The use of interactive maths apps to support early mathematical development in UK and Brazilian primary school children

Laura Ann Outhwaite

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2018

Abstract

Developing strong mathematical skills at the start of primary school is vital for children's later learning and development. Educational maths apps delivered on touch-screen tablet devices are suited to primary education and an emerging evidence base demonstrates their potential to support young children's early mathematical development. The educational maths apps at the focus of thesis draw on the principles of active, engaged, meaningful, and socially interactive learning combined with curriculum-based content and specific learning goals. Using a mixed-methods, pragmatic approach and a multi-level, ecological, determinant framework, this thesis aimed to address how does app-based mathematics instruction work, who does it work for, and under what circumstances does it work?

The UK Proof of Concept study (Chapter 4) showed children aged 4-5 years identified as low-achievers in mathematics (n = 12) who used the apps for 8 weeks made greater learning gains in mathematics compared to their typically attaining peers receiving standard mathematical practice (n = 15). No significant effect of children's socio-economic status was found but children with a poor memory capacity demonstrated stronger learning gains with the apps.

Similar results were found in the Brazil Proof of Concept study (Chapter 5). After a 10-week intervention period, results showed children aged 5-6 years made greater progress in mathematics with

the apps when delivered in their first (Brazilian Portuguese, *n* = 23) or second language (English, *n* = 20), compared to standard mathematical practice (*n* = 19). However, proficiency in the language of instruction was associated with increased learning gains.

To further evaluate the maths apps as a form of quality mathematics instruction, a pupil-level randomised control trial was conducted in the UK with 389 children aged 4-5 years old (Chapter 6). Consistent with previous results, children randomly assigned to use the maths apps for 30 minutes a day, over 12 weeks either as a supplementary teaching aid (*n* = 126) or instead of one daily mathematics activity (*n* = 131) made significantly greater learning gains compared to children who received standard mathematical practice (*n* = 132).

Results from the UK RCT Implementation Evaluation (Chapter 7) indicated that in order to maximise children's learning with the apps the implementation needs to be well organised, for example, integrating the apps into a consistent routine with well-organised equipment and a dedicated space in the classroom.

Overall, the evidence presented in this thesis suggests the maths apps can provide an individualised and targeted form of mathematics practice that can add value to a well-balanced early years curriculum and contribute to the long-term journey towards ensuring quality education for all.

Published articles

Work from this thesis has been published in the following articles:

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers & Education*, *108*, 43-58.

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2018). Raising early achievement in math with interactive apps: a randomized control trial. *Journal of Educational Psychology*. doi:10.1037/edu0000286.

Conference presentations

Work from this thesis was presented at the following conferences:

Outhwaite, L.A., Faulder, M., Gulliford, A., & Pitchford, N.J. (2018). Raising early achievement in math with interactive apps: a randomized control trial. Presented at *EARLI SIG 5 Conference,* Freie Universität, Berlin, Germany.

Outhwaite, L.A., Faulder, M., Gulliford, A., & Pitchford, N.J. (2017). Apps for All: Using tablet technology to raise maths attainment in disadvantaged children in the UK. Presented at *World Conference on Computers in Education,* Dublin, Ireland.

Pitchford, N.J., & **Outhwaite, L.A.** (2016). Unlocking Talent: Lessons learnt from Malawi and the UK. Presented at *Education Foundations Conference,* University of Malawi, Chancellor College, Zomba, Malawi.

Pitchford, N.J., & **Outhwaite, L.A.** (2016). Unlocking Talent: Lessons learnt from Malawi and the UK. Presented at *British Association for International and Comparative Education (BAICE) Conference*, University of Nottingham, UK.

Funding

This PhD was financially supported by the Economic and Social Research Council including additional funding from the Overseas Fieldwork Fund to support the project in Brazil.

Acknowledgements

Firstly, to my supervisors Professor Nicola Pitchford and Anthea Gulliford, thank you for guiding my PhD, for all your advice and constructive feedback, and for teaching me invaluable lessons in research. Thank you Nikki, for seeing my potential back in 2013 and for going above and beyond to support my academic career.

To Andrew Ashe and the *onebillion* team, thank you for supporting and investing in this research. Marc Faulder, thank you for welcoming me into your school and your professional network. Dr Maria Neves, thank you, for all your time and effort in supporting our project in Brazil. It has been a pleasure working with you all.

Thank you to all the schools, teachers, children, and parents who have taken part in this research. I am incredibly grateful for your time, expertise, and co-operation. A very special thank you to Rozário, for making Brazil feel like a second home. This research ultimately would not have been possible without you.

Thank you to my closest friends, especially Ali, Camilla, Lia, Katherine and Chiara. Thank you for always encouraging and believing in me.

Thank you to my family, especially my parents Chris and Jef, my grandparents Eileen and Nat, and my brother Sean. Thank you for unconditionally supporting and empowering me throughout this journey and beyond. Above all, thank you for building a life that has both encouraged and enabled me to achieve my goals.

Table of Contents

Chapter 1: Literature Review 1

Background 1

Section 1: Mathematical Development and Learning 5

Theories of mathematical development 5

Componential approach to mathematical development 5

Factors that influence mathematical development 7

Memory capacity and mathematics 8

Socio-economic status and mathematics 10

Language proficiency and mathematics 11

Early intervention 14

Theories of mathematical learning 15

Targeted practice and direct instruction 15

Play-based pedagogies 17

App Technology 19

Science of Learning principles 20

App technology affordances 24

Previous research 25

Section 2: Evidence-based education 26

Developing and scaling interventions 26

Initial evidence 27

Pilot studies 27

Efficacy trial 28
Effectiveness trial 28
Scale-up 29
RCTs in evidence-based education 30
Understanding implementation 33
Defining implementation 33
Quality 36
Adaptations 36
Participant responsiveness 37
Implementation variability and intervention outcomes 37
Implementation theories 38
Determinant frameworks 39
Conclusion 43
Chapter 2: Systematic Review 45
Background 45
Methods 48
Objective 48
Inclusion and exclusion criteria 48
Search strategy 56
Study selection 57
Study quality assessment 58
Identification 59

Screening 59
Eligibility 59
Included 59
Results 67
Study characteristics 67
Narrative synthesis 69
Discussion 88
Limitations of current review 89
Future research directions 90
How does the intervention work? 90
Who does the intervention work for? 92
Under what circumstances does the intervention work? 93
Thesis aims 94
Epistemological position 98
Conclusion 99
Chapter 3: Maths App Intervention 101
Overview of maths app intervention 101
Maths app curriculum content 104
Instructional features 114
Active learning 114
Engaged learning 115
Meaningful learning 115

Socially interactive learning 116

Support for language proficiency and working memory..... 121

Theories of learning 121

Chapter 4: UK Proof of Concept Study 123

Current study 123

Mathematical skill acquisition 124

Child-level influencing factors..... 125

Methods 126

Design 126

Participants..... 128

School context 129

Maths app intervention..... 130

Small group mathematics instruction 131

Whole class embedded mathematics activities..... 131

Mathematics assessments..... 132

Memory assessments..... 134

Socio-economic status..... 135

Procedure..... 136

Results..... 142

Mathematical skill acquisition 142

Child-level influencing factors..... 145

Discussion 147

Mathematical skill acquisition 148
Child-level factors 149
Limitations 152
Conclusion 154
Chapter 5: Brazil Proof of Concept Study 155
Current study 155
Mathematical skill acquisition 155
Language of instruction proficiency 156
Methods 157
Design 157
Participants 159
School context 161
Maths app intervention 163
Standard mathematics practice 173
Mathematics assessment 173
Progression through the maths apps 174
Language proficiency 175
Procedure 175
Results 179
Bilingual status of sample 180
Primary outcome variable: mathematics gains 180

Secondary outcome variable: progression through the maths apps 181
Language of instruction proficiency 181
Discussion 184
Mathematics learning 185
Language of instruction 186
Limitations 186
Conclusion 189
Chapter 6: UK Feasibility RCT 191
Current study 191
Effectiveness and implementation 191
Components of mathematical development 193
Method 195
Design 195
Participants 197
Maths app intervention 202
Small group mathematics instruction 202
Whole class embedded mathematics activities 203
Mathematics assessment 204
Procedure 205
Results 211
Preliminary analysis 211

Effectiveness and implementation 212
Components of mathematical development 213
Discussion 218
Effectiveness and implementation 218
Components of mathematical development 219
Limitations 220
Conclusion 223
Chapter 7: UK RCT Implementation Evaluation 224
Current study 224
Understanding implementation 225
Implementation themes and intervention outcomes 226
Aim 1: Understanding Implementation 226
Methods 226
Qualitative methodology 226
Narrative, direct observations 227
Self-report interviews 228
Assimilation of narrative, direct observations and self-report interviews 228
Participants 229
School context 230
Qualitative data collection protocols 236
Qualitative data analysis 242

Results & Discussion 247

Teacher level 247

Classroom level 248

Child level 248

Aim 2: Implementation themes and intervention outcomes 249

Methods 250

Structured judgment methodology 250

Procedure 250

Results 259

Preliminary analyses 259

Implementation themes and learning gains 259

Discussion 266

Classroom level 266

Teacher and child level 268

Future directions 269

Conclusion 270

Chapter 8: General Discussion 272

Are the maths apps an effective form of mathematics instruction? 275

Time on task 275

Under what circumstances do the maths apps work? 279

Comparing forms of implementation 279

Insights into implementation 280
Educational policy and practice implications 280
How do the maths apps work? 281
Mathematical development 282
Factors that influence mathematical development 284
Science of Learning principles 288
Theoretical implications 289
Who do the maths apps work for? 291
Socio-economic status 291
Low-achievers. 292
Educational policy and practice implications 295
Study quality assessment 296
Impact evaluation 299
Immediate follow-up evaluation 300
Two-year follow-up evaluation 301
Limitations and future research 302
Long-term follow-up 302
Additional benefits 303
Home learning environment 305
Teacher's belief system 306
Conclusion 306
References 309

Appendix 363

List of Tables

Table 1 Eight distinct but related dimensions of implementation (After: Durlak & DuPre, 2008; Forman et al., 2009)..... 35

Table 2 Inclusion and exclusion criteria used to select studies for the systematic review. 52

Table 3 Search terms included search strategy. Full search strategy code reported in Appendix A. 57

Table 4 The design of the quality assessment tool including definitions of different sources of bias in educational intervention research studies. 60

Table 5 Quality assessment of the seven studies included in the systematic review..... 65

Table 6 Analytical data extracted from the seven studies included in the systematic review. Within-group effects (Cohen's *d*) calculated from pre-test and post-test intervention group mean data reported in the individual studies..... 71

Table 7 Summary of general and specific research questions to be addressed in this thesis..... 96

Table 8 Maths apps content mapped to the UK mathematics curriculum for EYFS and KS1 and components of mathematical development..... 106

Table 9 Maths apps mapped to the principles of active, engaged, meaningful, and socially interactive learning. 117

Table 10 Description of maths intervention received and total maths-learning exposure for each group across the 16-week intervention period. 128

Table 11 Descriptive data, including mean age in months (SD, Min-Max), Gender (Female: Male) and mean IDACI score (SD, Min-Max) for the study sample and each participating group. 130

Table 12 Assessment protocol..... 139

Table 13 Description of intervention implementation over the 16-week intervention period. 141

Table 14 Raw score group mean (SD) at pre-test, post-test, and gain score (post- minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95% confidence interval) for each assessment of mathematical ability and composite short-term memory scores. 143

Table 15 Descriptive data, including mean age in months (SD, Min-Max), Gender (Female: Male) and mean L1 and L2 proficiency (SD, Min-Max) for the final study sample and each participating group. 160

Table 16 Comparison of the school daily routines, including total weekly time learning mathematics for the two treatment groups and control group. 162

Table 17 Maths apps content mapped to the mathematics curriculum and attainment measure..... 164

Table 18 Raw score group mean (SD) at pre-test and post-test with group mean (SD) for difference scores (post-test minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95%

confidence interval) for the assessment of mathematical ability (max score 54) and group mean (SD) and range (Min-Max) for number of app topics completed..... 183

Table 19 Description of maths intervention received and total maths-learning exposure for each group across the 12-week intervention period. 196

Table 20 CONSORT (2010) table describing the composition of the study sample at each stage of the RCT..... 199

Table 21 Descriptive data, including mean age in months (SD, Min-Max) and Gender (Female: Male) for the final sample in the trial.201

Table 22 Raw score group mean (SD) at pre-test and post-test with group mean (SD) difference scores (post-test minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95% confidence interval) for the assessment of mathematical ability and each component of mathematical proficiency. 215

Table 23 Descriptive data for participating teachers..... 230

Table 24 Profiles of the participating schools..... 232

Table 25 Guiding question used in the self-report interviews 238

Table 26 Details of the five measures taken to ensure qualitative validity and reliability..... 240

Table 27 Qualitative data analysis process followed (Creswell & Creswell, 2018; Clarke, 2004)..... 243

Table 28 Definitions for high, medium, and low ratings with guiding questions for each of the implementation themes. 253

Table 29 Within-group effect sizes of progress with the maths apps collapsed across Group 1 (treatment) and Group 2 (time-equivalent treatment)..... 261

Table 30 Group mean rating (SD, Min-Max) for each implementation theme at the teacher, classroom, and child level. Spearmans' rho correlations between implementation themes and learning gains. 263

Table 31 Model fits for the exploratory linear regression identifying significant predictors of learning gains. 265

Table 32 Summary of methodologies used and findings from each of the three quantitative studies..... 276

Table 33 Re-analysis of 'low-achievers' sub-sample in the UK Feasibility RCT (Chapter 6) 294

Table 34 Study quality assessment results for the experimental, quantitative studies reported in Chapters 4-6..... 297

List of Figures

Figure 1 Multi-level, ecological, determinant framework based on Bronfenbrenner's Ecological Systems Theory (1979). 42

Figure 2 PRISMA 2009 flow diagram of information throughout the systematic review (Moher et al., 2009) 59

Figure 3 Example item and task instructions for Topic 1, Sorting and Matching. 103

Figure 4 Example item and task instructions for Topic 1, Sorting and Matching with Brazilian Portuguese and English verbal instructions..... 172

Figure 5 Thematic map summarizing the six implementation themes identified from the qualitative observation and interview data..... 246

Figure 6 Summary of study results within the multi-level, ecological, determinant framework 274

List of Appendices

Appendix A Search strategy code used in the Systematic Review reported in Chapter 2..... 363

Appendix B Descriptive data extracted from the seven included studies reported in the Systematic Review in Chapter 2. 364

Appendix C Paper-and-pen assessment of Mathematical Curriculum Knowledge used in the UK Proof of Concept Study reported in Chapter 4. 366

Appendix D Assessment of child's language proficiency used in the Brazil Proof of Concept Study reported in Chapter 5..... 383

Appendix E Teacher Manual provided to all schools participating in the UK Feasibility RCT reported in Chapter 6. 386

Appendix F Teacher Feedback Questionnaire (see Chapter 8)... 404

Appendix G Two year follow-up evaluation Teacher Feedback Questionnaire (see Chapter 8) 409

Chapter 1: Literature Review

Background

All children have the right to education (United Nations Convention on the Rights of the Child, 1989; Article 28) in order to achieve their full potential (UN-CRC, 1989; Article 29.1). This includes access to inclusive and equitable quality education, as articulated in the Sustainable Development Goal 4 (United Nations, 2015). However, there are significant challenges to achieving this, both in the developed and developing world, particularly in mathematics. Recent statistics show an estimated 250 million children worldwide do not have the basic mathematical skills needed to live a healthy and productive life and contribute to economic growth (UNESCO-GMR, 2014). Many of these children live in developing countries, such as Brazil, where 47% of primary school aged children do not progress beyond the basic skill level for mathematics (UNESCO, 2011).

However, underachievement in mathematics is also evident in high-income countries, such as the United Kingdom (UK). In the latest PISA assessment of 15-year olds' maths ability, the UK ranked 27th out of the 34 participating countries (OECD, 2016). Furthermore, the development of early mathematics can lag behind early literacy achievement in the same group of children, indicating a significant discrepancy in the development of these two skills at the start of formal education (Pitchford, Papini, Outhwaite, &

Gulliford, 2016; Department for Education, 2016). Research also identifies a mathematics practice gap in education, relative to literacy skills (Stacy, Cartwright, Arwood, Canfield, & Kloss, 2017). In part, this may be due to the provision of a phonics-based literacy intervention in all primary schools in the country (Department for Education, 2010; 2016), whereas a similar strategy for mathematics is not currently implemented.

However, developing a strong foundation in mathematics is vital for a child's later educational success (Duncan et al., 2007; Geary, 2011a) and is associated with economic, health, and employment outcomes (Reyna, Nelson, Han, & Diechmann, 2009). In particular, longitudinal research suggests early mathematical skills at age 4 years predicts mathematical achievement through to age 15, even after controlling for early cognitive skills, reading ability, and family and individual child characteristics (Watts, Duncan, Siegler, & Davis-Kean, 2014). Furthermore, studies suggest early mathematical skills support the learning of other subjects, such as literacy (Razel & Eylon, 1990; Sarama, Lange, Clements, & Wolfe, 2012).

Collectively this evidence highlights the vital importance of developing strong mathematics skills at the start of schooling. In response, recent research and policy in the UK calls for a greater focus on learning mathematics at the start of school (APPG, 2014; Ofsted, 2018; Pitchford et al., 2016) through effective, efficient, and evidence-based interventions to support early mathematical

development and raise achievement (Butterworth, Varma, & Laurillard, 2011; Jordan & Levine, 2009). This is also reflected in international policy calling for educational interventions that can help improve children's learning outcomes in low and middle-income countries (LMICs), such as Brazil (UNESCO, 2011).

Educational maths applications (apps) delivered on touch-screen tablet devices offer an opportunity to support the development of early mathematical skills (Stacy et al., 2017). By combining the benefits of direct instructional theory (Kirschner, Sweller, & Clarke, 2006) and play-based pedagogies (Gray, 2015), they have the potential to deliver a child-centred, individualised, and scaffolded learning environment (Adams & Carnine, 2003; Grant, Tamim, Brown, Sweeney, Ferguson, & Brown, 2015; Mayo, 2009; Mayer, 2004). It is a pedagogical approach that is gaining popularity in educational settings worldwide, with schools and governments increasingly investing in such technologies (Clarke, 2014; Tamim, Borokhovski, Pickup, & Bernard, 2015). In particular, UK Educational Sectary, Damian Hinds recently called for "new technology to spearhead classroom revolution" through the use of innovative technology-based teaching practices that can support access, inclusion, and improve learning outcomes for all children (Department for Education, 2018).

However, concerns have also been raised about the impact of technology-based screen time on early child learning and development (Greenfield, 2015; Palmer, 2007; Sigman, 2012).

Formal research examining the impact of educational maths apps on children's mathematical development and the processes of scaling such interventions is nevertheless sparse. Research is needed in this area not only to inform theoretical understandings of how children learn with app-based learning environments but is also critical to shape opinions and inform evidence-based educational policy and practice. In response, this thesis aims to examine the reach and utility of a new maths app intervention in supporting the acquisition of early mathematical skills for young children at the start of formal education in the UK, a high-income country, and Brazil, a middle-income country.

Chapter 1 will review the relevant literature in two sections: (1) Mathematical development and learning and (2) Evidence-based education. Section 1 will discuss how mathematics develops from basic mathematical skills to higher-level mathematical skills. It will also present the current evidence on the influence of memory capacity, socio-economic status, and language proficiency on early mathematical development. This section will next consider key theories of learning mathematics. It will focus on targeted practice, direct instruction, and play-based pedagogies in order to consider how these theories can apply to app-based learning. This section will also discuss Science of Learning principles to expand the theoretical underpinnings of learning with educational maths apps. Finally, the prevalence, popularity, and potential benefits of app technology, as well as, the emerging evidence base will be discussed. A systematic review of all previous research in the area

of app-based mathematics instruction to support early mathematical development at the beginning of formal education is reported in Chapter 2.

Section 2 will consider the role of evidence-based education and discuss the process of developing and scaling educational interventions. This section will position the studies reported in this thesis (Chapters 4-7) within this broader scaling framework. Furthermore, it will consider the benefits, limitations, and critiques of different research designs, particularly issues surrounding the use of randomised control trial (RCT) designs in education. The role of implementation science and the value of mixed-methods to support a holistic evaluation of app-based mathematics instruction will also be presented.

Section 1: Mathematical Development and Learning

Theories of mathematical development

When examining interventions designed to support early mathematical learning it is important to understand the theoretical underpinnings of mathematical development.

Componential approach to mathematical development

The componential approach suggests mathematical development requires the acquisition of different component skills and processes that range in level of difficulty (Goswami, 2006; Holmes & Dowker, 2013). While there is no definitive list of components of mathematical knowledge (Delazer, 2003),

components can be grouped into four broad categories: (1) factual knowledge (for example, number bond combinations and the properties of shape and patterns) and (2) conceptual understanding (for example, identifying and applying mathematical procedures). Both of these components encompass basic mathematical skills. In contrast, (3) mathematical reasoning (for example, making deductions and inferences from mathematical information) and (4) problem solving (for example, combining and applying different areas of mathematics to solve a problem in a specific context) reflect higher-level mathematical skills, which require the application of basic mathematical knowledge to find a solution (Rutherford-Becker & Vanderwood, 2009; Thurber, Shinn, & Smolkowski, 2002).

Research shows the acquisition and automatisation of basic mathematical skills facilitates higher-level mathematical development (Codding, Archer, & Connell, 2010; Gersten & Chard, 1999; Mayfield & Chase, 2002; Van Der Heyden & Burns, 2008). For example, longitudinal research shows early factual and conceptual knowledge, including, number bond combinations, counting, pattern knowledge, shape composition, and calculation ability predict more complex skills, such as problem solving later in child development (Björn, Aunola, & Nurmi, 2016; Clements & Sarama, 2007; Fuchs et al., 2006; Rittle-Johnson, Fyfe, Hofer, & Farran, 2016). In contrast, poor fluency in basic maths skills is shown to be commonly associated with mathematical difficulties (Geary, 1993; Geary & Hoard, 2005; Jordan, Hanich, & Kaplan, 2003).

However, further research emphasises the importance of not viewing the different components of mathematics as a strict hierarchy. For example, children may perform well on an apparently difficult word problem solving task, while at the same time show poor performance on an apparently easier basic mathematical factual task, such as counting (Denvir & Brown, 1986; Holmes & Dowker, 2013).

Nevertheless, the current evidence supports theories of cumulative learning that propose the mastery of basic mathematical facts and concepts provide the essential foundation for the acquisition of higher-order, more complex mathematical skills (Gagné, 1968). It also emphasises the vital, foundational role of strong basic mathematical skills in successful mathematical development. However, many children struggle to acquire basic mathematical facts and conceptual knowledge (Geary, 2011b). This makes them vulnerable to persistent underachievement throughout their education (Duncan et al., 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009).

Factors that influence mathematical development

Children may also be vulnerable to underachievement in mathematics due to a range of other factors known to be associated with early mathematical development. This thesis will focus on: (1) poor working memory capacity as evidence suggests individual differences in working memory capacity account for approximately 25% of the variance in mathematical outcomes (Cragg, Keeble,

Richardson, Roome, & Gilmore, 2017), (2) low socio-economic status as research suggests children from lower SES backgrounds underachieve in mathematics, relative to their higher-SES peers (Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006), and (3) low levels of language proficiency in the language of instruction in the context of bilingual learners as evidence suggests poor language proficiency can hinder children's learning opportunities and achievements (Kempert, Saalbach, & Hardy, 2011; Snow & Uccelli, 2009).

Memory capacity and mathematics

Research suggests working memory capacity is closely associated with mathematics ability (Alloway, Alloway, & Wootan, 2014; LeFevre, DeStefano, Coleman, Shanahan, & Campbell, 2005; Peng, Wang, & Namkung, 2018). Working memory involves controlling, regulating, and actively maintaining relevant information to complete complex cognitive tasks, such as mathematics (Miyake & Shah, 1999). For example, in the context of mathematical learning, working memory allows children to hold critical information, such as interim totals in mind, keep track of counting steps, and retrieve number facts from memory (Bull & Lee, 2014). It also allows children to hold vital task instructions in mind, which are necessary for completing the learning activity (Gathercole, Lamont, & Alloway, 2006).

