical exceptions to this rule apply:

- i) A RCT with appropriate randomization procedure can be included without showing baseline data, as both experimental groups can be assumed to be equal at baseline by design.
- ii) A sophisticated quasi-experimental design such as PSM or RDD in theory could make the same claim to not require baseline data.

In both cases, the advice of an evaluation specialist will be sought as the researcher does not have the capacity to make an informed judgment in such specialist cases.

²The mixed-methods Critical Appraisal is facilitated for studies applying an explicit mixed-methods approach. The component is applied in addition to criteria for the qualitative component (I to VI), and appropriate criteria for the quantitative component (I to VI).

Appendix 3: Effects size calculations

Cohen's *d* as well as hedge's *g* will be calculated for each included study, with *g* being the effect size used in the meta-analysis. The usage of *g* has become established practice in statistical synthesis due to its ability to adjust for small sample bias prevailing in *d* (Deeks et al 2001). The following formula will be applied to calculate *g* and its standard error (SE):

$$g_i = \frac{m_{1i} - m_{2i}}{s_i} \left(1 - \frac{3}{4N_i - 9} \right)$$

with SE,

$$SE(g_i) = \sqrt{\frac{N_i}{n_{1i}n_{2i}} + \frac{g_i^2}{2(N_i - 3.94)}}.$$

where, m_i denotes the means values of experimental groups, s_i denotes the pooled standard deviation (SD), N_i denotes the combined sample, N_{1i} and N_{2i} the respective experimental groups.

The pooled SD will be calculated using,

$$s_i = \sqrt{\frac{(n_{1i} - 1)SD_{1i}^2 + (n_{2i} - 1)SD_{2i}^2}{N_i - 2}}$$

where, in addition to the above, SD_{1i} and SD_{2i} denote the respective SD of experimental groups. The pooled SD is the preferred value to calculate *g*. The size of *g* using the SD of the control group however will be reported in the appendices.

A common challenge in meta-analysis of continuous outcomes is whether to base one's effect size calculation on endline mean values or on the change between mean values from baseline to endline. Ideally, since this systematic review envisages to identify a majority of studies based on a quasi-experimental designs and thereby compromising the comparability of experimental groups somewhat, this meta-analysis hopes to calculate hedge's *g* using changes between mean values. However, reporting of this value, and in particular its SD, is rare in the target literature.

The inverse variance method will be used to conduct the meta-analysis⁶. Given the assumed heterogeneity in the true effects of studies across e.g. geographical reach, educational systems, and socio-economic contexts, a random effect model will be applied. However, fixed effects will also be calculated and reported in the appendices. EPPI reviewer 4's meta-analysis online tool will be used to calculate effect sizes (*g*) per study. The software will also be used to calculate the pooled effect size, construct forest and funnel plots, and lastly, to test for heterogeneity.

⁶ A detailed formula will be reported in the thesis.

Appendix 4: Deductive themes for thematic synthesis

Pre-defined descriptive themes include:

- *Perceptions/acceptance* of mobile technologies in education
- *Connectivity* of learners & teachers to peers/outside information
- *Personalisation/Ownership* of teaching/learning process
- *Appropriation* of the mobile technology
- *Mastery* of the mobile technology (pride; confidence; self-efficiency)
- *Relationship* between teacher and technology in the learning process
- *Authority/power relations* (students asking questions/challenging teacher)
- *Embeddedness* of the mobile technology in educational process
- *Change in pedagogy* (do teachers use the technology to teach in a new way?)
 - Interaction between teachers and learners
 - Participation of learners in the educational process
 - Collaboration/contextualisation
- *Change in teaching experience* (motivation, perceptions of teaching profession)