Young children have a very small working memory capacity, for example studies show children aged 5 years of age have a typical working memory capacity of two unrelated items (Alloway,

2011). There are also substantial individual differences in working memory capacity within the same age group of children (Alloway, 2006). Studies show these individual differences in working memory capacity can explain approximately 25% of the variance in mathematical outcomes (Cragg et al., 2017), above and beyond other cognitive measures of intelligence (Gathercole & Alloway, 2006; Raghubar, Barnes, & Hecht, 2010). This may be explained by evidence suggesting memory storage capacity in children underachieving in mathematics is used less efficiently compared to their peers (Passolunghi & Siegel, 2004).

While an association between working memory capacity and mathematics is well established (LeFevre et al., 2005), evidence of the nature and direction of this relationship is sparse (Raghubar et al., 2010). Working memory training involving the repeated practice of computer based adaptive tasks has been proposed to support children's mathematical development. This is based on the hypothesis that poor working memory causes poor mathematical ability and so improvements in working memory should transfer to improvements in mathematics (Melby-Lervåg & Hulme, 2013). Although some studies evaluating working memory training demonstrate some benefits (Holmes, Gathercole, & Dunning, 2009; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010), these benefits are typically restricted to performance on the specifically trained working memory task (Melby-Lervåg & Hulme, 2013) and do not transfer to improvements in mathematical ability (Roberts et al., 2015; St Clair-Thompson et al., 2010).

Similarly, mathematical training does not lead to improved working memory (Messer, Thomas, Holliman, & Kucirkova, 2018). This evidence suggests working memory is relatively stable. Instead, research suggests working memory capacity gradually increases as part of a maturational process throughout childhood and levels off in adolescence (Gathercole, Pickering, Ambridge, & Wearing, 2004). Collectively, this evidence suggests that in order to support mathematical learning, low working memory capacity needs to be accommodated for within the learning activity to ensure children's access to the mathematical learning opportunity (Gathercole & Alloway, 2007).

Gathercole and Alloway (2007) suggest demands on working memory can be reduced with learning strategies, such as regularly repeating key task instructions and information, breaking multi-step tasks into smaller, more manageable tasks, and using manipulatives, such as number lines, which can help anchor children's attention to relevant information. These strategies may be particularly beneficial when children are being introduced to new mathematical content and concepts (Rycroft-Smith, Gilmore, & Cragg, 2017).

Socio-economic status and mathematics

Research shows children with a low socio-economic status (SES), which can be assessed by the levels of income, employment, and deprivation within a geographical area are significantly more vulnerable to mathematical underachievement compared to their

higher-SES peers (Anders et al., 2012; Kalaycioğlu, 2015). More specifically, at the start of school, children from lower SES backgrounds may show significantly weaker basic mathematical skills, relative to their higher-SES peers (Jordan et al., 2006) and this income-achievement gap accelerates over time (Johnson, McGue, & Iacono, 2007).

Although further research shows working memory (see above) is more predictive of children's early mathematical development than SES (Alloway et al., 2014), the impact of SES on children's learning of mathematics remains a significant educational issue. In particular, the Department for Education in the UK established the Education Endowment Foundation (EEF) in 2011. The EEF is dedicated to breaking the link between family income and educational achievement through generating, disseminating, and supporting the use of high-quality evidence-based educational resources (EEF, 2017).

Language proficiency and mathematics

Language also plays an important role in the development and learning of mathematics (Morgan, 2013). For example, studies show children's vocabulary skills are a significant predictor of later mathematical development (Purpura, Hume, Sims & Lonigan, 2011). This is particularly relevant for children attending bilingual, immersion education programmes, which are growing in popularity in middle-income countries, such as Brazil (Coelho Liberali & Megale, 2016). Children with poor language proficiency in the

language of instruction, for example when instruction is delivered in the child's second language (L2), may be vulnerable to underachievement in mathematics (Kempert et al., 2011; Snow & Uccelli, 2009). As such, mathematical instructional interventions designed to support these children should consider the role of language proficiency.

In the context of bilingual, immersion education, where the language of mathematics instruction may differ from the child's first language (Gomez, Freeman, & Freeman, 2005) it is important to consider the relationship between the child's proficiency in the language of instruction and their ability to engage with mathematics instruction.

A child's communicative proficiency can be modelled as being related to two dimensions (Cummins, 1981). One dimension distinguishes between communication that is context embedded, for example, inclusion of situational and paralinguistic supports, such as gestures and intonation contrasting with context reduced communication, for example, the absence of additional linguistic cues, which might occur reading a difficult text. Communication that is context embedded draws upon the child's basic interpersonal communication skills (BICS). In contrast, communication that is context reduced typically relies on the child's cognitive academic language proficiency (CALP) to access curriculum content (Cummins, 2008). Another dimension maps the degree to which a task presents cognitive demand for the learner, ranging from

cognitively undemanding, for example, a situation where information can be processed relatively easily to tasks with high cognitive demand, for example, involving the need to process multiple challenging pieces of information quickly.

The Threshold Hypothesis (Cummins, 1976) suggests that the child's language proficiency, particularly the child's CALP, needs to be sufficiently developed to enable the child to access curriculum content and to cope with instructional demands (Cummins, 2008). Thus, for children whose first language (L1) differs from the school language of instruction (e.g. the child's L2)[1], as is typically the case for children attending immersion education programmes (Baker & Homberger, 2001), there can be a disadvantage when learning mathematics (Abedi, 2003). This disadvantage can arise because low proficiency, particularly CALP, in the language of instruction, unless accounted for, can negatively impact children's learning opportunities and therefore curriculum achievements (Kempert et al., 2011; Snow & Uccelli, 2009). These ideas can potentially inform the design of instructional and learning environments (Cummins, 2008).

The current evidence base examining the impact of educational interventions on bilingual children's mathematical development is fragmented and limited. Studies evaluating language training interventions found limited transfer benefits to

[1] The categorisation of *first* and *second* language is used to support the initial investigation reported in this thesis (Chapter 5). It is not implied that languages are acquired fully sequentially or assumed that children do not also use other languages.

mathematical ability (Hagena, Leiss, & Schwippert, 2017; Jordan, Glutting, Dyson, Hassinger-Das, & Irwin, 2012). This evidence suggests mathematical skills need to be trained explicitly in the bilingual classroom.

Small-scale classroom intervention studies that specifically targeted mathematical skills found all children made significant progress in basic mathematical skills regardless of whether classroom instruction was delivered in the child's L1 or L2 (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007; Hemsley, Holm, & Dodd, 2014). Similarly, small-scale research evaluating the same technology-based maths app intervention that is being studied in this thesis found children with English as an Additional Language (EAL) made as much progress in mathematics as their non-EAL peers after using the maths apps in English for 6 weeks (Outhwaite, Gulliford, & Pitchford, 2017). Benefits of L1 classroom-based mathematics instruction were only seen in areas of higher-level mathematical conceptual development (Hemsley et al., 2014; see above). Collectively, these studies emphasise the need for targeted mathematics interventions to support mathematical development in bilingual learners.

Early intervention

In order to support mathematical development for children, particular those that may be at risk of underachievement in mathematics, potential solutions need to engage children from a young age (Butterworth et al., 2011; Jordan & Levine, 2009).

Research shows early learning experiences are a significant predictor of later achievement (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2010) after controlling for early cognitive skills, reading ability, and family and individual child characteristics (Watts et al., 2014). Furthermore, early intervention approaches may also be particularly beneficial in the first years of school (Clements, Baroody, & Samara, 2013) when children show the fastest rates of mathematical development (Hill, Bloom, Black, & Lipsey, 2008).

Theories of mathematical learning

There are several theoretical positions that aim to explain how young children learn mathematics (Bransford, Brown, & Cocking, 1999; Tauson & Stannard, 2018). However, critics argue learning theories are not part of a monolithic, coherent system (Lowyck, 2014) and thus suggest a pragmatic approach is needed that utilises a collection of relevant and reliable learning theory concepts (Spector, 2008). As such, the theoretical framework of this thesis will focus on two key areas of theoretical principles relevant to mathematical learning: (1) targeted practice and direct instruction, and (2) play-based pedagogies.

Targeted practice and direct instruction

Studies suggest efficiency in basic mathematical skills can be facilitated through targeted practice and direct instruction (Chodura, Kuhn, & Holling, 2015; Kroesbergen & Van Luit, 2003; Swanson & Hoskyn, 1999). Targeted practice emphasises deliberate task

repetition for effective skill acquisition (Daly, Martens, Barnett, Witt, & Olson, 2007; Imbo & Vandierendonck, 2008). Whereas, direct instruction is characterised by deliberately sequenced small units of information taught explicitly (Kirschner et al., 2006). Both concepts are grounded in behaviourist approaches to learning, which emphasises the importance of reinforcing an association between a particular stimulus and a response through feedback (Naismith, Lonsdale, Vavoula, & Sharples, 2004).

Intervention research shows that individualised training that places the child at the centre of the their learning with game-like activities that include targeted practice and direct instruction can enhance the development and fluency of targeted basic mathematical skills (Fuchs et al., 2009; Kidd et al., 2013) to a greater extent than more conceptually abstract approaches (Walker, Mickes, Bajic, Nailon, & Rickard, 2013). This evidence provides additional support for theories of cumulative learning (Gagné, 1968) and further emphasises the importance of developing a strong foundation in basic mathematical skills.

Despite this evidence, critics argue that while drill and practice based learning activities may promote rote learning of facts it is unlikely to foster deeper conceptual understanding (Ravitch, 2010). However, it is important to distinguish that deliberate targeted and direct instructional practice is not equivalent to rote repetition, which is characterised by shallow memorisation and a lack of meaningful context (Brabeck, Jeffery, & Fry, 2011; Hirsh-

Pasek, Zosh, Golinkoff, Gray, Robb, & Kaufman, 2015). Furthermore, counter critiques argue basic knowledge and skills that are acquired through extensive practice (Ericsson, Krampe, & Tesch-Römer, 1993) also provide the essential foundations for more complex and deeper learning (Shuell, 1990). This argument is supported by intervention studies that found the benefits to targeted basic mathematical skills generalised to other higher-order, more complex mathematical components not explicitly targeted by the intervention (Fuchs et al., 2009; Kidd et al., 2013).

Play-based pedagogies

In addition, research shows mathematical learning can also be supported through play-based pedagogies, including free play and guided play (Gray, 2015; Weisberg, Hirsh-Pasek, Golinkoff, Kittredge, & Klahr, 2016). Play-based pedagogies are grounded in sociocultural theories of learning, which suggest learning is an active process that occurs within a social environment (Hall, 2007; Vygotksy, 1978).

Free play is characterised by self-regulated, self-directed, and voluntary playful activities (Gray, 2015; Sutton-Smith, 2001). It is grounded in the idea that children acquire knowledge through interactions with their surrounding world (Piaget, 1977). Studies show, for example, that in free play-based mathematical activities young children were able to learn basic mathematical concepts, such as the relation between the parts and whole of a number (Björklund, 2008; Ginsburg, Pappas, & Seo, 2001).

However, critiques argue that when young children independently play with mathematically related objects it is difficult to ascertain whether children have successfully acquired the intended mathematical concept (Sarama & Clements, 2009). For example, child interactions with shaped blocks in a free play situation may not necessarily support the discovery of factual knowledge of shape properties (Fisher, Hirsh-Pasek, Newcombe, & Golinkoff, 2013).

In contrast, guided play follows a discovery learning approach and is positioned as an intermediary between teacher-led direct instruction and child-centred free play by combining adult guidance and child autonomy (Golbeck, 2001; Weisberg et al., 2016). In guided play, teachers (or other adults) create well-planned learning activities and actively engage with children in these activities to scaffold their discovery of mathematical facts and concepts (Ash & Wells, 2006; Fisher et al., 2013) while maintaining the locus of control with the child (Weisberg et al., 2016). In this case the concept of scaffolding is defined as support that enables children to meaningfully participate in and develop skills in a task that was not possible to complete unaided (Belland, 2014). Moreover, scaffolding enables children to engage with activities that are just beyond their current ability, known as the zone of proximal development (Vygotsky, 1978).

Studies show guided play learning activities can effectively facilitate the development of basic mathematical knowledge

(Ramani & Siegler, 2008; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2014) that can also generalise to different situations (Haden, Cohen, Uttal, & Marcus, 2016). Furthermore, research directly comparing guided play, free play, and teacher-led direct instruction in the context of mathematics found children taught shape properties through guided play showed greater immediate and sustained learning gains in shape knowledge compared to children exposed to free play and teacher-led direct instruction (Fisher et al., 2013).

However, guided play faces a number of practical challenges. For example, it requires one to one or small group based interaction, which is often time consuming and resource intensive (Plowman & Stephen, 2005). Guided play may also lack an appropriate level of practice required to master a particular mathematical concept (Stacy et al., 2017), particularly for children that may be vulnerable to underachievement in mathematics based on the cognitive, social, and language factors discussed above.

App Technology

In a context where efficient and effective mathematical interventions are needed to optimise young children's early mathematical learning and development (Pitchford et al., 2016; Stacy et al., 2017) educational maths apps delivered on touch-screen tablet devices offer an opportunity for individualised and scaffolded instruction targeted to children's needs. Based on a

blended learning approach (Naismith et al., 2004), educational app technology grounded in learning science theory can offer a hybrid of direct instruction, targeted practice, and play-based principles, thus bridging the pedagogical space between the starkly different learning theories (Weisberg et al., 2016).

Science of Learning principles

The Science of Learning is an emerging field drawing on interdisciplinary expertise from psychology, linguistics, brain imaging and other areas and seeks to address how children most effectively learn (Benassi, Overson, & Hakala, 2014; Sawyer, 2006). In the context of app-based learning, Hirsh-Pasek et al. (2015) proposed a four pillar, child-centred framework. This framework applies the well-known Science of Learning principles of active, engaged, meaningful, and socially interactive learning to educational app design.

Active learning.

Active learning is distinguished as either physically or mentally active (Hirsh-Pasek et al., 2015). In the context of educational apps, physical action refers to the tap, swipe, and drag movements required to operate the touch-screen tablet device (Zosh, Lytle, Golinkoff, & Hirsh-Pasek, 2017). Studies show young children have the basic motor skills to use tablet technology (Nacher, Jaen, Navarro, Catala, & González, 2015; Vatavu, Cramariuc, & Schipor, 2015) and research suggests physical actions may benefit learning (Chi, 2009), compared to passively observing

or listening to educational content (Hargrave & Sénéchal, 2000; James & Swain, 2011; Kersey & James, 2013).

However, other studies suggest learning outcomes are maximised when tasks involve minds-on, mentally active learning (Benware & Deci, 1984; Grabinger & Dunlap, 1995). For an educational app to effectively support learning, it needs to go beyond just physical movements and encourage cognitively active processing (Wouters, Paas, & van Merriënboer, 2008). For example, a child completing a physical action by mindlessly dragging stimuli on a touch-screen tablet requires minimal minds-on, active thinking and is unlikely to have educational benefits. In contrast, a child actively engaging with the app software by purposively figuring out which item is next in the pattern sequence and dragging the stimuli to the appropriate place teaches pattern completion mathematical concepts (Hirsh-Pasek et al., 2015; Zosh et al., 2017).

Engaged learning.

Engaged learning refers to the extent to which the activity keeps the child on task and minimises distractions (Tauson & Stannard, 2018). At the most basic levels of engagement are contingent interactions; each tap, swipe, or drag movement with the app software is followed by an immediate response. Studies suggest this enables the child to feel in control and maintains their focus (Calvert, Strong, Jacobs, & Conger, 2007; Nielsen, 2014). Furthermore, this multi-touch nature of the tablet device affords cognitive embodiment, which suggests learning is grounded in the

physical body's interactions with the world (Shapiro, 2010) and has been shown to be influential in mathematical development and learning with tablet technology (Calder, 2016; Duijzer, Shayan, Bakker, Van der Schaaf, & Abrahamson, 2017; Wang, Xie, Wang, Hao, & An, 2016).

Based on theories of extrinsic motivation, a child's engagement and consequent learning is heightened through receiving meaningful feedback (positive or negative) and external rewards (Ryan & Deci, 2000). However, a child's long-term engagement is facilitated by intrinsic motivation. When a child is immersed in an activity, they experience a state of 'flow' where their levels of attention and concentration are increased to achieve a goal (Csikszentmihalyi & Csikszentmihalyi, 1992). For an app to be intrinsically motivating it needs to offer the child the opportunity to experience challenges and autonomy (Malone, 1981; Przybylski, Rigby, & Ryan, 2010). App software that incorporates these features is more likely to have the potential to positively impact intrinsic motivation, by going beyond extrinsically motivating rewards and feedback and make playing the app itself rewarding (Garris, Ahlers, & Driskell, 2002) and improve learning outcomes (Muis, Ranellucci, Trevors, & Duffy, 2015).

Meaningful learning.

Meaningful learning refers to the extent to which the learning activity is connected to the child's level of existing knowledge (Hirsh-Pasek et al., 2015). As such, meaningful learning, goes

beyond the shallow rote learning of facts, which does not link to previously learnt content (Ravitch, 2010).

To support meaningful learning, app software needs to incorporate a stepped curriculum that builds on children's prior knowledge (Magliaro, Lockee, & Burton, 2005) and challenges children with tasks that are just beyond their current ability level (Inal & Cagiltay, 2007; Vygotsky, 1978). App software that embeds a contextual narrative (Rogoff, 1995) with useful semantic context (Zimmerman, Christakis, & Meltzoff, 2007), and includes regular knowledge retrieval based assessments (Grimaldi & Karpicke, 2014) may also encourage meaningful learning.

Socially interactive learning.

Sociocultural learning theories consider learning to be an inherently social activity and propose high-quality social interaction is critical for effective learning (Sawyer, 2006; Vygotsky, 1978). While touch-screen tablet devices are not inherently socially interactive or adaptive compared to human interactions, educational apps can incorporate socially interactive learning in a number of ways.

Firstly, app software can integrate socially interactive learning through parasocial relationships with on-screen characters offering extrinsic rewards and social contingency through demonstrations (Troseth, Saylor, & Archer, 2006). Research has identified in-app social demonstrations can enhance young children's learning compared to when a stimulus is moved without an obvious

demonstrator (Fiorella & Mayer, 2016; Zimmerman, Moser, Lee, Gerhardstein, & Barr, 2017).

Secondly, the learning environment in which the app-based instruction is implemented can also engender socially interactive learning. Studies show educational apps can encourage a child's collaboration with peers (Hwang, Shadiev, Tseng, & Huang, 2015; Kucirkova, Messer, Sheehy, & Panadero, 2014) and interactions with adults (Kucirkova, Sheehy, & Messer, 2015).

App technology affordances

It is suggested that educational apps that incorporate these four Science of Learning principles with clearly specified learning goals can effectively support learning outcomes (Hirsh-Pasek et al., 2015). Furthermore, based on these principles, educational apps can combine the benefits of direct instructional approaches (Kirschner et al., 2006), in particular, a stepped curriculum, rehearsal, contingent feedback, and rewards, with features of play-based pedagogies (Gray, 2015), notably, choice, self-regulation, and control. This can help provide an efficient child-centred and scaffolded learning environment (Grant et al., 2015; Mayo, 2009; Mayer, 2004) tailored to individual needs (Slavin & Lake, 2008) enabling individualised and structured instruction (Adams & Carnine, 2003) without additional, time-consuming, teaching demands (Kucian et al., 2011; Hilton, 2016).

In addition to the theoretical affordances of educational app software, touch-screen tablet devices also offer a number of unique,

practical features suited for young children. In particular, young children typically find them motivating (Flewitt, Messer & Kucirkova, 2015) and intuitive to use (Cooper, 2005). Touch-screen tablet devices are also mobile, light weight, and do not rely on dexterity-based motor skills that are required to use a computer keyboard or mouse (Kucirkova, 2014). Studies show young children have the basic motor skills to successfully operate tablet devices (Nacher et al., 2015; Vatavu et al., 2015).

Furthermore, access to touch-screen tablet devices in educational setting is increasing (Tamim et al., 2015). For example, in the UK, 70% of primary schools have access to touch-screen tablets (Clarke, 2014). Initiatives are also underway to increase access to tablet technology to low-income countries, such as Malawi (Hubber, Outhwaite, Chigeda, McGrath & Hodgen, 2016) and middle-income countries, such as Brazil (Drinkwater, 2013; Johnson, Adams Becker, Cummins, Estrada & Meira, 2012). However, research suggests this increased investment from schools and governments has been driven by the popularity of tablet technology instead of by educational frameworks or research-based evidence (Tamim et al., 2015).

Previous research

Despite the prevalence, popularity, and potential benefits of using educational apps to support early mathematical development and learning, the current evidence base is fragmented (Haβler, Major & Hennessey, 2015) and lacks rigorous scientific investigation

(Cheung & Slavin, 2013). To evaluate the impact of educational apps practical and high-quality real-world research is needed (Cheung & Slavin, 2013) and should focus on the quality of educational app content (Blum-Ross & Livingstone, 2016; Falloon, 2013).

Recent reviews of emerging experimental research shows educational apps can have positive effects on early mathematical development (Herodotou, 2017), particularly for children experiencing mathematical difficulties (Cozard & Riccomini, 2016). Furthermore, increasing time spent learning mathematics using educational apps at home positively benefited children's achievement in school (Berkowitz et al., 2015). A more detailed systematic review of the emerging literature, including methodological limitations and directions for future research is reported in Chapter 2 (Systematic Review). In building the research evidence base on the effectiveness of educational maths apps for supporting early mathematical development a staged approach is needed (McDonald, Kessler, Kauffman, & Schneider, 2006).

Section 2: Evidence-based education

Developing and scaling interventions

In order to understand the impact of educational maths apps for supporting early mathematical development and the process of scaling such interventions, formal evaluation research is required.

There are five key stages proposed in the development and scaling of educational interventions (EEF, 2018a).

Initial evidence

Initial evidence involves establishing the proof of concept of a particular educational intervention. These studies aim to identify the potential for an intervention to improve learning outcomes, such as trialling the intervention in a particular school environment.

In this thesis, initial evidence for the benefits of the maths app intervention in different educational contexts will be demonstrated in the Proof of Concept UK Study (Chapter 4) and the Proof of Concept Brazil Study (Chapter 5).

Pilot studies

After establishing the initial evidence, the next stage is a feasibility trial. At this piloting stage, the intervention is still in an early, exploratory stage of development. Research is typically conducted on a small-scale, for example piloting the intervention in more than three schools (EEF, 2018a). Rigorous quantitative methods, such as RCTs are advocated in order to establish cause and effect between the educational intervention and learning outcomes (Torgerson & Torgerson, 2001; Connolly, Biggart, Miller, O'Hare, & Thurston, 2017). However, qualitative methods are also valued in developing and refining the intervention approach, ahead of the next stage in scaling (Humphrey, Lendrum, Ashworth, Frearson, Buck, & Kerr, 2016). A detailed critique on the use of RCTs in educational research is discussed later in this section.

In this thesis, the scalability of the maths app intervention progresses to the pilot study stage. Chapter 6 reports a pilot, feasibility trial of the maths apps using a pupil-level RCT in 12 UK schools. Chapter 7 then reports a qualitative evaluation of the maths app intervention implementation and examines the relationship between implementation and intervention outcomes.

Efficacy trial

Next, an efficacy trial aims to examine whether an educational intervention can work under optimal, developer-led conditions at a larger scale than the pilot studies (Revicki & Frank, 1999). The EEF typically conducts efficacy trials in 50 or more schools (EEF, 2018a).

Efficacy trials are usually conducted in an ideal, resource intensive setting with a highly selected, homogenous population group. Teachers implementing the intervention are typically highly experienced and trained on the intervention and the intervention protocol is standardised and strictly enforced (Singal, Higgins, & Waljee, 2014). As such, a greater emphasis is placed on internal validity, the extent to which cause and effect can be established, than generalisability, the extent to which results can apply to different settings, in efficacy trials (Revicki & Frank, 1999).

Effectiveness trial

In contrast, an effectiveness trial aims to examine whether an educational intervention can work in real-world classroom practice (Revicki & Frank, 1999). Specifically, effectiveness trials are usually

conducted in real-world, everyday settings with a heterogeneous population group. The intervention is also applied with flexibility and implementing teachers are representative of usual providers (Singal et al., 2014). Therefore, effectiveness trials focus more on the generalisability of an intervention compared to efficacy trials (Revicki & Frank, 1999).

In particular, an effectiveness trial aims to understand and refine the scalable model of the intervention beyond the research context (EEF, 2018a). In order to capture this generalisability, the EEF typically conducts effectiveness trials in more than 100 schools across three different geographical regions (EEF, 2018a). Furthermore, it is important to recognise that while there are differences in focus in efficacy and effectiveness trials, the relationship between these methods can be considered a continuum, rather than a strict dichotomy (Fritz & Cleland, 2003).

Scale-up

After an educational intervention has been shown to work under rigorous conditions and demonstrates the capacity to deliver at scale, the intervention is then expanded to deliver in a larger number of schools (EEF, 2018a).

In summary, each stage of the scaling process provides the required evidence base for the subsequent step in developing and scaling an educational intervention. Each stage utilises appropriate methodologies to address different research questions to gradually

add to the evidence base and contribute towards the full scale-up of an educational intervention.

As previously mentioned, RCTs are considered by many as the 'gold standard' method for determining whether a particular educational intervention causes measurable effect on children's learning and development (Connolly et al., 2017; Haynes, Service, Goldacre, & Torgerson, 2012; Torgerson & Torgerson, 2001). However, critiques argue the RCT design is too simple and ignores the complexities of school based contexts (Biesta, 2010; Morrison, 2001; Thomas, 2016). A critical evaluation on the role of RCTs in evidence-based education will now be discussed.

RCTs in evidence-based education

The start of the twenty-first century saw the emergence, growth, and increased investment in evidence-based education (Thomas & Pring, 2004; Hammersley, 2007). Evidence-based education seeks to understand 'what works' and shows a predilection for using experimental, quantitative, post-positivist methodologies, particularly RCTs (Connolly et al., 2017; Haynes et al., 2012; Torgerson & Torgerson, 2001).

In the context of educational research, the RCT design aims to measure children's progress in response to an intervention compared to a control group of children typically receiving standard teaching practice. Random allocation to group helps to ensure factors that may influence learning gains, such as age, gender, and pre-test ability are evenly distributed across the groups.

Sometimes group differences may still arise, such as through unsuccessful randomisation. However, baseline differences in the outcome variable can be taken into account by the statistical test used to analyse the data. For example, if there are differences in the outcome variable at pre-test, a difference score (post-test minus pre-test) based analysis can be used (Thomas & Zumbo, 2012) and is considered less biased and more appropriate than an analysis of co-variance (ANCOVA). This is because ANCOVA, which examines differences in post-test scores across the groups while controlling for pre-test scores, assumes the absence of baseline differences in the outcome variable (Van Breukelen, 2006).

Based on the logic of the RCT design, if on average the learning gains made by the children in the intervention group are shown to exceed those in the control group, it can indicate the educational intervention is successful. This is because the only assumed systematic difference between the two groups of children is that one received the intervention and the other did not. Consequently, a causal relationship between the intervention and learning gains can be established.

However, this predominant post-positivist approach has been subject to substantial trenchant criticism (Connolly, Keenan & Urbanska, 2018). It is argued that the purpose of RCTs to establish universal and replicable laws (Hodkinson & Smith, 2004) is too simplistic for complex interventions within complex environments, such as schools (Biesta, 2010). It is argued the emphasis on

statistical aggregation in RCTs removes educational interventions and their outcomes from the context in which they are situated (Elliott, 2004). Furthermore, it is argued that understanding the causal relationship between intervention and learning outcomes becomes more nuanced when more complex social and personal processes enter the situation, such as the involvement of more schools in a feasibility trial (Thomas, 2016).

In response to these criticisms a recent study systematically reviewed over 1000 RCTs that have been conducted across a range of educational settings, interventions, and outcomes (Connonlly et al., 2018). The review found 78% of reported RCTs included some discussion and critical reflexivity on the generalisability of the RCT findings to other situations, contexts, and participant groups. This counters the argument that RCTs in education rely too heavily on establishing universal causal laws.

While RCTs are acknowledged to form part of an epistemological ecosystem of educational inquiry, it is argued that positioning RCTs as superior to other research methodologies (e.g. Haynes et al., 2012) is narrow in focus and limits our understanding of successful educational interventions (Thomas, 2016). For example, small-scale, quasi-experimental designs that afford a basic experimental stance in real-world settings (Robson & McCartan, 2011) but do not include random allocation to groups (Campbell & Stanley, 1963) may be suitable for establishing initial evidence in the early phases of scaling an intervention (Carnine & Gersten,

2000), particularly when randomised control designs are not practically feasible (Kim & Steiner, 2016).

The need to pragmatically utilise other methods is also supported by further findings from Connelly et al.'s (2018) systematic review. Result showed only one in five RCTs evaluated intervention implementation. Research designs that include an emphasis on the social context are vital for understanding how the intervention works and under what circumstances (Connolly et al., 2018; Pawson & Tilley, 1997; Slavin, 2012). As such mixed-methods designs that optimally integrate the most appropriate quantitative and qualitative methods to understand the intervention impact and implementation are advocated (Thomas, 2016; Shaffer, 2011).

Understanding implementation

Implementation and process evaluation stems from health care research and aims to explain how and why the implementation of an intervention succeeds or fails. Understanding the implementation of an interventions helps to support the transition from controlled conditions, such as those characteristic of feasibility and efficacy trials to real life conditions seen in effectiveness trials (Glasgow, Lichtenstein, & Marcus, 2003).

Defining implementation

Implementation is a multi-dimensional construct (Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Durlak & DuPre, 2008)

and studies have identified eight distinct but related dimensions (Durlak & DuPre, 2008; Forman, Olin, Hoagwood, Crowe, & Saka, 2009; see Table 1). Previous implementation evaluation research has primarily focused on a single dimension of implementation, predominately fidelity or dosage (Berkel et al., 2011; Durlak & DuPre, 2008). This approach has been criticised for over-simplifying implementation, which is considered complex and multi-faceted (Biesta, 2010; Humphrey et al., 2016). As such, further research has called for an increased focus on other, multiple dimensions (Humphrey et al., 2016; Lendrum, Humphrey, & Greenberg, 2016). Consequently, the research reported in Chapter 7 (UK RCT Implementation Evaluation) will focus on the three dimensions highlighted in bold in Table 1. These three dimensions were chosen to give a broad overview of the intervention implementation within the practical feasibility of the study and the chosen methods; narrative, direct observations of children using the maths apps and self-report interviews with teachers.

Table 1 Eight distinct but related dimensions of implementation (After: Durlak & DuPre, 2008; Forman et al., 2009)

Dimension of Implementation	Definition
Quality	**E.g. How well was the intervention implemented as intended?**
Adaptations	**E.g. In what ways was the implementation of the intervention changed in a real-life context?**
Participant responsiveness	**E.g. To what extent are the recipients of the intervention engaged with the intervention activities?**
Fidelity	E.g. Is the intervention delivered as intended?
Dosage	E.g. How much of the intervention was implemented?
Reach	E.g. To what extent are the study participants representative of the intended target population?
Programme differentiation	E.g. To what extent is the intervention activities different from existing programmes?
Monitoring of control conditions	E.g. To what extent are the core components of invention activities present in control conditions?

Quality

Implementation quality refers to how well different aspects of the intervention are delivered (Durlak & DuPre, 2008). For example, research may focus on the process of implementing an intervention, such as teachers' responsiveness in the delivery of an educational intervention (O'Donnell, 2008). It is distinct from implementation fidelity, which focuses on the adherence to the structure and sequence of intervention activities (O'Donnell, 2008; Odom et al., 2010).

Adaptations

Adaptations refer to the ways in which the intervention may be changed during implementation by the implementers (Durlak & DuPre, 2008; Naylor et al., 2015). Adaptations are a natural process since implementers, such as teachers are 'active modifiers' not 'passive acceptors' of a particular intervention (Rogers, 2003). Adaptations can range from minor, surface level changes, such as, changing a cultural reference to appeal to the specific audience, to more substantial, deep changes, such as, removing core components of the intervention (Moore, Bumbarger, & Cooper, 2013). Other adaptations include the logistical fit, timing, and valence of an intervention (Moore et al., 2013). It is vital that implementation evaluation research examines adaptations to interventions in order to understand the relationship between any changes at the local, implementation level, and the intervention outcomes (Elliott & Mihalic, 2004).

Participant responsiveness

Participant responsiveness refers to the level of engagement with the intervention activities and content from the intended recipients (Dusenbury, Brannigan, Falco, & Hansen, 2003). For example, research may focus on how well the intervention appears to engage and maintain the attention of the participants (Durlak & DuPre, 2008).

Implementation variability and intervention outcomes

There is limited research linking implementation variability to intervention outcomes (Humphrey et al., 2016). Nevertheless, the few studies that have been conducted show a significant association between differences in intervention implementation and intervention outcomes (Askell-Williams, Dix, Lawson & Slee, 2013; Hansen et al., 2013).

For example, Askell- Williams et al. (2013) created an implementation index to quantitatively assess implementation quality, fidelity, and dosage of a mental health initiative implemented in primary schools. Results showed a significant relationship between high and medium rated intervention implementation and greater intervention outcomes over time, compared to low rated implementation.

Similarly, Hansen et al. (2013) assessed the frequency and nature of teacher adaptations in the delivery of a school based drug prevention programme through coded video observations. Results

showed teachers who made positive but fewer adaptations had a higher percentage of students that remained non-drug users compared to teachers who made more frequent adaptations, regardless of the valence of the adaptations.

An important caveat to consider with the current research, however, is that implementation variability is a natural process (Rogers, 2003) and not experimentally manipulated. As such, it is difficult to make causal inferences (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012) between implementation variability and intervention outcomes. Instead, the link should be described as a relationship, to reflect the correlative nature of the findings (Humphrey et al., 2016).

Furthermore, Humphrey et al. (2016) highlight that there is a serious and significant gap in the reporting of qualitative data in implementation evaluation research. For example, the description of and justification for the methodological approach used, and attempts to establish the quality, validity, and credibility of the qualitative data are frequently missing.

Implementation theories

Furthermore, previous implementation evaluation research has been criticised for a lack of focus on the theoretical underpinnings of implementation (Eccles, Grimshaw, Walker, Johnston, & Pitts, 2005). In response, Nilsen (2015) proposed a taxonomy that differentiates categories of implementation theories

depending on the stage of the intervention development and evaluation cycle. For example, process models focus on guiding the process of translating research into practice. Evaluation frameworks focus on aspects of implementation that could be evaluated to determine the success of an intervention. In contrast, determinant frameworks aim to identify factors that may act as barriers or enablers (independent variables) that influence intervention outcomes (dependent variables). As such, a determinant framework is the most relevant to the current body of work reported in this thesis.

Determinant frameworks

Determinant frameworks are typically developed by synthesising results from empirical studies that identify the barriers and enablers for implementation success (Nilsen, 2015). Furthermore, determinant frameworks are typically multi-level; the intervention is placed within an ecological system surrounded by different influencing factors at different levels of that system (Bronfenbrenner, 1979; Thomas, 1992). For example, Domitrovich et al.'s (2008) Implementation Quality model places the Implementation Quality of an intervention within an ecological system with multiple levels of influence, incorporating the individual-level (for example, professional characteristics), the school-level (for example, school culture) and the macro-level (for example, leadership and human capital).

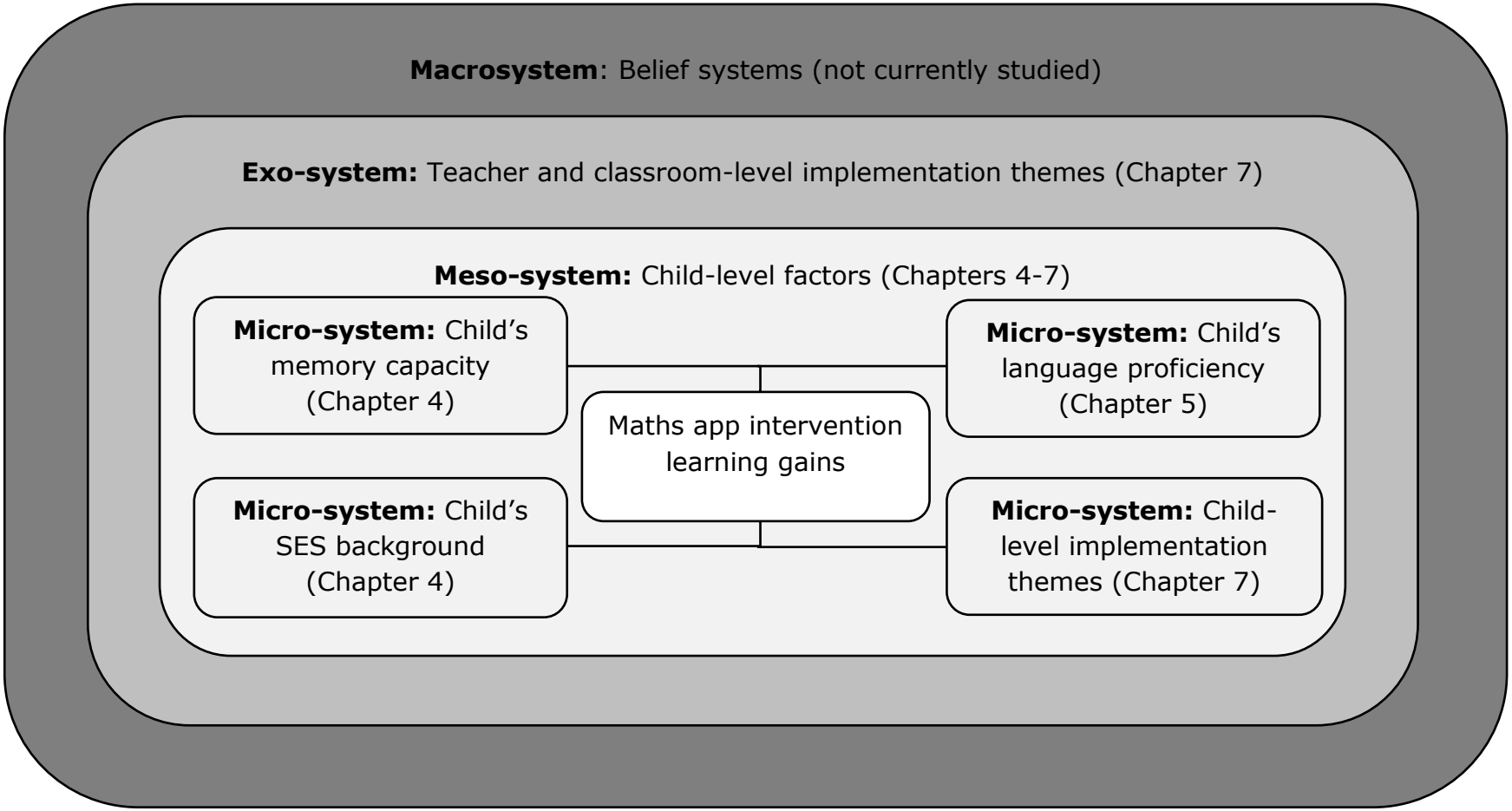
While there is a plethora of different implementation science theories, which aim to understand and explain how and why implementation succeeds or fails in a given context (Humphrey et al., 2016) many of the existing theories are not relevant to educational interventions and are beyond the scope of the current research in this thesis. For example, current models are largely based on health care interventions (e.g. Domitrovich et al., 2008) and focus on the diffusion of interventions beyond the research context (e.g. Greenhalgh et al., 2008). Limited progress has been made in the field specific to educational technology-based interventions. However, it is vital that implementation theories are developed in this context to holistically evaluate the maths app intervention.

Bronfenbrenner's (1979) Ecological Systems Theory provides a valuable structure in which to situate children's learning gains with the maths app intervention within micro-, meso-, exo-, and macrosystems (see Figure 1). The microsystem focuses on influencing factors at the individual level. Based on the literature reviewed so far, these child-level factors may include memory capacity, SES background, language proficiency, and child-level implementation themes. The mesosystem refers to the collection of microsystems. The exosystem consists of events, for example teacher and classroom level implementation themes, which may occur within and may influence the maths app intervention setting. Finally, the macrosystem consists of belief and values, for example of the teacher and senior management staff that may impact the

lower systems. For the purposes of the current research, this thesis focused on the micro-, meso-, exo-systems to develop the initial evidence evaluating the maths apps in a real-world context.

This multi-level, ecological, determinant framework (see Figure 1) offers the possibility of a holistic evaluation of the maths app intervention and has the potential to synthesise educational neuroscience (for example, working memory), child characteristics (for example, SES and language proficiency; see Section 1), and implementation science (Section 2), enabling the identification of multi-level factors that may influence children's mathematical learning gains with the apps.

Figure 1 Multi-level, ecological, determinant framework based on Bronfenbrenner's Ecological Systems Theory (1979).



Conclusion

In summary, Chapter 1 highlights mathematical skills are critical for later educational success and economic, health, and employment outcomes. The componential approach to mathematical development emphasises the importance of strong basic mathematical factual and conceptual knowledge for providing an essential foundation for higher-order mathematical reasoning and problem solving skills. Children with a weak memory capacity, from a low socio-economic background, and with a low proficiency in the language of instruction are all vulnerable to mathematical underachievement. Early interventions may be beneficial for securing young children's early mathematical learning and development.

Key theories of early mathematical learning include targeted practice, direct instruction, and play. Targeted practice and direct instruction emphasise deliberate task repetition with explicit instruction, feedback, and reward. These instructional approaches can support the acquisition of basic mathematical skills and transfer to higher-order mathematical components. Mathematical learning can also be facilitated through self-regulated and self-directed play-based experiences.

App technology offers a unique opportunity to combine the benefits of these pedagogical approaches to provide an innovative intervention for early mathematical learning and development.

Specifically, maths apps delivered on touch-screen tablet devices that are grounded in the principles of active, engaged, meaningful, and socially interactive with a specific learning goal have the potential to offer a child-centred and scaffolded learning environment.

A staged research process evaluating the use of app-based mathematics instruction is needed to inform educational policy and practice, as well as theoretical understandings of how children most effective learn this technology. Mixed-methods approaches, including the use of randomised control trials are valued in applied educational research. However, an increased focus on implementation evaluation and a multi-level, ecological, determinant framework, which considers factors at the child, classroom, and teacher level, are required to understand how app technology works, who it works for, and under what circumstances it can work.

Overall, this thesis aims to examine the reach and utility of a new maths app intervention (see Chapter 3) in supporting the acquisition of early mathematical skills for young children at the start of formal education in two countries that differ in educational context and income status, the UK and Brazil.

Chapter 2: Systematic Review

Background

As highlighted in Chapter 1, previous research examining the use of educational apps to support early mathematics is emerging but the field is currently fragmented and limited. Five notable reviews have been conducted relevant to this research area, which synthesise the current body of evidence and highlight important methodological issues (Cheung & Slavin, 2013; Cozard & Riccomini, 2016; Crompton, Burke, & Gregory, 2017; Haβler et al., 2015; Herodotou, 2017).

Haβler et al. (2015) reviewed 23 studies examining the impact of touch-screen tablet devices across a broad range of subject domains. Results showed previous studies mostly reported positive learning outcomes for children aged 5 to 18 years. However, the current evidence base was identified as fragmented with a paucity of rigorous methodologies. This conclusion is consistent with previous research that showed technology-based mathematical intervention studies were frequently brief in duration, lacked an adequate control group, did not establish initial equivalence between groups, and used measures inherent to the experimental condition (Cheung & Slavin, 2013). These issues can increase the risk of inflated effect sizes compared to more rigorous methods and thus reduce the validity of the study findings (Slavin & Madden, 2011).

In a similar vein, a recent systematic review that synthesised the focus and methodologies used in studies examining the impact of tablet technology in education for children aged 5-18 years showed of the 113 studies included, only 2% followed an experimental research design, such as a RCT. 11% of studies used a quasi-experimental design. Results also indicated mathematics was relatively understudied in comparison to literacy and science. Furthermore, the tablet technology-based learning activities predominantly focused on the use of app-based instructional content rather than children using touch-screen tablets to become producers, collaborators, and creators of knowledge (Crompton et al., 2017).

Building on this evidence, a recent review focusing on the use of app-based instructional content with children aged 2-5 years showed more rigorous evidence is emerging. Of the 19 studies identified, 14 utilised RCT, quasi experimental or mixed-methods designs. Five of these focused on mathematics, thus also demonstrating an increased focus on app-based mathematical instruction. Consistent with the previous reviews, results showed app technology had positive benefits on children's mathematical learning and development (Herodotou, 2017). Similarly, a review of eight studies showed digital-based mathematical interventions, delivered on touch-screen tablets or computers supported the development of fluency in basic mathematical skills for children aged 7-14 years with mathematical difficulties (Cozard & Riccomini, 2016).

Collectively these five reviews are indicative of the current research and potential benefits of app technology. However, the current reviews are broad in focus. In particular, the reviews examined a wide range of mathematical interventions (Cheung & Slavin, 2012) and a large age range of participants (Crompton et al., 2017; Haβler et al., 2015). Reviews that were more specific to app-based mathematics instruction were limited to children under 5 years of age in formal, school and informal, home settings (Herodotou, 2017) and children with mathematical difficulties (Cozard & Riccomini, 2016). It is important to consider the impact of app-based mathematics instruction for all young children at the start of formal education, as studies suggest typically attaining and low-achieving pupils can both demonstrate difficulties in acquiring different components of mathematical skills (Dowker, 2005; Geary & Hoard, 2005).

Emerging experimental research that adopts this focus demonstrates the benefits of different high-quality maths apps with young children in a classroom setting (e.g. Outhwaite et al., 2017; Pitchford, 2015; Schacter & Jo, 2016; 2017; van der Ven, Segers, Takashima, & Verhoeven, 2017). In order to assess this emerging evidence base and identify directions for future research, a systematic review with a narrative synthesis of the literature was conducted.

Methods

Objective

The aim of this systematic review is to evaluate the impact of app-based educational interventions for supporting mathematical development and learning in young children during the first three years of formal education.

Inclusion and exclusion criteria

The PICOS method (Population, Intervention, Comparison, Outcome, Study Design) adapted from the original PICO method (Richardson, Wilson, Nishikawa, & Hayward, 1995) was used to develop strict and clear inclusion and exclusion criteria summarised in Table 2. This criteria was pragmatically developed to intentionally focus on studies directly relevant to the overarching aim of this thesis (see page 4). This was intended to afford important and specific lessons to be established from the previous research.

Population.

For the purpose of this thesis and to expand on previous reviews (Herodotou, 2017), the population of interest was young children attending the first four years of formal education relevant to the country in which the study took place, for example Kindergarten to Third Grade in the USA. A strict age criteria was not applied as some countries, for example Malawi follow an ability-based educational system.

In order to capture a broad range of participants, studies that focused specifically on children from special populations, such as

children underachieving in mathematics, children with special educational needs, or children from low-income settings were also eligible for inclusion.

Intervention.

Eligible intervention studies must evaluate the impact of an educational maths app delivered on touch-screen tablet devices in a school based setting. It is preferred that teachers implement the intervention to ensure high ecological validity and to support the generalisability of the intervention beyond the research context (Clements, Sarama, Wolfe, & Spitler, 2015). However, this is not always practically possible and so in order to maximise the number of studies eligible for the systematic review, interventions may have been implemented by the class teacher or the research team.

Similarly, while it is recommended that interventions are implemented for a minimum of 12 weeks (Higgins, Xiao & Katsipataki, 2012), a study duration of a minimum of one week was applied to ensure sufficient studies could be included in the systematic review. However, in order to address previous criticisms surrounding the brief duration of interventions (Cheung & Slavin, 2013) and examine its association with learning gains, included studies must explicitly state sufficient details to calculate time on task. Specifically, how long the intervention was implemented for, including the number of sessions, the time duration of the sessions, and the full length of the intervention period.

Comparison.

To address other previous criticisms surrounding the lack of an adequate control group in technology-based intervention studies (Cheung & Slavin, 2013) all eligible studies were required to include some form of a control group. However, based on the relative infancy of the research field, exact comparators were not specified. This enabled all potential comparison conditions to be identified. This was designed to facilitate a discussion on the nature of research questions asked in specific studies and thus identify directions for future research.

Outcome.

For the purposes of this systematic review, all eligible studies must quantitatively measure some aspect of mathematical ability as the primary outcome variable. Quantitative assessments of mathematical ability were required to take place before (pre-test) and immediately after (post-test) the intervention period. Additional delayed post-tests were also eligible, but were considered a secondary outcome.

Although previous reviews have criticised the use of researcher developed measures to assess the learning gains associated to a particular intervention (Cheung & Slavin, 2013) this restriction was not included in the inclusion criteria. This was due to the global perspective of this systematic review and so standardised assessments may be biased towards a Western population and not

appropriate for use in low-income countries (Laher & Cockcroft, 2017; Sabanathan, Wills, & Gladstone, 2015).

Study design.

Based on the emergence and value of experimental methods in this research area (Connolly et al., 2017; Herodotou, 2017; Haynes et al., 2012; Torgerson & Torgerson, 2001), included studies must follow a quasi-experimental or RCT design at the pupil or class level. Random allocation to condition is not necessary but as previously stated all studies must include an adequate control group. Mixed-methods, such as studies that include an implementation process evaluation (Connolly et al., 2018) are eligible for inclusion.

Table 2 Inclusion and exclusion criteria used to select studies for the systematic review.

Item	Inclusion criteria	Exclusion criteria
Participants	Young children must be in the first four years of formal education relevant to the country in which the study took place. Studies that focused specifically on children underachieving in mathematics or children with special educational needs are also eligible for inclusion.	Older children in the fifth year or above of formal education relevant to the country in which the study took place (e.g. USA; Fourth Grade or above).
Intervention	Studies must primarily examine the impact of educational app(s) delivered on touch-screen tablet devices on some aspect of young children's mathematical development.	Studies that examine the general affordances of touch-screen tablets and not specifically educational app(s). Computer assisted interventions and subject domains other than mathematics are not eligible for inclusion.

Item	Inclusion criteria	Exclusion criteria
Study setting	Studies must be conducted in a school based setting. Given the global nature of this thesis, the geographical location of studies was not restricted.	Studies that are not conducted in school based setting, for example, studies conducted in the home learning environment (e.g. Berkowitz et al., 2015).
Study duration	Studies must be conducted for a minimum of one week. Studies must also explicitly specify how long the intervention was implemented for, including the number of sessions, the time duration of the sessions, and the full length of the intervention period.	Articles that do not explicitly state the study duration with the relevant level of detail. Studies implemented for less than one week.
Comparison	Studies must include a control/comparison group. No restriction was placed on the activities of the control group, for example, use of other maths apps or standard mathematical practice.	Studies that do not include a control group.

Item	Inclusion criteria	Exclusion criteria
Outcome measures	Studies must primarily measure some aspect of mathematical ability before (pre-test) and immediately after (post-test) the intervention period. No restriction is placed on type/ design of assessment. Studies that include an additional delayed post-test are also eligible for inclusion.	Studies that do not include a measure of mathematical ability before and after the intervention period. Studies that not use mathematical ability as the primary outcome measure are also not eligible for inclusion. For example, studies that primarily assess teacher perspectives on technology are not eligible.
Study design	Quasi-experimental or RCT designs with allocation to group at the pupil or class (cluster) level. Random group allocation is not necessary but studies must include a control/comparison group. Studies may be mixed-methods, but there should be a strong quantitative focus.	Studies which are predominately qualitative in focus. Studies that do not include a control group.

Item	Inclusion criteria	Exclusion criteria
Year published	Articles must be published between January 2010 and July 2018.	Studies conducted prior to 2010.
Language	Articles must be written in English.	Articles that are not written in English.
Publication type	Articles must be peer reviewed.	Articles that are not peer reviewed, for example, book chapters and PhD theses.

Search strategy

A literature search using the search items outlined in Table 3 (search strategy code reported in Appendix A) was performed in six databases including Web of Science, PsycINFO, PsycARTICLES, Medline, PubMed, and the Educational Resource Information Centre (ERIC). The literature search aimed to identify articles published in peer-reviewed journals in English between January 2010 and July 2018. This date range was chosen as touch-screen tablet devices (e.g. the Apple iPad) were launched in 2010.

This search was supplemented with a backwards citation chaining search of additional articles that were included in previous systematic reviews (Cozard & Riccomini, 2016; Herodotou, 2018) that were not identified in the original search of the six databases. Although citation chaining is time consuming it is considered effective best practice when conducting systematic reviews (Dundar & Fleeman, 2017).

Table 3 Search terms included search strategy. Full search strategy code reported in Appendix A.

Population	Intervention	Outcome
"Young children"	Tablet	Mathematics
Kindergarten	"Tablet technology"	Maths
Preschool	"Educational apps"	Math
Elementary	"Interactive apps"	Achievement
Primary		Development
		Learning

Study selection

The PRISMA 2009 flow diagram in Figure 2 summarises the flow of research studies throughout the different phases of the systematic review (Moher, Liberati, Tetzlaff & Altman, 2009).

In total, the described literature search identified 203 articles. 42 duplicate articles were removed and the remaining 161 articles were screened in line with the inclusion and exclusion criteria described above and outlined in Table 2. Article screening was primarily based on the title and abstracts of the 161 articles. However, if these details were unclear, the full article was consulted.

Of the 161 articles initially screened, 130 studies did not satisfy the inclusion criteria and were consequently excluded. Full

articles for the remaining 31 studies were further assessed in line with the inclusion and exclusion criteria (see Table 2). Of the 31 articles further assessed, 24 did not meet the inclusion criteria and were excluded. In total, seven studies met the inclusion requirements for the systematic review.

Study quality assessment

Prior to data extraction, the quality of the seven included studies was assessed using a quality assessment tool designed specifically for the systematic review. The assessment tool was designed to assess five areas of potential bias relevant to school based research (see Table 4; Greenhalgh & Brown, 2011; Torgerson & Torgerson, 2003).

Figure 2 PRISMA 2009 flow diagram of information throughout the systematic review (Moher et al., 2009)

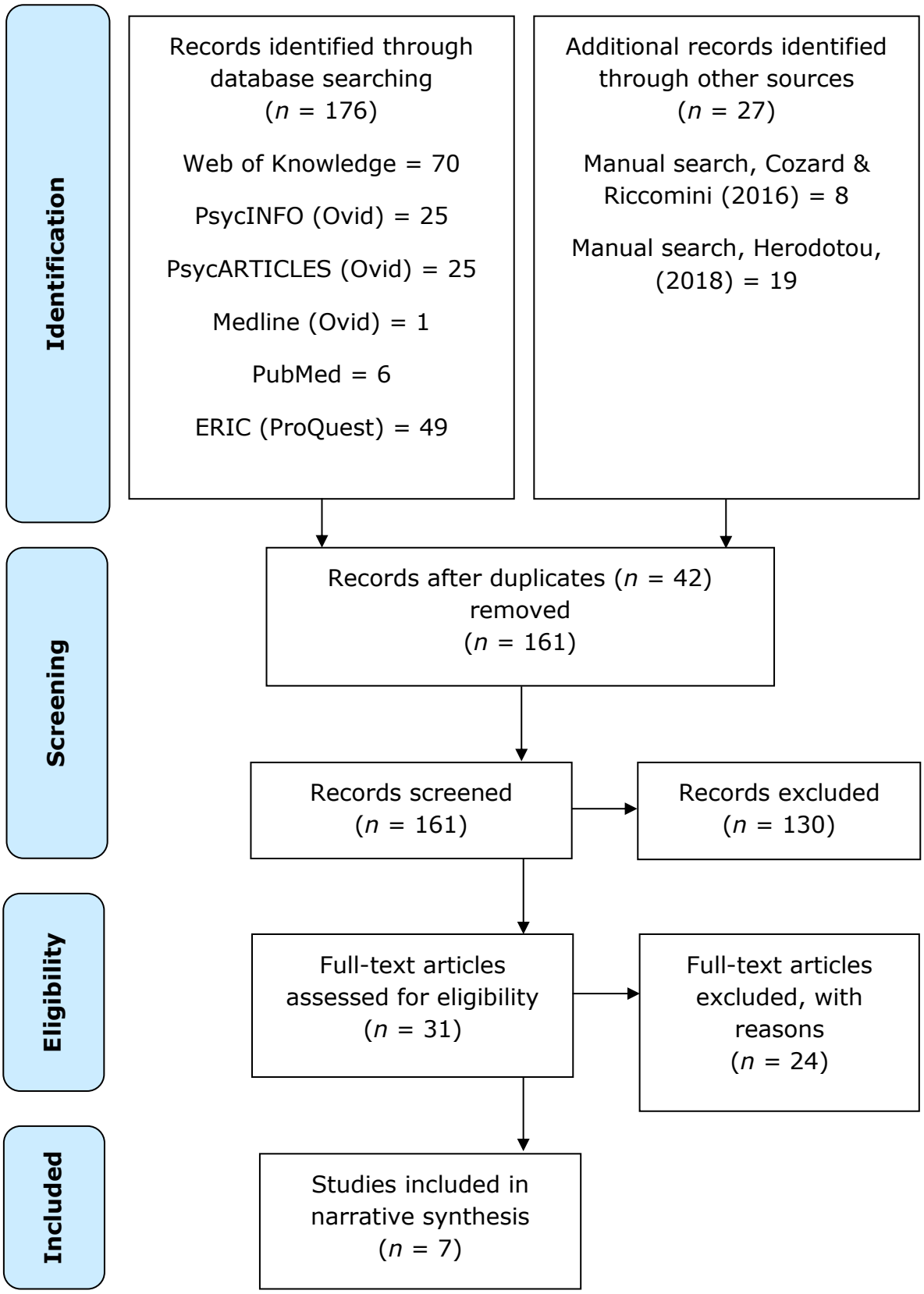


Table 4 The design of the quality assessment tool including definitions of different sources of bias in educational intervention research studies.

Type of bias	Definition	Quality assessment tool items
Allocation bias	A systematic difference in how participants were allocated to the intervention or control groups (Segdwick, 2013).	(1) Were the participants randomly allocated to groups? (2) Were details of the random allocation method included in the article? (3) Was the number of participants allocated to each group clearly stated?
Confounding bias	The potential presence of other variables which may influence the results, such as group differences in gender frequency, age, and mathematical ability at pre-test (Cheung & Slavin, 2013).	(1) Were the details of the comparability between the groups at pre-test reported in the article? (2) Were groups equivalent at pre-test with reference to an equal gender ratio, no significant age differences, and no significant differences in pre-test mathematics ability?

Type of bias	Definition	Quality assessment tool items
Performance bias	The possibility that the researchers or teachers implementing the intervention were aware of child's group allocation (Greenhalgh & Brown, 2011).	(1) Was the implementing team (researchers or teachers) blinded to group allocation throughout the study duration?
Detection bias	The possibility that the researchers conducting the study assessments are aware of the child's group allocation (Greenhalgh & Brown, 2011).	(1) Were the assessors at pre-test and post-test blinded to group allocation?
Attrition bias	Attrition refers to the proportion of participants in each group that withdraw from the study. Reasons for withdrawal should be stated as	(1) Were more than 80% of the participating children randomised to group (or in the initial sample if randomisation did not occur) included in the final data analysis?

Type of bias	Definition	Quality assessment tool items
	attrition may, in part be related to the intervention, for example non-compliance and thus may bias the study results (Torgerson & Torgerson, 2003).	(2) Were the reasons for children's withdrawal from the study stated?

Quality assessment results.

Table 5 summarises the quality of the seven included studies. Overall, the risk of bias was considered moderate (Higgins & Green, 2006). Specifically, the allocation bias risk was rated as moderate. While six of the seven studies randomly allocated participating children to either the treatment or control conditions, only two studies reported sufficient details of the random allocation procedure used. As such more detail is required in future study reports.

The risk of confounding bias was rated as low as all studies reported the in satisfactory detail the comparability between groups at pre-test based on gender, age, and pre-test mathematics ability. However, initial equivalence, particularly in terms of baseline achievement levels were not always established (Cheung & Slavin, 2013). There is always a chance that despite randomisation efforts, groups may still be different on key variables, such as pre-test mathematical ability (Torgerson & Torgerson, 2003). Certain methods, such as pseudo-randomisation (Uhry & Shepherd, 1993) can be utilised (e.g. Park et al., 2016; Szkudlarek & Brannon, 2018; van der Ven et al., 2017), however, these methods are costly and vulnerable to error (Torgerson & Torgerson, 2003). Alternative methods for example the choice of statistical test (Thomas & Zumbo, 2016; Van Breukelen, 2006) can be used to address the lack of initial equivalence between treatment and control groups (see Chapter 1).

The risk of performance and detection bias was rated as high as none of the seven included studies stated or achieved sufficient blinding of the assessors and/or the implementing team to group allocation. This level of blinding can be difficult to achieve in educational intervention studies due to the practical, hands on nature of the maths app interventions. However, measures should be taken to minimise this risk as much as possible.

Finally, the risk of withdrawal bias was considered low across the seven studies. The final sample only fell below 80% of the initial sample in one study; Schacter and Jo (2016) which experienced a 26% attrition rate (74% of the initial sample included in the final sample). Although detailed reasons for withdrawal were not given, the rate of attrition was considered in line with those reported by the participating preschools in the previous academic year (between 19% and 34%). Furthermore, the risk of potential withdrawal bias was reduced as the rate of attrition was consistent across the intervention (29%) and the control (28%) groups (Torgerson & Torgerson, 2003).

Table 5 Quality assessment of the seven studies included in the systematic review

Study	Randomisation (Allocation Bias)			Group Comparability (Confounding Bias)		Blinding (Performance & Detection Bias)		Withdrawals (Attrition Bias)	
	Random allocation	Detailed methods	Number stated	Comparability details	Initial equivalence	Implementers blinded	Assessors blinded	>80% in final analysis	Reasons stated
Schacter & Jo (2016)	N	NA	Y	Y	Y	NS	NS	N	P
Schacter & Jo (2017)	Y	NS	Y	Y	NS	NS	NS	Y	NS
Schacter et al. (2015)	Y	NS	Y	Y	Y	NS	NS	Y	NS
Park et al. (2016)	Y	Y	Y	Y	Y	P	NS	Y	Y

Study	Randomisation (Allocation Bias)			Group Comparability (Confounding Bias)		Blinding (Performance & Detection Bias)		Withdrawals (Attrition Bias)	
	Random allocation	Detailed methods	Number stated	Comparability details	Initial equivalence	Implementers blinded	Assessors blinded	>80% in final analysis	Reasons stated
Szkudlarek & Brannon (2018)	Y	NS	Y	NS	NS	NS	P	Y	Y
Van der Ven et al. (2017)	Y	NS	Y	Y	Y	NS	NS	Y	NS
Pitchford (2015)	Y	Y	Y	NS	NS	NS	NS	Y	Y

Note. Y = Yes; N = No; P = Partially; NS = Not stated; NA = Not applicable

Results

Study characteristics

Descriptive data was extracted from the seven included studies to synthesise study characteristics. A tabulated summary of the descriptive data is presented in Appendix B.

Study location and design.

The seven included studies were published between 2015 and 2018. Five of the studies were conducted in the USA, one in Europe, and one in Africa.

86% of the studies followed a RCT design where random allocation to condition was based either at the pupil (n = 4) or class level (also referred to as a cluster RCT; n = 2). The remaining one study followed a quasi-experimental design and did not randomly assign children to group. However, only two studies reported details of the methods used to randomly allocate participants.

Final sample size and characteristics.

All seven studies were conducted in a school based setting and had relatively large sample sizes, ranging from 86 to 433 children. The age of participating children ranged from 3 to 13 years. The inclusion of older children was as a result of one study conducted in Malawi, which follows an ability-based educational system and so repetition of school years is frequent (Hubber et al., 2016). 56% of studies worked with children aged 5 years or below.

Interventions and implementation.

Six of the studies examined the impact of researcher developed maths apps and the remaining one study evaluated a series of commercially available maths app software. Researchers typically implemented the maths app interventions (*n* = 4); teacher implementation was reported in only three studies. Teacher led implementation is critical for addressing ecological validity and generalisability of the intervention beyond the research context (Clements et al., 2015).

Control conditions.

The impact of the maths apps was typically compared to one control group (*n* = 5) with a range of control conditions, including other maths app software (*n* = 1), non-maths app software (*n* = 1), standard mathematics practice (*n* = 2), or specifically designed small group mathematics activities (*n* = 1). Two of the studies included two control conditions comparing the maths app intervention to non-maths apps and standard practice (*n* = 1) and to other maths and non-maths apps (*n* =1). The wide range of control conditions is beneficial for gaining a broad understanding of the impact of app-based mathematics instruction compared to other forms of instruction. How the different comparators address different research questions will be discussed in the narrative synthesis.

Study duration.

Across the seven studies there was a large range in study durations from 2 to 22 weeks. Time on task also varied across the included studies, ranging from 120 to 1200 minutes. Previous research shows increased time on task is associated with increased learning gains (Cheung & Slavin, 2013). This relationship will be explored in the narrative synthesis.

Measures of mathematical ability.

Mathematical ability was typically assessed using a researcher developed measure (n = 6). Only one study used an assessment measure independent from the maths app intervention at the focus of the evaluation. While it is not always practically possible to use standardised measures, particularly in low-income country settings (Laher & Cockcroft, 2017; Sabanathan et al., 2015), this does pose a threat to the internal validity of the current evidence as findings may be biased towards the intervention condition (Cheung & Slavin, 2013; Slavin & Madden, 2011).

Narrative synthesis

The differences across the studies (see Appendix B), including diversity in the comparison conditions, differences in time on task, and the inconsistency in reporting of outcomes limits the capacity for performing a valid and reliable meta-analysis (Brown & Richardson, 2017). Instead a narrative synthesis of the seven included studies was conducted.

As reported in Table 6, for each of the seven studies within-group effect sizes (Cohen's *d*) across the intervention period were calculated. Typically, an effect size of 0.2 is considered small, 0.5 is medium, and 0.8 or above is large (Cohen, 1988). This enabled some objective comparison of learning gains across the individual studies. The association between learning gains and time on task was also explored.

In addition, the emerging body of research was synthesised through a detailed evaluation of the research questions asked and the methodologies employed. This enabled important directions for future research to be highlighted, which will in part be addressed in the studies reported in this thesis.

Table 6 Analytical data extracted from the seven studies included in the systematic review. Within-group effects (Cohen's *d*) calculated from pre-test and post-test intervention group mean data reported in the individual studies.

Study	Participants (enrolled)	Intervention	Comparison	Effect size *d* (95% CI)	Total time on task (mins)
Schacter & Jo (2016)	227 low-income children aged 4-5 years	Math Shelf (researcher implemented)	Standard practice (teacher implemented)	1.22 (0.90 – 1.55)	300
Schacter & Jo (2017)	433 children aged 4-5 years	Math Shelf (teacher implemented)	Small group maths (teacher implemented)	NS	440
Schacter et al. (2015)	100 low-income children aged 4-5 years	Math Shelf (researcher implemented)	Other maths apps (researcher implemented)	0.90 (0.32 – 1.48)	180
Park et al. (2016)	109 children aged 3-5 years	Max's Math (researcher implemented)	Memory training game (researcher	0.35 (-0.20 – 0.90)	120

Study	Participants (enrolled)	Intervention	Comparison	Effect size *d* (95% CI)	Total time on task (mins)
			implemented)		
Szkudlarek & Brannon (2018)	158 children aged 3-5 years	Max's Math (researcher implemented)	123 Ninja & ABC Ninja (researcher implemented)	0.25 (-0.29 – 0.79)	120
Van der Ven et al. (2017)	103 children aged 6-7 years	Addition and subtraction fluency app (teacher implemented)	Standard practice (teacher implemented)	0.48 (-0.07 – 1.04)	300
Pitchford (2015)	318 children aged 6- 13 years	*Onebillion* maths apps (teacher implemented)	Non-maths apps & standard practice (teacher implemented)	CK = 1.38 (0.96 – 1.81) MC = 0.63 (0.24 – 1.02)	600 (S1)* 1200 (S2 & S3)

Note. NS = Not stated/ insufficient descriptive date reported to calculate within-group effect sizes for the intervention group; CK = curriculum knowledge assessment; MC = mathematical concepts assessment.

*S1 = Standard 1; S2 = Standard 2; S3 = Standard 3.

Schacter & Jo (2016).

Using a quasi-experimental design with non-randomised groups this study examined the impact of Math Shelf, a touch-screen tablet based app software, compared to standard mathematics practice with a final sample of 162 preschool children aged 4-5 years in the USA.

Math Shelf was developed by the first author and specifically targeted number skills, ranging from one-to-one counting to addition, subtraction, and place value (Clements & Sarama, 2009) using virtual manipulatives (Baroody, Eiland & Thompson, 2009; Jordan et al., 2012; Ramani & Siegler, 2008) based on the Montessori Method for early mathematics (Montessori, 1967). Similar to the direct instruction and play-based pedagogical theories discussed in Chapter 1 the Montessori Method incorporates deliberately sequenced activities with clear goals and feedback that are just beyond the child's current ability level (Lillard, 2005). Moreover, the software provides the opportunity for learner control and choice (Lillard, 2011).

173 children were allocated to the intervention condition and used Math Shelf for a total of 30 sessions lasting 10 minutes each, over a 15-week intervention period. The intervention was implemented by researchers. Results showed children who used Math Shelf made significantly greater learning gains on the researcher developed number skill assessment compared to children who received standard mathematics practice (n = 54).

This research is the first in a series of studies evaluating the Math Shelf software. As such a quasi-experimental design was appropriate at this stage in scaling (Carnine & Gersten, 2000). However, the lack of randomisation to intervention or control conditions can increase the risk of allocation bias (Greenhalgh & Brown, 2017) and consequently reduce the internal validity of the study findings (Shadish, Cook & Campbell, 2002).

Furthermore, it is unclear whether the app-based intervention was implemented as supplementary teaching aid, in addition to standard mathematics practice. Potential differences in time on task across the intervention and control groups render it difficult to disentangle the effects of the intervention from potentially increased time learning mathematics (Foster, Anthony, Clements, Sarama & Williams, 2016; Ginsburg & Smith, 2016). Nevertheless, this study is strengthened by the 15-week intervention period. In order to assess the full potential of educational interventions, a minimum 12-week implementation period is recommended (Higgins et al., 2012).

Schacter & Jo (2017).

To address some of the methodological issues in Schacter and Jo (2016), this study followed a cluster RCT design with a larger final sample of 378 preschool children aged 4-5 years in the USA and implemented the Math Shelf intervention for a longer 22-week period. In this study, classes of children were randomly allocated to either the treatment condition and received the Math Shelf app

software implemented by teachers (n = 260) or to the control condition and received research-based, hands-on mathematics activities delivered by the class teacher (n = 173; Van de Walle, Lovin, Karp & Bay-Williams, 2014). Both groups of children completed 44 sessions lasting 10 minutes each, twice a week for 22 weeks.

Consistent with Schacter and Jo (2016), results showed Math Shelf supported significant learning gains in number skill acquisition compared to small group based mathematics activities based on the same number concepts. This evidence suggests the extensive, individualised, child-centred, targeted mathematics practice, which are characteristic features of app-based mathematics instruction facilitated the mastery of basic number skills above and beyond the small group, teacher led mathematical activities. These findings are similar to previous research demonstrating the importance of targeted practice for early mathematical development (Fuchs et al., 2009; Kidd et al., 2013).

The strong methodological approach followed in this study, in particular, the equivalent time on task between the intervention and control groups enables the impact of the maths app to be directly compared to other forms of mathematics instruction. However, it is not clear whether the two conditions were implemented instead of standard mathematical practice or whether all children had increased exposure to mathematics instruction. This has critical practical implications. Although a mathematics practice gap has

been identified relative to literacy (Stacy et al., 2017) and research suggests early mathematics supports the learning of other subjects (Razel & Eylon, 1990; Sarama et al., 2012), it is important that a balance between mathematics instruction and other subjects and skills is achieved (Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2012). As such, research studies need to examine the efficiency of app-based mathematics instruction as part of a well-balanced curriculum (Evangelou, Sylva, Kyriacou, Wild & Glenny, 2009).

Furthermore, the cluster RCT study design can address potential spill over, contamination effects, by containing the intervention and control conditions in separate locations (Torgerson, 2001). Spill over effects occur when children allocated to the control condition may receive some exposure to the intervention and thus contaminate the results. However, as randomisation occurred at the class level, this study design is vulnerable to potential implementer effects. For example, studies suggest differences in teachers' experiences, confidence, and opinions are associated with technology use in the classroom (Blackwell, Lauricella, & Wartella, 2014). As such, this may influence the study outcomes and consequently pose a threat to the internal validity of the study (Shadish et al., 2002). Potential implementer effects in cluster RCTs can be addressed using statistical analyses that can control for clustering (Thomas, Crutcher, & Lorenzetti, 2007). Alternatively, randomisation can be conducted at the pupil-level within participating classes. This methodological approach was utilised in another study of the Math Shelf intervention.

Schacter et al. (2015).

In a pupil-level RCT, low-income, American, preschool children aged 4-5 years were randomly allocated to either the Math Shelf intervention group (n = 50) or to a control group (n= 50) that used a selection of other maths app software on touch-screen tablet devices.

Similar to Schacter and Jo (2017) the time on task was equal across the intervention and control groups. The allocated interventions were implemented by researchers in 18 sessions, lasting 10 minutes each, three times a week across a 6-week intervention period. Consistent with the previous Math Shelf studies, results showed in the final sample of 86 low-income children who used the Math Shelf intervention showed significantly greater number based learning gains compared to children in the control group who used a selection of five other maths app software programmes.

The inclusion of five other maths apps in the control condition addresses an important research question surrounding the educational quality of Math Shelf. The other studies have suggested the intervention is effective in comparison to standard practice (Schacter & Jo, 2016) and small group based mathematics instruction (Schacter & Jo, 2017). This study helps to address whether the learning gains are specific to the Math Shelf software or to app-based mathematics instruction more generally.

However, while the initial design can be commended, the current study is limited in its capacity to answer this question based on two caveats. Firstly, the five other maths apps were chosen based on popularity and positive reviews from technology based brands and parent forums. While teachers may typically rely on these forms of assessment as an indication of app quality, more formal evaluation is needed (Korat, Shamir & Segal-Dror, 2014; Wood, Grant, Gottardo, Savage & Evans, 2017). Assessments of the content and quality of app software should incorporate: (1) a mapping of the educational content to mathematical development and the school curriculum (e.g. Grant, Wood, Gottardo, Evans, Phillips, & Savage, 2012) and (2) a mapping of the app instructional features in line with Science of Learning principles that have been shown to underpin successful learning (e.g. Hirsh-Pasek et al., 2015).

Secondly, the use of five other maths apps in the control group, in comparison to one intervention, Math Shelf, may potentially dilute any learning gains specifically associated with the other individual maths app software. This internal threat to validity may be further amplified as children in the control condition were allowed to pick which maths app to use without any structure or scaffolding, which is characteristic of the Math Shelf software and is known to be important in the development of early mathematical skills (Vygotsky, 1984; Weisberg et al., 2016).

Overall, the three studies that have evaluated Math Shelf provide encouraging evidence with large effect sizes, Cohen's *d* >0.90, for the benefits of app-based mathematics instruction for supporting the acquisition of basic number skills in preschool children aged 4-5 years.

Schacter and Jo (2017) was conducted in private Catholic preschools that were not subject to the same regulations as federal or state funded preschools, which may increase the risk of selection bias and consequently reduce the generalisability of the individual study findings. Nevertheless, overall the three reported studies were conducted across a range of geographical and demographic settings in the USA and found consistent results. As such, collectively these studies can be considered transferable to the target preschool population.

Park et al. (2016).

A pupil-level RCT with preschool children aged 3-5 years in the USA was conducted to examine the impact of an app-based non-symbolic approximate arithmetic training programme (Max's Math, *n* = 51) compared to an app-based short-term memory training game (*n* = 52). Both training programmes were researcher developed. Children were allocated to conditions using a pseudo-randomisation procedure conducted by the first author, who remained blind to group allocation throughout the study.

Max's Math was designed to provide individualised, adaptive, and targeted practice in addition and subtraction number skills using

virtual manipulatives. Training tasks increased or decreased in difficulty based on the child's performance. In response to correct responses, children received positive feedback.

The allocated interventions were implemented by experimenters for 12 sessions, lasting 10 minutes each over a period of 2-3 weeks. In the final sample of 103 children, results showed children who used the app-based non-symbolic approximate arithmetic training intervention made significantly greater learning gains in basic symbolic mathematical ability in comparison to the control group receiving the app-based short-term memory training. In contrast, greater gains in short-term memory were seen in the control group compared to the intervention group. These results are consistent with previous research on app-based mathematics instruction (Schacter et al., 2015; Schacter & Jo, 2016; 2017) and targeted memory training programmes (Holmes et al., 2009; Melby-Lervåg & Hulme, 2013; St Clair-Thompson et al., 2010).

Furthermore, gains in other cognitive skills, including receptive vocabulary and executive function were not reported for either condition. This evidence suggests app-based instruction targeted to different domains (i.e. mathematics or short-term memory) supports benefits to the specifically trained skill only (Messer et al., 2018; Roberts et al., 2015; St Clair-Thompson et al., 2010).

Szkudlarek & Brannon (2018).

In a similar pupil-level RCT this study evaluated the same app-based non-symbolic approximate arithmetic training programme (Max's Math) with a final sample of 157 preschool children aged 3-5 years in the USA. This research extended Park et al.'s (2016) study through the inclusion of two comparison conditions. Specifically, the impact of the Max's Math app software (n = 53) was directly compared to two commercially available apps targeted at training number identification (123 Ninja; n = 52) and letter identification (ABC Ninja, n = 52). Consistent with the previous study, the total time on task was 120 minutes in each of the three conditions and the relevant intervention activities were implemented by the researchers.

This study conducted analyses at the whole sample level and with a sub-sample of children identified as 'low-achievers' based on their mathematics performance at pre-test.

Contrary to Park et al. (2016), results at the whole sample level, showed no significant differences in mathematical learning gains across the three groups. The authors offer two explanations for these divergent findings. Firstly, the comparison app software used in this study was suggested to be more engaging than the short- term memory training used in Park et al. (2016). However, without assessing the content and quality of each app (Korat et al., 2014; Grant et al., 2012; 2017; Hirsh-Pasek et al., 2015), this assertion is purely speculative.

Secondly, the primary outcome measure utilised in this study was based on the Number Sense Screener (NSS; Glutting & Jordan, 2012). The NSS assessment is an age appropriate, valid assessment of early number skills for children over the age of 5 years (Jordan, Glutting, Ramineni & Watkins, 2010). However, children in this study were younger, aged 3-5 years. Although this study modified the NSS in an attempt to make it suitable for younger children, the authors suggest this measure may not have been age appropriate and sufficiently sensitive for the study sample. In comparison, Park et al. (2016) used the Test of Early Mathematics Ability- Third Edition (TEMA-3; Ginsburg & Baroody, 2003), which is a valid, reliable, and age appropriate measure early mathematical skills for children aged between 3 and 9 years. As such, the divergent findings may be due to the outcome measure selection. Moreover, it emphasises the importance of using age appropriate measures to assess young children's mathematical ability in intervention research.

In contrast, in the sub-sample analyses this study reported children with low mathematical skills who used Max's Math showed significantly stronger basic mathematical learning gains compared to the control groups. The authors concluded the approximate arithmetic training is particularly beneficial for 'low-achieving' children. However, there is significant caveat with this assertion based on the regression to the mean phenomenon.

Children were identified as 'low-achievers' (n = 87) if their pre-test score was in the lower half of mathematics performance across the whole sample. This is problematic because using the same variable to define the low-achievers and to assess the impact of the intervention for these children renders the analysis vulnerable to regression to the mean and poses a threat to the internal validity of the findings (Barnett, van der Pols & Dobson, 2004). This is because extreme values, such as a low mathematics score at pre-test may be due to measurement error and are more likely to be followed by less extreme scores closer to the participant's true mean when assessed again at post-test (Barnett et al., 2004; Stigler, 1997). This renders disentangling a regression to the mean effect from a genuine intervention effect very difficult (Yudkin & Stratton, 1996). In order to minimise the regression to the mean problem, studies should utilise an external method for defining low-achievers that is independent from the outcome variable, for example, another measure of mathematical achievement, such as teacher reports (Yudkin & Stratton, 1996).

Overall, the two studies evaluating Max's Math report small effect sizes for the app-based mathematics intervention, Cohen's d >0.35. These findings add to the emerging evidence base on mathematical development and learning with app technology interventions for young children aged 3-5 years.

These studies are strengthened by their holistic, child-centred approach through examining the relationship between mathematics,

working memory, and targeted interventions (see Chapter 1). Consistent with previous research, both studies found short-term memory capacity did not improve in response to the maths app intervention (Messer et al., 2018). However, these studies, particularly Szkudlarek & Brannon (2018) should be treated with caution based on the methodological limitations.

Van der Van et al. (2017).

A cluster RCT conducted in the Netherlands, pseudo-randomised children aged 6-7 years to one of two conditions. The intervention group received an app-based addition and subtraction fluency training programme (n = 52). The control group received standard mathematics practice (n = 51). Both conditions were implemented by the class teacher.

Similar to Max's Math, the app-based addition and subtraction fluency training programme was designed to provide individualised, adaptive, and targeted practice in addition and subtraction number skills using virtual manipulatives in a game like format. This software also included positive and negative feedback and challenging learning activities.

In the final sample of 102 children, immediately following a 5-week intervention period, results showed the app-based mathematics instruction supported significantly stronger learning gains in dot- subtraction efficiency compared to the control group. The study concludes that the app-based mathematics instruction benefitted arithmetic fluency by improving children's skills in the

calculation process, instead of speeding up or increasing the retrieval of basic mathematical facts. This is consistent other intervention studies that have targeted basic mathematical skill acquisition and demonstrated evidence of generalisation to higher-level mathematical skills (Fuchs et al., 2009; Kidd et al., 2013).

However, this study, as well as the studies discussed so far, target number based early mathematical skills only. No studies have considered other components of basic factual and conceptual knowledge, such as pattern recognition and shape composition, which have been shown to underpin more complex mathematical understanding (Clements & Sarama, 2007; Rittle-Johnson et al., 2016). This broader view on early mathematical development was addressed in Pitchford (2015).

Pitchford (2015).

Pitchford (2015) conducted a pupil-level RCT in the first three years of formal education in Malawi. As Malawi follows an ability based educational system children's age ranged from 6 to 13 years. Children were randomly allocated to one of three conditions. The intervention group used a series of maths apps developed by *onebillion*, an educational not-for-profit organisation (n = 104). The maths apps were based on core, basic mathematical concepts in number, shape, space, and measure. A placebo control group used a selection of other non-maths, educational apps that did not involve mathematical concepts taught in the *onebillion* maths app intervention (n = 79). Both technology based conditions were

implemented by teachers with technical support from the charity Voluntary Service Overseas (VSO), who was a project implementing partner.

A standard practice control group continued to receive business as usual teaching (n = 100). In Malawi, standard mathematics practice typically consists of teacher led, face-to-face instruction. However, the educational system in Malawi suffers from a poor supply or lack of essential teaching and learning materials, a severe shortage of qualified teachers (Hubber et al., 2016) and very large classes sizes, with on average 76 children per class (Pitchford, 2015).

Following an 8-week intervention period and with a final sample of 283 children, results showed children who used the *onebillion* maths app intervention made significantly stronger learning gains in targeted curriculum knowledge and mathematical concepts not explicitly targeted by the intervention, in comparison to the two control groups. In line with the other six studies in the systematic review, this evidence suggests app-based mathematics instruction can effectively support early mathematical development in different educational contexts.

Furthermore, these results are consistent with other studies that found this maths app intervention supported the acquisition of early mathematical skills for children with special educational needs in Malawi (Pitchford, Kamchedzera, Hubber, & Chigeda, 2018) and with children aged 4-7 years at the start of school in the UK

(Studies 1-3 Outhwaite et al., 2017). However, these studies were not explicitly included in this systematic review due to the correlational design (Pitchford et al., 2018) and the lack of an adequate control group (Outhwaite et al., 2017; Pitchford et al., 2018).

Pitchford's (2015) study is strengthened by its novel and rigorous three-group RCT design. The inclusion of two control groups enabled the impact of the mathematical educational app software to be differentiated from the touch-screen tablet device hardware (placebo control group) and the effects of natural maturation (standard practice control group). This supports the internal validity of the study findings.

However, the mathematical assessments used in this study were delivered on tablets and so may have been biased in favour of the technology based conditions. To address this issue, older children in Standard 3 (8-9 years) also completed a paper and pen version of the curriculum knowledge assessment. Results showed the learning gains found in the tablet based assessment generalised to the different delivery format.

Within-group effect sizes and time on task.

To explore the relationship between learning gains in response to the specific maths app intervention and time on task a partial correlation analysis was conducted controlling for the number of children who used the relevant maths app. This control was implemented to minimise potential inflation bias, as previous

reviews shows smaller samples sizes are more likely to report larger effect sizes (Cheung & Slavin, 2013). Results showed a small, positive correlation between greater learning gains and increased time on task (r (4) = .35, p = .503). The lack of statistical significance was most likely due to an underpowered study sample (n = 7). Nevertheless, this exploratory finding is consistent with previous research that suggests time on task impacts learning gains (Cheung & Slavin, 2013).

Discussion

This chapter reports a systematic review evaluating the impact of app-based educational interventions for supporting mathematical development and learning in young children during the first three years of formal education. Seven studies met the inclusion criteria and were included in the narrative synthesis. Overall, the results showed stronger mathematical learning gains for children who used the app-based mathematics instruction compared to a range of comparison groups. Moreover, increased time on task was shown to be associated with increased learning gains. Collectively, the seven included studies demonstrate the potential of app-based mathematics instruction as a form of early intervention for young children at the start of school (Cooper, 2005; Flewitt et al., 2015; Kucirkova, 2014).

Limitations of current review

This systematic review gives a useful insight into the current literature on app-based mathematics instruction. However, the review may be limited in two ways.

First, other relevant studies may have been overlooked as the search strategy focused on peer reviewed journal articles from six databases (Web of Science, PsycINFO, PsycARTICLES, Medline, PubMed, and ERIC). The additional backwards citation search of two relevant previous reviews (Cozard & Riccomini, 2016; Herodotou, 2018) aimed to mitigate this issue. However, future reviews should expand the search strategy to include grey literature. The inclusion of grey literature in systematic reviews reduces the risk of publication bias, which may influence the review findings (Connolly et al., 2018) as studies with negative or non-statistically significant results are not commonly published (Rosenthal, 1979). Grey literature sources relevant to education include ProQuest (unpublished theses database), the Educational Endowment Foundation, and the What Works Clearinghouse (Connolly et al., 2018).

Second, despite the clear, precise inclusion criteria, the use of one reviewer means the risk of subjectivity influencing study selection still remains (McDonagh, Peterson, Raina, Chang, & Shekelle, 2013). This potential selection bias can be addressed with a dual review process where two reviewers independently complete the study selection procedures and the level of agreement between

researchers can be formally assessed (Eden, Wheatley, McNeil, & Sox, 2008). This dual review process helps to clarify the inclusion criteria and resolve any potential misunderstandings, therefore ensuring the included studies are truly reflective of the review aims and selection criteria (McDonagh et al., 2013).

Future research directions

Despite the emerging experimental evidence base discussed in this systematic review, important questions still need to be addressed surrounding how does app-based mathematics instruction work, who does it work for, and under what circumstances does it work (Connolly et al., 2018; Pawson & Tilley, 2004; Slavin, 2012)? This thesis will address these questions through evaluating the same *onebillion* maths app intervention reported in Pitchford (2015). A detailed overview of the maths apps is given in Chapter 3.

How does the intervention work?

In order to understand how the maths app intervention works it is important to consider componential theories of mathematical development and factors known to be associated with early mathematical development, specifically memory capacity and proficiency in the language of instruction.

Theories of mathematical development.

Theories suggest mathematical development requires the acquisition of different component skills and processes that range in level of difficulty (Goswami, 2006; Holmes & Dowker, 2013; see

Chapter 1). Consequently, there is a need to examine which components of mathematical development are supported by educational maths apps. As reported in this systematic review, previous research evaluating maths app interventions have frequently used assessments closely aligned with the intervention content, which typically focuses on specific aspects of mathematical knowledge. For example, the studies evaluating the Math Shelf intervention measured learning gains in number skills targeted by the intervention using a researcher developed assessment (Schacter & Jo, 2016; 2017; Schacter et al., 2015). Future studies are required to take a broader view of mathematics and consider how educational apps may support the acquisition of targeted components of mathematical knowledge and whether this generalises to higher-level skills. This will help elucidate how mathematical development is supported by interactive, individualised, educational apps.

Factors that influence mathematical development.

A range of child-level, cognitive and language factors have been shown to be associated with early mathematical development, including poor memory capacity (Cragg et al., 2017; Gathercole & Alloway, 2006; Passolunghi & Siegel, 2004; Raghubar et al., 2010) and low proficiency levels in the language of instruction (Kempert et al., 2011; Snow & Uccelli, 2009; Morgan, 2013). No studies included in this systematic review considered the potential impact of cognitive factors on children's learning gains with the app-based

mathematics intervention. Therefore, based on a determinant, ecological framework (see Chapter 1), future studies need to consider how child-level factors may be associated with children's learning gains.

Who does the intervention work for?

To understand who the maths app intervention works for most effectively prior research suggests it is important to examine the relationship between a child's socio-economic status and progress with the maths apps. Further research with different achievement based groups of children is also needed.

Socio-economic status.

Three studies included in the systematic review specifically targeted the app-based mathematics instruction to children from low-income areas (Pitchford, 2015; Schacter and Jo, 2016; Schacter et al., 2015). Although demographic data was included in all of the included studies, which gives an indication of the generalisability of the study findings, no previous research has examined the association between a child's SES and mathematical learning gains with the app-based interventions. This line of enquiry is of interest given the SES related achievement gap evident at the start of formal schooling (Johnson et al., 2007; Jordan et al., 2006).

Different groups of children.

Previous intervention research presented in this systematic review predominantly focused on the specific study populations and

did not consider for whom the intervention may work most effectively. One study attempted to address this and re-analysed a sub-sample of children identified as 'low-achievers.' This analysis aimed to examine whether the maths app instruction was particularly beneficial for children struggling to learn mathematics relative to their peers (Szkudlarek & Brannon, 2018). However, as outlined earlier this analysis method is vulnerable to the regression to the mean phenomena (Barnett, et al., 2004; Stigler, 1997). Future research needs to utilise an independent, external method for defining low-achievers, such as teacher reports (Yudkin & Stratton, 1996) and examine how these children respond to the new maths app intervention.

Under what circumstances does the intervention work?

When considering the circumstances in which app-based mathematics instruction is most effective it is vital to look at how the maths apps are integrated and implemented in a classroom setting.

Intervention integration.

As highlighted in this systematic review, the app-based mathematical interventions (and control conditions) have been typically integrated into the classroom routine as a supplementary teaching aid and are typically implemented by researchers (Park et al., 2016; Schacter & Jo, 2016; Schacter et al., 2015; Szkudlarek & Brannon, 2018). This entails greater instructional time on learning mathematics compared to the comparison groups, making it difficult

to disentangle the effects of the intervention from the effects of extra time learning mathematics (Foster et al., 2016; Ginsburg & Smith, 2016). As such, future studies need to include a time-equivalent, business as usual control group (e.g. Holmes & Dowker, 2013). Furthermore, it is critical that teachers implement the intervention, to ensure high ecological validity and to support the generalisability of the intervention beyond the research context (Clements et al., 2015).

Implementation variability within school ecosystems.

Consistent with previous systematic reviews (e.g. Connolly et al., 2018), none of the seven studies in this systematic review included an implementation process evaluation component. As discussed in Chapter 1, intervention process evaluation is important for understanding the association between implementation variability within individual school ecosystems (Durlak & DuPre, 2008) and intervention outcomes (Askell-Williams et al., 2013; Hansen et al., 2013; Humphrey et al., 2016). Future studies should adopt a mixed-methods approach to holistically understand the impact and implementation of the particularly intervention (Connolly et al., 2018; Thomas, 2016; Shaffer, 2011).

Thesis aims

To address the avenues for future research highlighted in this systematic review, this thesis aims to examine the reach and utility of a maths app intervention. This research aims to go beyond

asking ‘are the maths apps an effective form of mathematics instruction?’ (Haynes et al., 2012; Torgerson & Torgerson, 2001) by adopting a multi-level, ecological, determinant framework (see Chapter 1). The thesis aims to evaluate the maths apps in line with the three general themes of; how does the intervention work, who does the intervention work for, and under what circumstances does the intervention work (Connolly et al., 2018; Pawson & Tilley, 1997; Slavin, 2012)? Table 7 summarises the general and specific research questions asked in this thesis.

Table 7 Summary of general and specific research questions to be addressed in this thesis

Research Questions	Thesis Chapter
Under what circumstances do the maths apps work?	
How are the maths apps most effectively implemented in a classroom setting compared to standard instructional practice?	Chapter 6: UK Feasibility RCT
How are the maths apps implemented as part of an individual school's ecological system and is there a relationship between the identified implementation themes and children's learning gains with the maths apps?	Chapter 7: UK RCT Implementation Evaluation
How do the maths apps work?	
Is short-term memory capacity associated with children's progress with the maths apps?	Chapter 4: UK Proof of Concept
Is app-based mathematics instruction more effective at supporting the acquisition of basic mathematical skills when delivered in the child's first (L1) or second language (L2), in comparison to standard classroom practice only in	Chapter 5: Brazil Proof of Concept

Research Questions	Thesis Chapter
a dual immersion, bilingual educational context and does proficiency of the language of instruction influence mathematical learning gains and progress within the maths apps?	
Which components of mathematical developments are supported by the maths apps?	Chapter 6: UK Feasibility RCT
Who do the maths apps work for?	
Does the maths app intervention support the acquisition of early mathematical skills for children aged 4-5 years old and identified as low-achievers compared to their typically attaining peers receiving standard instructional practice?	Chapter 4: UK Proof of Concept
Is SES associated with children's progress with the maths apps?	Chapter 4: UK Proof of Concept

Epistemological position

In order to examine the research questions outlined in Table 7 in a real-world research context this thesis adopted a mixed-methods approach, orientated around the philosophy of pragmatism to holistically evaluate the new maths app intervention (Teddlie & Tashakkori, 2010).

As discussed in Chapter 1, RCTs and quasi-experimental methods are valued in evidence-based education to understand 'what works' (Connolly et al., 2017; 2018 Haynes et al., 2012; Torgerson & Torgerson, 2001). These primarily quantitative, experimental methods typically draw upon positive or post-positivist paradigms. The post-positivist worldview holds a deterministic philosophy whereby the relationship between cause (e.g. the intervention) and effect (e.g. learning gains) can be established through empirical observation and measurement (Creswell & Creswell, 2018). However, as also outlined in Chapter 1 this approach has been criticised for removing educational interventions and outcomes from the complex context in which they are situated in (Biesta, 2010; Elliott, 2004; Thomas, 2016).

Within the pragmatic worldview the choice of research methods is driven by what works best to answer specific research questions (Johnson & Onwuegbuzie, 2004; Mertens, 2015). As such Chapters 4-6, which report quasi-experimental and RCT studies conducted in the UK and Brazil, follow a primarily positivist or post-positivist paradigm. In contrast, Chapter 7, which reports an

evaluation of the maths app intervention, adopts a more relativist, primarily qualitative approach.

This use of complimentary, pluralistic approaches aims to maximise the strengths and minimise the weaknesses of the individual research methods utilised (Johnson & Onweugbuzie, 2004). It affords a greater depth of enquiry compared to a single method used in isolation (Sammons, 2010) and can address the multiple dimensions of the same phenomena (Sale, Lohfeld, & Brazil, 2002); how does the intervention work, who does the intervention work for, and under what circumstances does the intervention work (see Table 7)?

Conclusion

In summary, Chapter 2 reports a systematic review on the use of educational maths apps to support early mathematical development for children at the start of formal schooling. Consistent with previous reviews (e.g. Cozard & Riccomini, 2016; Crompton et al., 2017; Haβler et al., 2015; Herodotou, 2017), the narrative synthesis of the seven eligible studies provides further support for the potential of app-based mathematics instruction as a form of early intervention (Cooper, 2005; Flewitt et al., 2015; Kucirkova, 2014).

The systematic review also identified important directions for future research, which provide the grounding for the research questions investigated in this thesis (see Table 7). Specifically,

through a mixed-method, pragmatic approach, this thesis adopts a multi-level, ecological, determinant framework and aims to further address how does the new maths app intervention work, who does the intervention work for and under what circumstances it does it work (Connolly et al., 2018; Pawson & Tilley, 2004; Slavin, 2012)?

Chapter 3: Maths App Intervention

This thesis aims to examine the reach and utility of a new maths app intervention in supporting the acquisition of early mathematical skills for young children at the start of formal education in the UK and Brazil. The intervention at the focus of this investigation consisted of two maths apps, 'Maths 3-5' and 'Maths 4-6.' The maths apps were developed by *onebillion,* an educational not-for-profit organisation (www.onebillion.org)[2]. This chapter will give an overview of the curriculum aims of the *onebillion* maths apps, followed by a detailed assessment the content and quality of the app software.

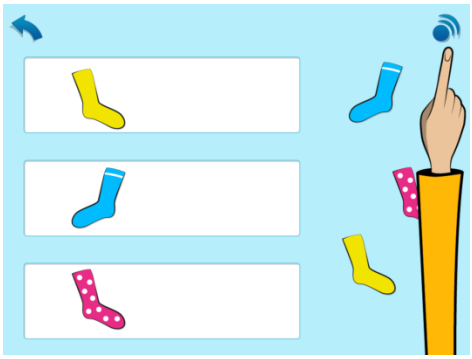
Overview of maths app intervention

The maths apps are designed to deliver child-centred tuition through interactive picture, audio, and animation formats with clear objectives, instructions, and immediate formative feedback, consistent for all users. Children work through the apps individually with headphones, at their own pace, and have the opportunity to repeat instructions and activities as often as desired. To complete a maths app topic (see Table 8), children need to achieve 100% pass rate on the end of topic quiz included in the software. The quizzes are designed to assess children's knowledge of the mathematical concepts covered in the topic activities.

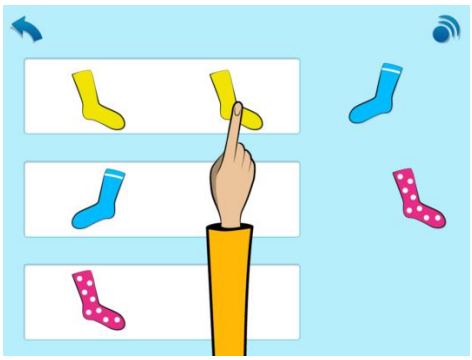
[2] The relationship with *onebillion* was established by the primary supervisor, Professor Nicola Pitchford. *Onebillion* supported this research through in-kind donations of the maths app software to all participating schools. The research team were independent from and did not have a financial interest in *onebillion.*

For example, in Topic 1 of ‘Maths 3-5’ children are taught the concept of sorting and matching through a range of activities involving sorting and matching different items by type, shape, size, and colour. Screenshots of example activity items and task instructions for Topic 1 are illustrated in Figure 3. After completing seven sets of activities, children reach the end of topic quiz that includes 10 questions from the previous learning tasks. When children pass the quiz they are awarded a certificate and progress to the next topic. Overall, the maths apps provide an opportunity for targeted mathematical practice.

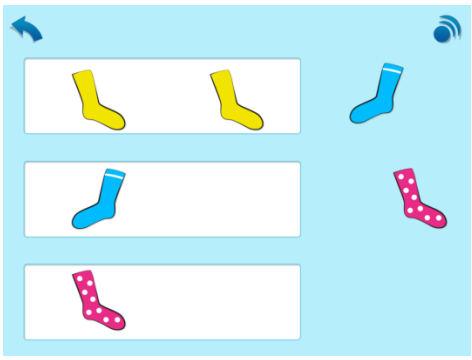
Figure 3 Example item and task instructions for Topic 1, Sorting and Matching.



"Remember you can touch this to hear me again"



"Now look, yellow sock, yellow sock"



"Your turn, drag the other socks into the boxes to match them"

Repeatable instruction:

"Drag the other socks into the boxes to match them"

In order to assess the content and quality of the maths app software a two strand mapping exercise was undertaken by the researcher: (1) the maths app educational content was mapped to theories of mathematical development and the UK based school curriculum (Grant et al., 2012). An additional curriculum mapping was conducted specific for the private bilingual school that participated in the Brazil Proof of Concept study, reported in Chapter 5. (2) The maths app instructional features were also mapped to the Science of Learning principles (Hirsh-Pasek et al., 2015) and learning strategies proposed to reduce demands on working memory (Gathercole & Alloway, 2007) and support bilingual children's developing language proficiency (Cummins, 2008).

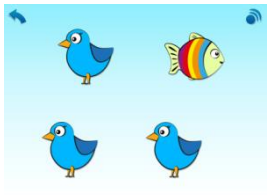
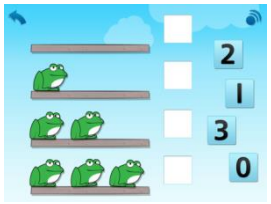
Maths app curriculum content

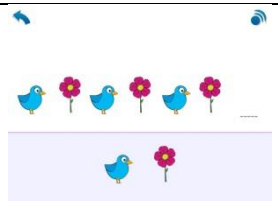
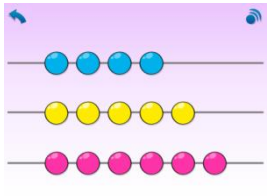
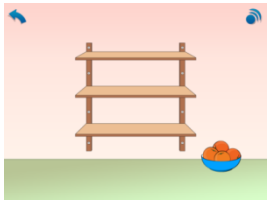
The specific learning goals within the maths apps primarily target the acquisition of basic mathematical facts and conceptual knowledge, for example number bond combinations, properties of shape and patterns, and simple numerical operations. A detailed outline of each maths app topic and how it aligns with componential theories of mathematical development (Goswami, 2006; Holmes & Dowker, 2013) is reported in Table 8.

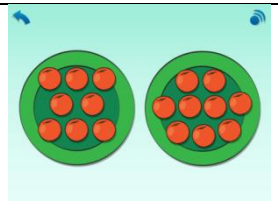
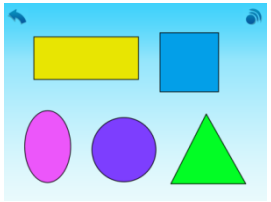
Table 8 also outlines how the maths apps topics are aligned with the mathematics curriculum followed in the UK at the start of formal education. In particular, the apps cover core mathematical concepts in Number, Shape, Space, and Measure included in the Early Years Foundation Stage (EYFS) Profile for children aged 4-5

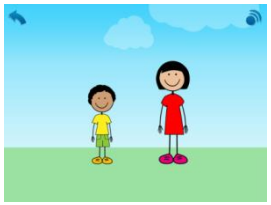
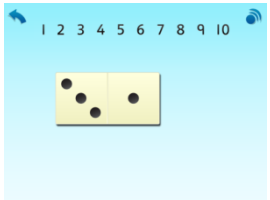
years old (Department for Education, 2013). The maths apps also start to introduce children to topics included in the UK National Primary Curriculum for Key Stage 1 for children aged 5-7 years old (Department for Education, 2014). Specific alignment between the maths apps and the mathematical curriculum followed by the Brazilian private school participating in the Brazil Proof of Concept Study is reported in Chapter 5.

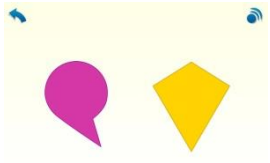
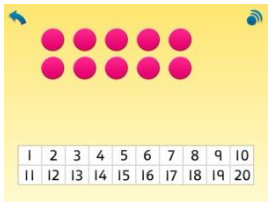
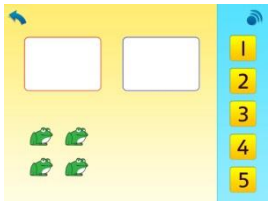
Table 8 Maths apps content mapped to the UK mathematics curriculum for EYFS and KS1 and components of mathematical development.

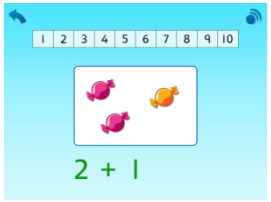
App Topic	Example Activity Item and Verbal Instructions	EYFS Profile Maths Curriculum	Maths Component
Maths 3-5 (10 topics with seven activities each)			
Sorting and Matching	 *"Which is the odd one out? Cross it out"*	Shape, Space, and Measure • Interest in shape and space by playing with shapes and making arrangements with objects • Awareness of similarities in shape	Factual knowledge
Counting to 3	 *Put the numbers in place"*	Number • Uses some number names • Recognise and recite numbers in order • Count objects that cannot be moved	Factual knowledge

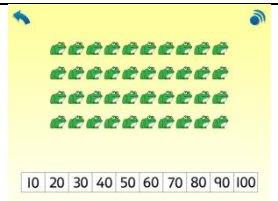
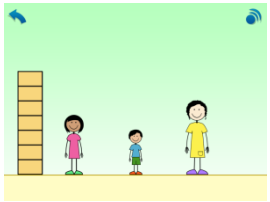
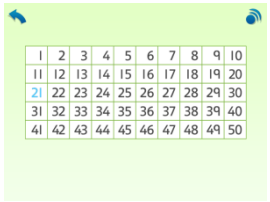
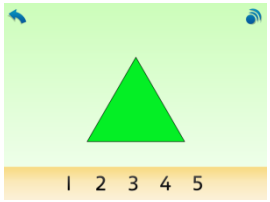
App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
Lines and Patterns		*"Which one comes next?"*	Shape, Space, and Measure • Uses objects and shapes to create and recreate patterns and build models	Factual knowledge
Counting 4 to 6		*"Touch the wire with four beads"*	Number • Begin to represent numbers with external objects • Select correct numerals 1 to 5 • Count up to 6 objects using correct number name for each item	Factual knowledge
Where is it?		*"Put the bowl of oranges on the top shelf"*	Shape, Space, and Measure • Uses relative positional language e.g. behind/ next to	Factual knowledge

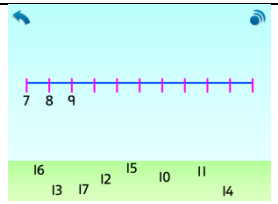
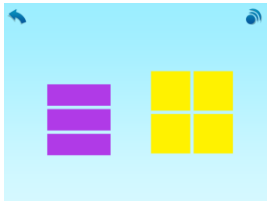
App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
Counting 7 to 10		*"Now, which plate has eight oranges? Touch it"*	Number • Identifying numbers in a set • Match number and quantity	Factual knowledge
Patterns and Shape		*"Choose the triangle"*	Shape, Space, and Measure • Select a named shape	Factual knowledge
Numbers 1 to 10		*"Which number is one more than 8?*	Number • Recite numbers, select correct numerals and count objects 1 to 10 in regular and irregular arrangements • One more and one less than a given number up to 10 • Compare two groups that have the	Factual knowledge

App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
			same number	
Comparing		*"Which child is taller?"*	Shape, Space, and Measure • Order items by length, height, weight, and capacity • Use 'more' and 'fewer' to compare two sets of objects	Conceptual understanding
Adding and Taking Away		*"Count all the spots on the domino, then touch the number"*	Number • Interest in number problems • Find total number of items in two groups by counting • Addition and subtraction vocabulary	Conceptual understanding

App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
Maths 4-6 (12 topics with six activities each)				
Shape and Position		*"Which shape is symmetrical?"*	Shape, Space and Measure • Describe position and distance • Identify line of symmetry in shapes*	Factual knowledge
Counting to 20		*"How many counters are there? Touch the number"*	Number • Count reliably 1 to 20	Factual knowledge
Sharing		*"Four frogs, share them into two equal sets…how many frogs are there in set?"*	Number • Separate a group of objects in different ways and recognise the total is the same	Conceptual understanding

App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
More Counting		*"Choose the fifth house"*	Number • Place numbers in order • Count in multiples of 2* • Odd and even numbers*	Factual knowledge
Telling the Time		*"…Put the pictures in the right order"*	Shape, Space and Measure • Use everyday language related to time • Measure short periods of time in simple ways • Order and sequence familiar events	Mathematical reasoning
Add and Subtract		*"Add 1 sweet to the set, how many sweets are there now?"*	Number • Add and subtract two single digit numbers	Conceptual knowledge

App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
Count in 10s and 5s		*"The frogs are in 10s, how many frogs are there?"*	Number • Count in multiples of 5s and 10s*	Factual knowledge
How tall, how long?		*"Which child is 4 bricks tall?"*	Shape, Space and Measure • Recognise shapes and objects as 'tall' • Compare length and height	Conceptual understanding
Count to 100		*"Find 22"*	Number • Count to 100*	Factual knowledge
2-D Shapes		*"How many corners does the triangle have?"*	Shape, Space and Measure • Use mathematical names to describe 2D shapes	Factual knowledge

App Topic	Example Activity Item and Verbal Instructions		EYFS Profile Maths Curriculum	Maths Component
Number lines		*"Put the other numbers in the right places"*	Number • Count reliably 1 to 20 • One more and one less than a given number up to 20	Factual knowledge
Fractions		*"Chose the one you cut into quarters"*	Number • Recognise, find, and name fractions of object, space or quantity*	Conceptual understanding

Note. EYFS = Early Years Foundation Stage; Asterisk refers to Key Stage 1 topics.

Instructional features

The *onebillion* maths apps also map highly onto the four Science of Learning principles of active, engaged, meaningful, and socially interactive learning (Hirsh-Pasek et al., 2015; see Table 9). In addition, these instructional features are aligned with learning strategies suggested to reduce demands on working memory (Gathercole & Alloway, 2007) and support developing language proficiency (Cummins, 2008).

Active learning

As illustrated in Table 9, active, minds-on learning in the maths apps is fostered through the direct manipulation of virtual objects, verbal labels, and numerical representations (Lindahl & Folkesson, 2012). The simultaneous presentation of auditory and visual input engenders multi-sensory learning, which research shows can help facilitate children's understanding (Pavio, 1986; Carr, 2012). Furthermore, the multi-touch nature of the tablet device affords interactive, physical experiences with mathematical content, characteristic of cognitive embodiment (Clark, 2008). Studies indicate cognitive embodiment supports mathematical development and learning (Calder, 2016; Duijzer et al., 2017). In the example illustrated in Table 9, children must purposively engage with the maths apps by tapping the correct answer to the question and progress to the next item.

Engaged learning

Engaged learning is supported by immediate feedback (positive or negative) given after every interaction with the maths apps and external rewards for correct responses (see Table 9). This highlights the potential of app-based maths instruction for supporting children's extrinsic motivation (Couse & Chen, 2010).

When using the maths apps children can also regulate their pace of learning within their own in-app profile (see Table 9). The child's individual in-app profile saves their progress throughout the apps, which can be monitored by the implementing teachers and the research team. As discussed in Chapter 1, these features can support learner autonomy shown to be critical for intrinsic motivation (Malone, 1981; Przybylski et al., 2010) and effective for supporting children's educational outcomes (Morrison, Ross, & Baldwin, 1992). It can also provide effective scaffolding for pupils with differing needs and serves to create an individualised learning environment (Adams & Carnine, 2003; Slavin & Lake, 2008).

Meaningful learning

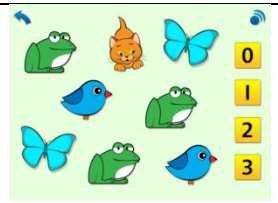
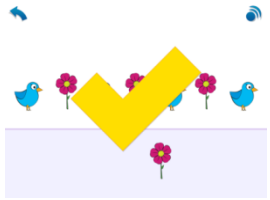
Meaningful learning through the maths apps is promoted through a staged curriculum that builds on the child's previous knowledge (Magliaro, Lockee, & Burton, 2005) and extends children beyond their current ability level (Vygotsky, 1978; Inal & Cagiltay, 2007; see Tables 8 & 9). The maths apps also include continuous assessment of knowledge acquired through the different topics taught through the end of topic quizzes (see Table 9). This feature

engenders retrieval-based learning, shown to improve learning outcomes (Grimaldi & Karpicke, 2014; Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Furthermore, the end of topic quizzes help to assess whether a child has successfully acquired the intended mathematical concept, which is sometimes challenging in play-based pedagogies (Sarama & Clements, 2009).

Socially interactive learning

Socially interactive learning is also evident in the maths apps through the on-screen teacher proving demonstrations and short, clear, task instructions (see Table 9; Troseth et al., 2006). These instructional features can help offer the efficient and effective delivery of one-to-one instruction, which has been shown to be a vital component of effective mathematical interventions (Holmes & Dowker, 2013).

Table 9 Maths apps mapped to the principles of active, engaged, meaningful, and socially interactive learning.

Visual stimuli	Verbal stimuli	Science of Learning principles
 Maths 3-5, Topic 2 'Counting to 3,' Activity 5	"*How many frogs are there? Count then touch the number.*"	**Active learning** Direct manipulation of virtual objects, verbal labels, and numerical representations with auditory and visual presentation of information. Children purposively tap the correct answer to progress to the next item.
 Maths 3-5, Topic 3 'Lines and Patterns,' start of Activity 1	Positive tone for correct answer. Negative tone for incorrect answer.	**Engaged learning** Children receive immediate positive feedback after each correct response. Children hear a negative tone for incorrect responses and children must try again to achieve correct response and progress to next activity item.

Visual stimuli	Verbal stimuli	Science of Learning principles
 Maths 3-5, Topic 3 'Lines and Patterns,' end of Activity 1	Other children cheering.	**Engaged learning** Children receive meaningful rewards upon successful completion of maths app activity. Other rewards include a certificate upon successful completion of the end of topic quiz.
 Example visuals of in app profile for Maths 3-5, Topic 4, 'Couting 4 to 6'	*"Choose the flashing button."*	**Engaged learning** Children have their own in-app profile which saves their progress and may promote self-paced learning and autonomy.

Visual stimuli	Verbal stimuli	Science of Learning principles
 In app visuals of Maths 3-5 topic structure	*"Welcome back to Maths. Now you can play some more. This is about counting from 4 to 6."*	**Meaningful learning** The maths apps are structured around a staged curriculum, which builds on children's prior knowledge (see Table 8 for detailed outline).
 Maths 3-5, Topic 3 'Lines and Patterns,' example end of topic quiz question	*"Which comes next in this pattern? Drag it into place and then touch the arrow."*	**Meaningful learning** The end of topic quizzes included in the maths apps provide continuous assessment of knowledge and engender retrieval based learning.
	*"Now look. It's a heart. Look at this half. This half is the same but*	**Socially interactive learning** The maths apps include an in-app teacher, who provides clear task demonstrations and instructions, which can be

Visual stimuli	Verbal stimuli	Science of Learning principles
Maths 4-6, Topic 1 ‘Shape and Position,’ example In-app teacher demonstration	*the other way around. One half is the reflection of the other, like in a mirror. We say this shape is symmetrical.”*	repeated as required.

Note. Italicised quotations report transcript of in-app teacher.

Support for language proficiency and working memory

In a bilingual educational context, for example the Brazil Proof of Concept study (Chapter 5), the active and socially interactive learning app features discussed above may be potentially supportive of the child's cognitive academic language proficiency (CALP). For example, the visual task demonstrations and verbal instructions provided by the on-screen teacher, combined with the interactive virtual objects, verbal labels, and numerical representations may afford a relatively high degree of contextual support for children to derive meaning about the learning activity.

The maths apps also have the potential to support task delivery with an appropriate level of cognitive demands of the learning activity. For example, the apps include clear, specified learning goals with simple, stepwise task instructions from the on-screen teacher, which children can repeat if required (see Figure 3). This may provide a suitable learning strategy beneficial for bilingual children (see Cummins, 2008) and children with a low memory capacity (Gathercole & Alloway, 2007) by providing easy access to mathematical instruction.

Theories of learning

Overall, the *onebillion* maths apps can been seen to provide targeted mathematics practice and combines direct instruction principles (Kirschner et al., 2006), in particular, a stepped curriculum, rehearsal, contingent feedback, and rewards, with aspects of play-based pedagogies (Gray, 2015), such as, choice,

self-regulation, and control. The apps offer intensive, systematic, and explicit instruction, which is characteristic of effective intervention approaches, such as Response to Intervention (RTI).

RTI is a multi-tiered intervention approach employed in the USA that increases in intensity as children move across the tiers and is targeted for children at risk of educational underachievement (Fuchs & Fuchs, 2006). It reflects the graduated response models employed in the UK (See, Gorard, & Siddiqui, 2015). Greater intensity can be achieved by systematic and explicit instruction and small group or individually based delivery: features included in the new maths app intervention. However, unlike the higher end of the RTI approaches the maths apps require minimal implementation requirements and so offer the opportunity for increased intervention access for all children.

Chapter 4: UK Proof of Concept Study

Current study

As outlined in Chapter 1, in order to understand the impact of technology, specifically educational maths apps for improving young children's learning outcomes and supporting early mathematical development (Department for Education, 2018), empirical research is required. Results from the systematic review reported in Chapter 2 suggests that app-based mathematics instruction can support early mathematical development. However, no study to date has adopted a determinant systems approach (see Chapter 1) and investigated how additional factors known to influence mathematical development, such as short-term memory and SES, impact the efficacy of app-based mathematics instruction in early education. Furthermore, previous initial evidence examining the same maths apps at the focus of this thesis (see Chapter 3) lacked an adequate control group (Studies 1-3, Outhwaite et al., 2017).

To address these limitations and build the initial evidence, the current study addressed two important questions: (1) does the maths app intervention support the acquisition of early mathematical skills for children aged 4-5 years old and identified as low-achievers compared to their typically attaining peers receiving standard instructional practice? (2) Is short-term memory capacity and SES associated with children's progress with the maths apps?

Mathematical skill acquisition

First, there is a need to understand whether the maths apps are supportive of early mathematical skill acquisition in comparison to standard practice (Cheung & Slavin, 2013) when time learning mathematics is equated across the treatment and control group conditions (Foster et al., 2016; Ginsburg & Smith, 2016). Previous intervention research evaluating the same maths app intervention in the UK did not include an adequate control group (Studies 1-3 Outhwaite et al., 2017) and app-based mathematics interventions have been frequently implemented as a supplementary teaching aid (Berkowitz et al., 2015; Schacter & Jo, 2016).

Furthermore, it is important to understand who may benefit from the maths app intervention based on achievement levels (Connolly et al., 2018; Pawson & Tilley, 2004; Slavin, 2012), for example children identified by their teachers as low-achievers in mathematics. Previous studies have examined the benefits of other technology based mathematical intervention for low-achievers and showed positive benefits (Cozard & Riccomini, 2016; Schacter & Jo, 2016; Szkudlarek & Brannon, 2018), including reducing the achievement gap in mathematics between low and higher-achieving children (Burns, Kanive & Degrande, 2012; Tan & Tan, 2015).

Therefore, this study asked, do low-achieving children aged 4-5 years make more progress in mathematics with the maths app intervention (Group 1: time-equivalent treatment) compared to relatively typically attaining children from the same class receiving

standard mathematical practice (Group 2: control)? The time learning mathematics was equated across the two conditions.

Child-level influencing factors

Second, within a multi-level, ecological, determinant systems approach (see Figure 1, Chapter 1), there is a need to examine which child-level factors are associated with children's progress with the maths apps (Bronfenbrenner, 1979; Thomas, 1992). As outlined in Chapter 1, poor memory capacity (Alloway et al., 2014; LeFevre et al., 2005; Peng et al., 2018) and a low SES (Anders et al., 2012; Kalaycioğlu, 2015) have been shown to be associated with poor mathematical ability. Previous studies have shown mathematics interventions, including app-based mathematics instruction did not lead to improvements in short-term memory capacity (Messer et al., 2018; Park et al., 2016; Szkudlarek & Brannon, 2018). This is to be expected given the relative stability of memory capacity early in child development (Gathercole et al., 2004). However, the instructional features within the maths app intervention, for example, the clear, specified learning goals with simple, stepwise task instructions from the on-screen teacher, which children can repeat as required (see Chapter 3) may provide suitable scaffolded, learning strategies for children with a low memory capacity and thus increase access to learning (Gathercole & Alloway, 2007).

Furthermore, it is vital to understand who may benefit from the maths app intervention based on the child's SES background. Previous studies demonstrate a positive impact of app-based

mathematics instruction for children from low SES backgrounds (Park et al., 2016; Schacter & Jo, 2016). Therefore, this study also asked, is short-term memory capacity and SES associated with children's progress with the maths apps?

Methods

Design

A between-groups, quasi-experimental research design was adopted to evaluate the impact of the new maths app intervention for supporting early mathematical skill acquisition compared to standard mathematics instruction with children aged 4-5 years, in the first year of compulsory education in the UK known as the Early Years Foundation Stage 2 (hereafter referred to as EYFS).

As outlined in Table 10, a group of low-achieving children were allocated to receive the maths app intervention over a period of 16 weeks with 50% exposure time to the maths apps. This was equivalent to 8 weeks of full time exposure, consistent with Pitchford's (2015) study of the same maths app intervention in Malawi (see Chapter 2). In the remaining 50% exposure time, children received standard mathematics practice. Children's mathematical learning gains were compared to that of a control group of typically attaining children, identified as progressing well in mathematics by their teachers, within the same EYFS class receiving standard mathematics practice only. This ensured that overall exposure to mathematics instruction was equivalent for both

groups. This between-groups, quasi-experimental design with the inclusion of a control group that were matched in exposure to mathematics learning activities addressed the main limitation of Studies 1-3 reported in Outhwaite et al. (2017; see Chapter 2).

Children's mathematical abilities were assessed on a paper-and-pen assessment of Mathematical Curriculum Knowledge (see Appendix C) and the Mathematical Reasoning sub-test from the Wechsler Individual Achievement Test Second Edition (WIAT-II; Wechsler, 2005), immediately before (pre-test) and immediately after (post-test) the 16- week intervention period. Children's memory capacity was also assessed on the Number Recall and Word Recall sub-tests from the Kaufman Assessment Battery for Children Second Edition (KABC-II; Kaufman & Kaufman, 2004) at pre-test and post-test.

The School of Psychology Ethics Committee at the University of Nottingham granted ethical approval for the study. Opt-in informed parental consent was obtained for all participating children in line with the British Psychological Society ethical guidelines.

Table 10 Description of maths intervention received and total maths-learning exposure for each group across the 16-week intervention period.

Mathematics Activities	Group 1 (time-equivalent treatment)	Group 2 (control)
Maths app intervention (50% exposure)	✓	
Small group-based maths instructions (50% exposure)	✓	
Small group-based maths instructions (100% exposure)		✓
Whole class embedded maths activities	✓	✓
Total time learning maths	Typical	Typical

Participants

In total 27 children aged 4-5 years took part in the study. 12 children identified by their teachers as low-achievers in mathematics (Yudkin & Stratton, 1996) were assigned to Group 1 (time-equivalent treatment). These children received the maths app intervention instead of a daily small-group based mathematics activity that is given as part of standard mathematical practice for 50% of the time during the 16-week intervention period. The remaining 15 children were identified by their teachers as typically

attaining in mathematics and were assigned to Group 2 (control). These children received regular mathematics teaching only; they did not use the maths app intervention.

All 27 children were pre-tested on the two mathematics measures and the two memory measures. The 27 children were available at post-test and were given a different version of the paper-and-pen assessment of Mathematical Curriculum Knowledge (see Appendix C), the same Mathematical Reasoning sub-test from the WIAT-II (Wechsler, 2005) and the same Number Recall and Word Recall sub-tests from the KABC-II (Kaufman & Kaufman, 2004). Table 11 details descriptive data for the sample of 27 children.

School context

This study was conducted in one EYFS class in a larger than average sized, mixed gender primary school in a middle-class SES area of Nottingham, UK. The school was rated 'Good' in the latest Ofsted report (Ofsted, 2014). Ofsted (Office for Standards in Education, Children's Services and Skills) is responsible for inspecting and regulating the quality of educational institutions in the UK. The large majority of children were from White British backgrounds. Children eligible for pupil premium financial support (additional government funding for children from low SES backgrounds), children with special educational needs, and children with English as an additional language was below the national average (Ofsted, 2014).

Table 11 Descriptive data, including mean age in months (SD, Min-Max), Gender (Female: Male) and mean IDACI score (SD, Min-Max) for the study sample and each participating group.

Descriptive Data	Total (*n* = 27)	Group 1 (time-equivalent treatment; *n* = 12)	Group 2 (control; *n* = 15)
Age (months)	55.96 (4.00)	53.92 (3.70)	57.60 (3.54)
	50.00 – 61.00	50.00 – 60.00	50.00 – 61.00
Gender (F:M)	14: 13	7: 5	7: 8
SES	7.44 (7.09)	9.67 (9.72)	5.67 (3.42)
(IDACI score)	1 - 36	1 -36	1 -8

Maths app intervention

Children allocated to Group 1 (time-equivalent treatment) used the app 'Maths 3-5' developed by *onebillion,* an educational not-for-profit organisation. How the maths app maps onto the mathematical concepts covered in the EYFS Profile (Department for Education, 2013) and the Science of Learning principles is outlined in detail in Chapter 3.

As highlighted in Table 10 children in Group 1 (time-equivalent treatment) used the maths app intervention in 30-minute sessions over a 16-week intervention period with 50% exposure time (equivalent to 8 weeks of full time exposure). This intervention dosage was due to practical reasons associated with the accessibility of the touch-screen device hardware.

Small group mathematics instruction

Standard mathematical practice included a small group-based mathematics activity, delivered by the class teacher and focused on a particular mathematical concept from the EYFS Profile (Department for Education, 2013). Example activities were obtained through observations made at the start of the study. For example, in an activity delivered on the interactive whiteboard, children practiced counting a set number of teddy bears onto the bus. Children responded by counting aloud as a group and were supported by the teacher, who gave corrective feedback before moving onto the next item. As highlighted in Table 10, children in Group 1 (time-equivalent treatment) received this small group maths instruction for 50% of the 16-week intervention period. In comparison, children in Group 2 (control) received this standard mathematics practice for 100% of the 16-week intervention period.

Whole class embedded mathematics activities

In the UK children in the first year of school (age 4-5 years) are typically taught through play. In a whole class setting teachers embed mathematical concepts from the Early Years Foundation Stage Profile into play-based activities. Examples of whole class activities were obtained through observations at the participating school. For example, children were introduced to the concepts of pattern recognition and completion in a cut and paste activity situated on one of the options tables. During free play time children could choose to participate in this activity, where they were required to organise a collection of shapes and objects to complete the

predefined pattern. Children were supervised by the teaching staff to ensure child engagement and provide additional supportive scaffolding. These embedded, child-directed, play-based activities were consistent with guided play principles (Ash & Wells, 2006; Fisher et al., 2013; Weisberg et al., 2016; see Chapter 1). As shown in Table 10, all children participating in this study received this form of play-based embedded instruction.

Mathematics assessments

Children's mathematical ability was assessed using a paper-and-pen assessment of Mathematical Curriculum Knowledge (see Appendix C) and the Mathematical Reasoning sub-test from the WIAT-II (Wechsler, 2005).

The paper-and-pen assessment of Mathematical Curriculum Knowledge was a specifically designed and experimental measure. It included 50 novel items based on the mathematical content taught in the maths apps. Basic mathematical skills assessed in this measure included counting, understanding and using numbers, simple addition and subtraction, and shape, space and measure recognition. Items increased in difficulty in line with progress through the apps. No discontinuation rule was applied; all questions were administered. This Mathematical Curriculum Knowledge assessment has been used in previous studies evaluating the same maths app intervention (Studies 2 & 3, Outhwaite et al., 2017) and a tablet version of the assessment (version A only) has also been in

the UK (Study 1, Outhwaite et al., 2017) and Malawi (Pitchford, 2015).

To control for the possibility of practice effects over repeated administration of the assessment, two versions of the paper-and-pen Mathematical Curriculum Knowledge assessment were created (see Appendix C). In each version, questions were designed to assess the same basic skills but included different numbers and/or stimuli. As such, this paper-and-pen based assessment measured generalisation of Mathematical Curriculum Knowledge gained through the maths app intervention to new items and a different format. The researcher administered the assessment to children on an individual basis by reading aloud each of the questions in turn and then asking the child to respond. A raw score of number correct (maximum raw score = 50) was determined for each child. Reliability analysis of the paper-and-pen Mathematical Curriculum Knowledge assessment showed good internal consistency between pre-test and post-test scores, r = .47, Cronbach's alpha = .62.

The Mathematical Reasoning sub-test is a standardised and age-adjusted assessment of children's mathematical ability, suitable for children aged 4 years and above. It is designed to assess a range of children's basic and higher-order mathematical skills, including tasks such as, counting, identifying geometric shapes, problem solving, and graph interpretation. The researcher administered the sub-test individually to children in accordance with the test manual. Discontinuation rules are applied after six

consecutive fails. A raw score of number correct (maximum score = 67) was calculated for each child. Reliability analysis of the Mathematical Reasoning sub-test showed high internal consistency between pre-test and post-test scores, r = .60, Cronbach's alpha = .74.

To calculate an overall mathematics composite score, percentage correct on the Mathematical Curriculum Knowledge assessment (maximum score = 50) and the Mathematical Reasoning sub-test (maximum score = 67) were mean averaged.

Memory assessments

Span tasks are typically used to assess children's memory capacity (Conway et al., 2005). Complex span tasks require participants to maintain a collection of information over a brief period of time in addition to a demanding secondary cognitive task component (Unsworth & Engle, 2007a). For example, in a backwards corsi block task, children are required to recall a specific visuospatial pattern in the opposite order in which it was presented (Milner, 1971). These tasks typically assess working memory but are often considered too difficult for young children to perform reliably and may not be considered developmentally appropriate as working memory starts to develop from approximately 4 years of age (Gathercole et al., 2004).

In contrast, simple span tasks typically measure short-term memory and require participants to maintain a collection of information over a brief period of time (Unsworth & Engle, 2007b).

Research shows there is a high degree of overlap between short-term memory and working memory in young children (Aben, Stapert & Blokland, 2012; Wiebe, Espy & Charak, 2008).

As such, the simple span tasks, Number Recall and Word Recall sub-tests from the KABC-II (Kaufman & Kaufman, 2004) were chosen as age-appropriate assessments of children's short-term memory capacity. Both of these measures are standardised and appropriate for children aged three years and above.

In the Number Recall sub-test children were required to verbally repeat a series of one digit, one syllable numbers presented verbally by the researcher. Sequences of digits ranging from two to nine were administered with increasing difficulty. The test discontinued after three incorrect trials.

In the Word Recall sub-test children were required to touch a series of common object silhouettes in the same order as verbally presented by the researcher. Task difficulty increased progressively and a discontinuation rule was applied after three consecutive fails. For both tasks the number of correct trials completed was determined for each child and then averaged to give a composite memory score. Reliability analysis of the memory composite score showed high internal consistency between pre-test and post-test scores, r = .64 Cronbach's alpha = .77.

Socio-economic status

Definitions of socio-economic status in relation to research with children and young people can incorporate parental income,

education, and occupation (Hauser, 1994). As such, in this study a child's SES was indicated through their residential postcode using the Income Deprivation Affecting Children Index (IDACI). This measure reflects neighbourhood deprivation based on the proportion of local children aged 0-15 years living in low-income families receiving means tested benefits within a particular postcode area (McLennan et al., 2011). Postcode data was collected for each child through school records.

Procedure

Recruitment.

The participating school was recruited to take part in this study by the research team through the Apple Distinguished Educator Network. The researcher and first supervisor attended meetings with the EYFS teachers and the school Head Teacher to discuss the study design and procedures. The participating school already had access to the required hand-held tablet device hardware (iPads) and were given access to the maths apps free of charge, by *onebillion.* Children were recruited through the participating school via a parent information sheet and a letter from the school. Opt-in parental consent was obtained for all participating children before study commencement.

Support system.

The support system for intervention implementation (Domitrovich et al., 2008) comprised of pre-intervention teacher training and continued email communication with the researcher

throughout the intervention period. During pre-intervention teacher training, all EYFS teaching staff were trained in how to use the maths app intervention. The training focused on the knowledge and skills they needed to use the maths app, for example, how to set up individual children's in-app profiles, demonstrations on how children progress through the app and how to deal with common difficulties. The participating school also had continued email communication support from the researcher, in case of any further questions, queries or problems during the 16-week intervention period.

Group allocation.

Children's allocation to the time-equivalent treatment group (Group 1) or the control group (Group 2) was based on teacher's assessments of children's prior mathematical abilities. These assessments were conducted at the start of the academic year as standard EYFS practice and independent of the research. The assessment was based on observations of children's number, shape, space and measure abilities at the start of the academic school year as part of the EYFS profile (Department for Education, 2013). Children meeting the EYFS profile typical behaviour criteria for 30-50 months or below for mathematics were identified as low-achievers. Children meeting the EYFS profile typical behaviour criteria for 40 to 60+ months or above for mathematics were identified as typically attaining. Although group allocation remained fixed throughout the duration of the study, it was not practically possible for the researcher to remain blind to condition.

Assessment administration.

Before and after the 16-week intervention period, children were individually assessed on the two measures of mathematical ability (Curriculum Knowledge and Mathematical Reasoning) and the two measures of short-term memory (Number Recall and Word Recall). Assessments were administered in the same order in three short blocks (see Table 12). Each assessment block lasted approximately 10-20 minutes and was administered on a one-to-one basis by the researcher over consecutive school days in a quiet area, free from distraction, in the child's familiar school environment. Before each assessment, the researcher explained the task to the child, ensuring the child felt comfortable and explicitly asked if the child would like to take part in the assessment so as to gain verbal assent. Completed assessments were scored by the researcher.

Table 12 Assessment protocol

Assessment Block	Pre-test	Post-test
Block 1	Mathematical Reasoning	Mathematical Reasoning
Block 2	Number Recall	Number Recall
	Word Recall	Word Recall
Block 3	Mathematical Curriculum Knowledge (Version A)	Mathematical Curriculum Knowledge (Version B)

Maths app intervention implementation.

Children in Group 1 (time-equivalent treatment) used the maths app intervention in 30-minute sessions. Dues to practical reasons associated with the accessibility of the touch-screen tablet device hardware, each of the low-achieving children used the maths app for 50% of the time over the 16-week intervention period (equivalent to 8 weeks of full-time intervention implementation; Pitchford, 2015).

The maths app intervention was embedded into the daily classroom routine and was scheduled just before lunchtime. As outlined in Table 13, for the first 8 weeks of the intervention period, the first wave of six children from Group 1 (time-equivalent treatment) used the maths app under the supervision of the teaching assistant. At the same time, the remaining six children in Group 1 (time-equivalent treatment) and the 15 children in Group 2

(control) received the daily small group-based mathematics activity lead by the class teacher. During the last 8 weeks of the intervention period, the second wave of six children in Group 1 (time-equivalent treatment) used the maths app and the other six children from Group 1 (time-equivalent treatment) and all of the children in Group 2 (control) received standard teacher-led mathematical practice. As such, all children had the same amount of exposure time to mathematics instruction over the 16-week intervention. After the 16-week intervention period, the 15 typically attaining children used the maths apps. At the end of the study, the participating school was given continued access to the maths apps, so that the apps were available to all children.

All children that were allocated to use the maths app intervention (Group 1: time-equivalent treatment) started using the 'Maths 3-5' app on topic one and used the same Apple iPad each day, as they had their own profile in the app, which saved their progress. To support classroom organisation, the iPads were colour coded by the teaching staff. Children used the maths app independently within their own in-app profile and with headphones. Children were seated in a quiet area of their normal classroom, separate from the other mathematics activities occurring at the same time and were supervised by the teaching assistant. The main role of the teaching assistant in the implementation of the intervention was to provide technical support and behavioural management to the children when necessary.

Table 13 Description of intervention implementation over the 16-week intervention period.

Study Timeline	Group 1 (time-equivalent treatment) Wave 1 (*n* = 6)	Group 1 (time-equivalent treatment) Wave 2 (*n* = 6)	Group 2 (control)
Weeks 1-8	Maths apps	Normal maths practice	Normal maths practice
Weeks 9-16	Normal maths practice	Maths apps	Normal maths practice
After 16 weeks	Normal maths practice	Normal maths practice	Maths apps

Time on task.

In their allocated implementation wave, children in Group 1 (time-equivalent treatment) used the maths apps for 30 minutes a day, five days a week, for eight consecutive weeks. This equated to 1200 minutes (20 hours) exposure to the maths app intervention. As the maths apps were implemented instead of the small group-based mathematics activity and all children continued to receive whole class embedded mathematics activities (see Table 10), the total time learning mathematics was equated across Group 1 (time-equivalent treatment) and Group 2 (control).

Results

Mathematical skill acquisition

A change from baseline (pre-test) analysis was deemed appropriate due to the quasi-experimental research design and differences in pre-test mathematics assessment scores across the two groups (see Table 14; Thomas & Zumbo, 2012; Van Breukelen, 2006).

To examine whether the maths app intervention supported early mathematical skill acquisition for low-achieving children (Group 1: time-equivalent treatment) compared to the typically attaining control group who received standard mathematics practice only (Group 2: control), mean performance on the Mathematical Curriculum Knowledge and Mathematical Reasoning assessments for each of the two groups (see Table 14) were compared using two separate 2 (Time: pre-test, post-test) x 2 (Group: treatment, control) mixed ANOVAs.

Table 14 Raw score group mean (SD) at pre-test, post-test, and gain score (post- minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95% confidence interval) for each assessment of mathematical ability and composite short-term memory scores.

Outcome Variable	Group 1 (time-equivalent treatment; *n* = 12)	Group 2 (control; *n* = 15)
Mathematical Curriculum Knowledge (max. 50)		
Pre-test Mean (SD)	22.83 (3.76)	29.67 (3.58)
Post-test Mean (SD)	37.25 (4.81)	38.13 (3.27)
Gain Score Mean (SD)	14.42 (4.10)	8.47 (3.36)
% Gain	28.8	16.9
Cohen's *d* (95% CI)	3.34 (1.59 – 5.09)	2.47 (1.12 – 3.81)
Mathematical Reasoning (max. 67)		
Pre-test Mean (SD)	10.50 (1.93)	12.73 (3.10)
Post-test Mean (SD)	15.92 (1.93)	17.07 (2.49)
Gain Score Mean (SD)	5.42 (1.38)	4.33 (2.85)
% Gain	8.1	6.5
Cohen's *d* (95% CI)	2.81 (1.21 – 4.40)	1.54 (.39 – 2.70)
Mathematics Composite Score		
% Pre-test Mean (SD)	30.67 (4.38)	39.17 (5.35)
% Post-test Mean (SD)	49.13 (5.41)	50.87 (4.04)
% Gain Score Mean (SD)	18.46 (4.49)	11.70 (4.32)
Cohen's *d* (95% CI)	3.75 (1.87 – 5.63)	2.47 (1.13 – 3.81)

Outcome Variable	Group 1 (time-equivalent treatment; *n* = 12)	Group 2 (control; *n* = 15)
Short-Term Memory Composite		
Pre-test Mean (SD)	9.38 (2.49)	11.60 (1.38)
Post-test Mean (SD)	11.79 (1.97)	12.67 (1.78)
Gain score Mean (SD)	2.42 (2.11)	1.07 (1.19)
Cohen's *d* (95% CI)	1.07 (-.14 – 2.28)	0.67 (-.37 – 1.71)

For Mathematical Curriculum Knowledge results showed a significant interaction between Time and Group, F (1, 25) = 17.32, $p < .001$. Post-hoc paired samples t-tests revealed that both groups made significant progress in Mathematical Curriculum Knowledge over time: Group 1 (time-equivalent treatment), t (11) = 12.18, $p < .001$, Group 2 (control), t (14) = 9.83, $p < .001$. However, larger within-group effect sizes[3] (Cohen's *d* with 95% CI) reflecting the magnitude of progress were observed for Group 1 (time-equivalent treatment) compared to Group 2 (control; see Table 14).

Further analysis of simple main effects with independent t-tests showed that at pre-test, the typically attaining children in

[3] Within-group effect sizes were calculated by subtracting the post-test group mean from the pre-test mean and dividing the result by the pooled standard deviation. The calculation was conducted using this online calculator: https://www.psychometrica.de/effect_size.html. This approach was applied to all within-group effect size calculations reported in this thesis.

Group 2 (control) significantly outperformed their low-achieving peers in Group 1 (time-equivalent treatment), t (25) = 4.82, p < .001, between-group effect size[4] = 1.87, CI = .96 – 2.78. This is to be expected as children in Group 2 (control) were identified at pre-test as being of relatively higher mathematical ability than children identified as low-achievers in Group 1 (time-equivalent treatment). Importantly, at post-test no significant difference between groups was observed, t (25) = .57, p = .576, between-group effect size = .22, CI = -.54 – .98.

For Mathematical Reasoning results showed no significant Time x Group interaction, F (1, 25) = 1.46, p = .239.

Child-level influencing factors

Memory capacity.

To examine the relationship between mathematical learning gains and short-term memory capacity Pearson's correlational analyses were conducted for Group 1 (time-equivalent treatment) between difference scores for the mathematics composite measure and the raw memory composite scores taken at pre-test.

Results showed a small, negative but not statistically significant correlation relationship between children's pre-test short-

[4] Between-group effect sizes were calculated by subtracting the mean difference score for Group 1 from the mean difference score for Group 2 and dividing the result by the pooled standard deviation controlling for the sample sizes. The calculation was conducted using this online calculator: https://www.psychometrica.de/effect_size.html. This approach was applied to all between-group effect size calculations reported in this thesis.

term memory capacity and mathematical learning gains (difference scores on the mathematics composite measure) for Group 1 (time-equivalent treatment), r = -.20, p = .536.[5]

To examine whether short-term memory improved over the 16-week intervention period, a 2 (Time: pre-test, post-test) x 2 (Group: treatment, control) mixed ANOVA was conducted.

Results showed a significant Time x Group interaction, F (1, 25) = 4.41, p = .046. Post-hoc paired samples t-tests revealed that both groups made significant progress in short-term memory over time: Group 1 (time-equivalent treatment), t (11) = 3.97, p = .002, Group 2 (control), t (14) = 3.46, p = .004. However, larger within-group effect sizes (Cohen's d with 95% CI) were observed for Group 1 (time-equivalent treatment) compared to Group 2 (control; see Table 14).

Further simple main effects analysis with independent t-tests showed at pre-test, the typically attaining children in Group 2 (control) significantly outperformed their low-achieving peers in Group 1 (time-equivalent treatment), t (25) = 2.95, p = .007, between-group effect size = 1.14, CI = .32 – 1.96. In contrast, at post-test no significant difference between groups was observed, t (25) = .15, p = .703, between-group effect size = .47, CI = -.30 – 1.24.

[5] Non-parametric Spearman's Rho correlation analyses were also conducted and showed the same pattern of results, r^s = .24, p = .450.

SES.

To examine the impact of SES on mathematical learning gains observed across the 16- week intervention period Pearson's correlational analyses were conducted for each group between IDACI scores (see Table 11) and the difference scores (post-test minus pre-test) for the mathematics composite measure (see Table 14).

Results showed small, negative and not statistically significant correlations between children's SES and difference scores on the mathematics composite measure for Group 1 (time-equivalent treatment; r = -.13, p = .679) or Group 2 (control; r = -.19, p = .493)[6].

Discussion

This chapter examined the efficacy of the new maths app intervention for children aged 4-5 years old identified as low-achievers and the role of short-term memory and SES in learning with the apps. Through the inclusion of a control group that were matched in exposure to mathematics activities, the current study addresses the main limitation of the previous similar research (Studies 1-3, Outhwaite et al., 2017).

[6] Non-parametric Spearman's Rho correlation analyses were also conducted and showed the same pattern of results: Group 1 r^s = .17, p = .586; Group 2 r^s = .18, p = .520.

Mathematical skill acquisition

Consistent with the previous studies (Outhwaite et al., 2017), this study showed that low-achieving children who used the maths app intervention made significantly stronger learning gains in mathematics compared to the control group who received standard practice. The progress in Curriculum Knowledge made by the low-achievers over the 16-week intervention periods was significantly greater (Cohen's *d* = 3.3) than that made by their relatively higher-achieving, typically attaining peers in the control group (Cohen's *d* = 2.5) and was characterised by a very large between-group effect size (Cohen's *d* = 1.61, 95% CI = .74 – 2.48).

Furthermore, the significant gap in achievement that was evident at pre-test closed following the 16-week intervention period. At pre-test the Curriculum Knowledge performance of the typically attaining control group (Group 2) was 13.7% higher than that of the low-achieving time-equivalent treatment group (Group 1). In contrast, at post-test the performance levels were similar across groups. At post-test, the mean performance of the typically attaining control group (Group 2) was only 1.8% higher than that of the low-achievers (Group 1).

For the standardised measure of mathematical concepts, although the Time x Group interaction was not significant, effect sizes calculated per group across the intervention period were greater for the low-achieving children that used the maths app intervention (Cohen's *d* = 2.8) compared to the typically attaining

controls (Cohen's *d* = 1.5). Comparisons of difference scores, reflective of progress showed a small to medium between-group effect size (Cohen's *d* = .47, 95% CI = -.30 – 1.24).

Consistent with previous studies, this evidence collectively, suggests app-based mathematics instruction is effective for supporting the acquisition of early mathematical skills (Outhwaite et al., 2017; Park et al., 2016; Pitchford, 2015; Schacter & Jo, 2016; 2017; Schacter et al., 2015; Szkudlarek & Brannon, 2018; Van der Ven et al., 2017) and can reduce the achievement gap between low and higher-achieving children (Burns et al., 2010; Tan & Tan, 2015).

Child-level factors

Within a determinant systems approach, this study also considered the influence of memory capacity and SES on children's learning gains with the maths apps.

Memory capacity.

A negative relationship was found between short-term memory composite scores at pre-test and mathematical learning gains (difference score) for children low-achieving in mathematics. This suggests low-achievers with a weaker memory capacity that used the maths apps made more progress in mathematics over the duration of the 16-week intervention period than those with a relatively higher memory capacity. This finding is consistent with similar previous research (Study 3, Outhwaite et al., 2017) that found a moderate, negative, significant relationship between maths

app learning gains and pre-test short-term memory capacity for low-achieving children aged 5-7 years (r = -.38, p = .048). However, in the current study this relationship did not reach significance (r = -.20, p = .536). The difference in strength and significance of this relationship across studies could be due to the relative lack of power in the current study compared to Study 4, as sample sizes differed across studies (Current low-achievers n = 12; Study 3 low-achievers n = 27), as did the duration of the intervention (Current study 16 weeks of intervention at 50% exposure totalling 1,200 minutes/20 hours app use; Study 3 13 weeks of intervention at 100% exposure totalling 1,950 minutes/32.5 hours app use). Nevertheless, the negative relationship found across both studies suggests that the maths app intervention is particularly suited for use with low-achieving children with poor memory capacity.

Interestingly, this study also showed that children's short-term memory capacity significantly increased over the 16-week intervention period for both the treatment and control groups. In particular, children using the maths apps showed stronger memory gains (Cohen's d = 1.07) than the control group (Cohen's d = 0.67). Furthermore, pre-test group differences in memory capacity were no longer evident at post-test.

Initially, this evidence may suggest that there are secondary, cognitive benefits of the new maths app intervention. However, this finding is inconsistent with previous research, which showed no

improvement in memory capacity in response to both app-based and non-technology based mathematical training (Messer et al., 2018; Park et al., 2016; Szkudlarek & Brannon, 2018). Previous research also suggests memory capacity is relative stable in early in child development (Gathercole et al., 2004), meaning that an observed improvement in memory capacity is unlikely.

In the current study the short-term memory findings follow the same pattern of results as those seen for the mathematical, Curriculum Knowledge outcome variable. Given the close relationship between memory and mathematics (LeFevre et al., 2005), particularly for low-achieving children (Cragg et al., 2017; Gathercole & Alloway, 2006; Raghubar et al., 2010), the current memory findings could be a proxy for and thus reflective of mathematical learning gains.

Alternatively, the current observed gains in short-term memory could be attributed to task familiarity, as children completed the same simple span task before and after the intervention period. Without further investigation utilising different versions of the simple span tasks, which would address practice effects, it is difficult to separate and address these three contrasting possible explanations.

SES.

Consistent with similar previous research examining the same maths app intervention (Outhwaite et al., 2017), SES was shown to have no significant association with children's progress in

mathematics over the duration of the intervention period for either group.

Limitations

The current study builds on the previous studies evaluating the same maths app intervention with UK children (Studies 1-3, Outhwaite et al., 2017) and adds to the growing evidence base examining app-based mathematics instruction in early education (see Chapter 2). While the inclusion of a control group in this between-group, quasi-experimental research design addresses potential threats to internal validity associated with natural maturation (Shadish et al., 2002) there are two important limitations to consider when interpreting the current findings and directing future research.

First, the two groups were not equal in terms of mathematical ability at pre-test. Specifically, children allocated to receive the maths app intervention (Group 1: time-equivalent treatment) had significantly lower abilities in mathematics relative to their typically attaining peers allocated to receive standard practice (Group 2: control). Threats to internal validity associated with regression to the mean were addressed through teacher assessment based group allocation (Yudkin & Stratton, 1996) and this design enables the examining of interesting research questions associated with who the maths app intervention may work best for, as demonstrated in this study.

However, critiques have argued that studies with large pre-test differences still pose significant threats to validity, even if appropriate statistical procedures are used as the underlying distributions may be fundamental different across the treatment and control groups (Cheung & Slavin, 2013; Shadish et al., 2002). To address this issue, future research needs to implement the maths apps at the whole class level, whereby children may be allocated to receive the intervention regardless of their pre-test ability level. This approach will be taken in Chapter 5 (Brazil Proof of Concept) and 6 (UK Feasibility RCT).

Second, while contributing to the initial cumulative evidence for the new maths app intervention (Outhwaite et al., 2017) the current study is relatively small-scale and the findings are relatively restricted to one school, located in a middle-class area of Nottingham. In order to scale the maths app intervention and further understand how the apps work, who they work for, and under what circumstances (see Chapter 1), further larger scale research is needed, including different population groups, a greater number of schools and children, and different forms of intervention implementation.

Future research needs to work with different population groups, such as bilingual children. This will help elucidate the association between proficiency in the language of instruction and learning with the maths apps (see Chapter 5, Brazil Proof of Concept). Through working with a greater number of schools and

pupils, studies will help establish the efficacy of this intervention when implemented at scale. How the maths app intervention is most effectively implemented at scale also needs to be investigated, for example whether the intervention is more beneficial as a supplementary aid to mathematics classroom teaching (Studies 1-3, Outhwaite et al., 2017) or whether it should be fully embedded as part of the standard mathematics curriculum, as it was in the current study (see Chapter 6, UK Feasibility RCT).

Conclusion

In summary, Chapter 4 reported a small-scale, proof of concept study conducted in one UK primary school that showed children aged 4-5 years identified by their teachers as low-achieving in mathematics, made more progress in mathematics with the maths app intervention in comparison to their relatively typically attaining peers who received standard mathematical practice. The time learning mathematics was equated across both group, only the method of delivery; app-based mathematics instruction and standard teacher-led practice, was manipulated. No significant effect of children's socio-economic status was found but children with a poor memory capacity demonstrated stronger learning gains. This study provides a useful foundation for further research evaluating the new maths app intervention, prior to teaching practice and educational policy recommendations.

Chapter 5: Brazil Proof of Concept Study

Current study

As highlighted in Chapter 1, language proficiency plays a significant role in the development and learning of mathematics (Morgan, 2013). In order to further build the initial evidence of the new maths app intervention for supporting early mathematical development and to understand some of the mechanisms underpinning the maths apps, additional research examining language factors is required. Furthermore, in order to further explore for whom the maths apps work best, studies need to be conducted with different population groups. This study focuses on bilingual learners.

Consequently, this chapter addresses two important questions: (1) is app-based mathematics instruction more effective at supporting the acquisition of basic mathematical skills when delivered in the child's first (L1) or second language (L2), in comparison to standard classroom practice only in a dual immersion, bilingual educational context? (2) Does proficiency of the language of instruction influence mathematical learning gains and progress within the maths apps?

Mathematical skill acquisition

First, there is a need to evaluate whether the maths apps are effective for supporting early mathematical development compared to standard practice. Current experimental research evaluating app-

based mathematical interventions for bilingual children is sparse (see Chapter 1). To date, only one small-scale, quasi-experimental study has been conducted, which evaluated the same maths app intervention as this thesis (Study 1, Outhwaite et al., 2017). Results showed children with English as an Additional Language (EAL) made just as much progress in basic mathematical skills with the maths apps delivered in English, their assumed L2, compared to their non-EAL peers.

However, as the *onebillion* maths apps are available in 52 different languages (www.onebillion.org), this presents a unique opportunity to examine the impact of app-based mathematics instruction when delivered in the child's first or second language. As such, this study asked, are the maths apps more effective at supporting the acquisition of basic mathematical skills when delivered in the child's first (L1, Group 1: Brazilian Portuguese) or second language (L2, Group 2: English), in comparison to standard classroom practice only (Group 3: control) in a dual immersion, bilingual educational context?

Language of instruction proficiency

Second, there is a need to examine the relationship between proficiency in the language of instruction and learning with the maths apps. Research shows a child's cognitive academic language proficiency (CALP) needs to be sufficiently developed to access curriculum content and cope with instructional demands (Cummins, 2008; see Chapter 1). The design of high-quality app-based

mathematics instruction may provide additional contextual and cognitive support, which may enable young bilingual children to derive meaning about the learning activity (see Chapter 3).

As such, this study asked, does proficiency of the language of instruction influence mathematical learning gains and progress within the maths apps? This will help elucidate theoretical understanding of how app-based mathematics instruction works for young bilingual children, as well as, illuminate important practical considerations for implementing the maths app intervention beyond the research context.

Methods

Design

A quasi-experimental study was conducted to evaluate the efficacy of the maths app intervention delivered in the child's first (L1, Brazilian Portuguese) or second language (L2, English)[7] compared to standard mathematics practice (delivered in L2) for supporting the acquisition of basic mathematical skills with children aged 5-6 years, attending Kindergarten in a bilingual private school in Recife, in the North East of Brazil. A non-randomized, between-groups, design was adopted with three predetermined groups based on the three Kindergarten classes at the school.

[7] The typical language of mathematics instruction in this study was English, which for all participating children was their second language (L2).

To examine if the maths app intervention was more effective when delivered in the child's L1 (Brazilian Portuguese) compared to the child's L2 (English) the maths apps were implemented across two classes; one class received the app intervention in Brazilian Portuguese (treatment group 1) and one class received the app intervention in English (treatment group 2). Teaching staff implemented the maths apps in 20-minute sessions, 4 times a week, across 10 consecutive weeks. As the maths app intervention was implemented by teachers it optimised ecological validity. The maths app sessions replaced a small group, embedded mathematics activities, used in standard practice. The third class continued to receive mathematics teaching practice as usual, which for this school was delivered in English (L2) and so formed a control group to enable comparison of the impact of the maths apps intervention to standard practice and maturation. Group allocation to condition was determined by senior teaching staff at the school and was based on the school timetable and tablet device availability.

All children were assessed on mathematical skills, before (pre-test) and immediately after (post-test) the 10-week intervention period. Language proficiency in L1 and L2 for each child was assessed with a teacher-rated questionnaire at pre-test. Ethical approval for this study was granted by the Federal University of Pernambuco, the regional ethics committee, Comitês de Ética em Pesquisa (CEP) and the national ethics committee, Comissão Nacional de Ética em Pesquisa (CONEP). Opt-in parental consent was gained for all participating children. Parental consent to take

part in the study was given by 92% of all available children across the three classes.

Participants

In total, 62 children, aged 5-6 years, across the three Kindergarten classes took part in the study: 23 children in Class 1 were allocated to receive the maths app intervention in Brazilian Portuguese (treatment group 1), 20 children in Class 2 were allocated to receive the maths app intervention in English (treatment group 2), and 19 children in Class 3 received standard mathematics teaching practice only (control group). All children completed the pre-test (before) and post-test (after) assessments.

Exclusion criteria.

In order to enable children to show learning gains in the region of previous research (Outhwaite et al., 2017; Outhwaite, Faulder, Gulliford, & Pitchford, 2018) any child with a pre-test score on the primary outcome measure that was equal to or above 75% was excluded from the analysis. Similarly, time on task has been shown to impact learning gains (Outhwaite et al., 2017; Chapter 2). As such, any child that missed equal to or above 20% of the intervention sessions was also excluded from the analysis. According to these criteria one child was excluded from the analysis as they achieved 83% on the primary outcome measure at pre-test and was absent for 30% of the intervention sessions due to family travel. Descriptive data for the final sample structure (n = 61) is summarized in Table 15.

Table 15 Descriptive data, including mean age in months (SD, Min-Max), Gender (Female: Male) and mean L1 and L2 proficiency (SD, Min-Max) for the final study sample and each participating group.

Descriptive data	Total (*n* = 61)	Group 1 (Brazilian Portuguese; *n* = 23)	Group 2 (English; *n* = 19)	Group 3 (control; *n* = 19)
Age (months)	67.79 (3.49)	67.74 (3.54)	68.89 (3.14)	66.74 (3.59)
	62.00 – 73.00	62.00 – 73.00	64.00 – 73.00	62.00 – 71.00
Gender (F:M)	27 : 34	10 : 13	8 : 11	9 : 10
L1 proficiency	16.07 (3.83)	15.65 (3.07)	17.16 (3.83)	15.47 (4.57)
	8.00 – 21.00	10.00 – 21.00	8.00 – 21.00	8.00 – 21.00
L2 proficiency	13.31 (5.29)	13.96 (3.70)	15.53 (2.50)	10.32 (7.42)
	.00 – 21.00	8.00 – 21.00	12.00 – 19.00	.00 – 21.00

School context

The study was conducted in a Brazilian Portuguese-English bilingual private school in Recife, in the North East of Brazil. The school followed a one-way, dual language programme with 50/50 immersion (see Gomez et al., 2005). Separate class teachers delivered L1 and L2 instruction and time allocated to each language was split equally. The first half of the school day focused on all subjects, including mathematics, taught in the child's L2 (English). The second half of the school day focused on language skills, including reading and writing, in the child's L1 (Brazilian Portuguese). Table 16 illustrates the standard daily school routine (see Group 3) and how the apps intervention was embedded into this routine (see Group 1 & 2).

Table 16 Comparison of the school daily routines, including total weekly time learning mathematics for the two treatment groups and control group.

School Day	Group 1& 2 (maths app intervention)	Group 3(control)
First half	L2 instruction with English speaking Brazilian teacher: • All subjects • Maths app intervention delivered in L1 or L2 (4 x 20 mins per week)	L2 instruction with English speaking Brazilian teacher: • All subjects • Embedded mathematics activities delivered in L2 (2-3 x 20 mins per week)
Second half	L1 instruction with native Brazilian Portuguese teacher: • Brazilian Portuguese Language including reading and writing.	L1 instruction with native Brazilian Portuguese teacher: • Brazilian Portuguese Language including reading and writing.
Time on task		
Per week	~80 minutes	~40-60 minutes
Across study	~ 13 hours	~6-10 hours

Maths app intervention

Children allocated to Group 1 (Brazilian Portuguese) and Group 2 (English), used the two maths apps, 'Maths 3-5' and 'Maths 4-6' developed by *onebillion.* Table 17 detailed the topics covered in each app and how the app content maps onto the mathematical concepts covered in the Kindergarten (age 5-6 years) curriculum followed by the participating bilingual school.

As detailed in Chapter 3, the apps map strongly onto the principles of active, engaged, meaningful, and socially interactive learning (Hirsh-Pasek et al., 2015) and, as noted, provide a good degree of contextual support and graduated cognitive demand (Cummins, 1981). In order to complete a maths app topic, children need to achieve 100% pass rate on an end of topic quiz included in the software. The quizzes are designed to assess children's knowledge of the mathematical concepts covered in the topic activities (see Table 17). The number of quizzes was recorded as an indication of children's progress throughout the maths apps (Pitchford et al., 2018).

Table 17 Maths apps content mapped to the mathematics curriculum and attainment measure.

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
Maths 3-5		
Sorting and Matching	Data handling • Match pairs of identical or related objects • Select criteria for sorting • Sort and label real objects into sets by attributes (colour, type, size, shape)	No
Counting to 3	Number • Read, write and model numbers to 3 • Count to determine the number of objects in a set (one-to-one correspondence)	✓
Lines and Patterns	Pattern & Function • Recognise and describe simple patterns (sounds, actions, objects, nature) • Create simple patterns using real objects and drawings	✓
Counting	Number	✓

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
4 to 6	• Read, write and model numbers to 6 • Count to determine the number of objects in a set (one-to-one correspondence) • Connect the number names and numerals up to 6 to the quantities they represent	
Where is it?	Shape & Space • Use everyday language to describe position and direction (left, right, behind, next to)	No
Counting 7 to 10	Number • Read, write and model numbers to 10 • Count to determine the number of objects in a set (one-to-one correspondence) • Connect the number names and numerals up to 10 to the quantities they represent	✓
Patterns and Shape	Shape & Space • Sort, describe and name familiar 2D shapes (circle, square, rectangle,	No

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
	triangle, diamond, heart, oval) • Create patterns using familiar 2D shapes	
Numbers 1 to 10	Number • Count to 10 forwards and backwards by naming numbers in sequences • Order a set of selected numbers • Explore the conservation of number through the use of manipulatives • Subitise small collections of objects (up to 5 objects without counting) • Compare two numbers using more and less	✓
Comparing	Measurement • Introduce the language of length (long/short), mass (heavy/light) and capacity (full/empty, more/less) • Identify, compare and describe attributes of real objects, e.g. longer/shorter, heavier/ lighter, full/empty • Put sets of real objects in order of size	No

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
	and length	
Adding and Taking Away	Number • Introduce vocabulary involved in adding one more, add, and) • Introduce vocabulary involved in subtracting (one less, take away) • Represent practical situations to model addition and subtraction within 10 on concrete materials or pictures	✓
Maths 4-6		
Shape and Position	Not covered in Kindergarten curriculum	No
Counting to 20	Number • Read and model numbers to 20 • Count to determine the number of objects in a set (up to 20 objects; one-to-one correspondence)	✓
Sharing	Number • Explore the conservation of number through the use of manipulatives	No
More	Number	✓

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
Counting	• Use ordinal numbers to 10 in real life situations	
Telling the Time	Measurement • Begin to read o´clock time • Identify, describe and sequence events	No
Add and Subtract	Number • Represent practical situations to model addition and subtraction within 10 on concrete materials or pictures	✓
Count in tens and fives	Not covered in Kindergarten curriculum	No
How tall, how long?	Not covered in Kindergarten curriculum	No
Count to 100	Not covered in Kindergarten curriculum	No
2-D Shapes	Not covered in Kindergarten curriculum	No
Number	Number	No

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
lines	• Compare two numbers using more and less	
Fractions	Not covered in Kindergarten curriculum	No
*Weight it!	Measurement • Introduce the language of mass (heavy/light) • Identify, compare and describe attributes of real objects, e.g. heavier/ lighter • Put sets of real objects in order of size	No
*More Number Work	Number • Compare two numbers using more and less • Order a set of selected numbers	No
*3D Shapes	Not covered in Kindergarten curriculum	No
*Measure Time	Measurement • Begin to read o´clock time • Identify, describe and sequence events	No
*More +	Not covered in Kindergarten curriculum	No

App Topic	School Kindergarten Curriculum (Age 5-6 years)	Assessed in EGMA
and -		
*How much can it hold?	Measurement • Introduce the language of capacity (full/empty, more/less) • Identify, compare and describe attributes of real objects, e.g. full/empty	No

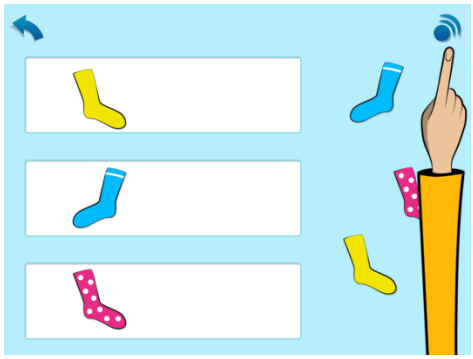
Note. *Last 6 topics available in English only.

As highlighted in Table 17, there are 28 topics in total across the two maths apps. The first 22 topics are available in Brazilian Portuguese; the last 6 topics are available only in English. The difference in the number of available topics could not be controlled or restricted. However, none of content from the last 6 topics was included in the primary outcome variable; mathematics ability as measured by EGMA. As the apps were self-paced children could also choose the order in which they worked through the app topics, but they were encouraged to complete the app topics systematically as presented in the apps. To address this, the children's progress through the apps and access to the different topics was recorded. No child in either app group completed the full 22 topics available in both languages over the 10-week intervention period (see Table 18). Two children in treatment group 2 (English maths app) completed one of the six topics that was available only in English.

In addition, the adequacy of the translation between the app content in Brazilian Portuguese and English was evaluated by a native Brazilian Portuguese speaker and a native English speaker. For each of the 22 topics (available in Brazilian Portuguese and English), one activity was randomly selected from the 7 activities available in each topic in 'Maths 3-5' and 6 activities available in each topic in 'Maths 4-6'. The Brazilian Portuguese and English transcripts of the first trial for each of the randomly selected activities were compared, such as the one illustrated in Figure 4.

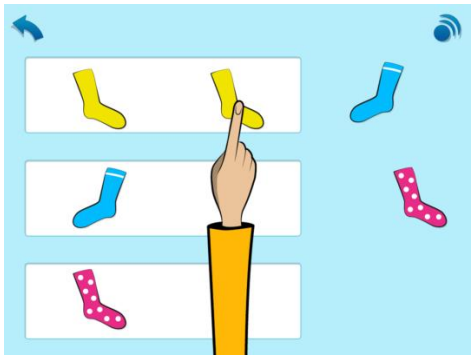
All of the selected app trials were judged to accurately convey the same meaning in Brazilian Portuguese and English, except for one, where the opposite meaning was conveyed. In this particular trial, the visual information positioned the numbers '10, 20, 30, 40' in ascending order and the English verbal tasks instructions asked children to continue placing the numbers in the 'right order.' However, the Brazilian Portuguese verbal task instructions asked children to continue placing the numbers in 'descending order.' Overall, there was high confidence in the translation adequacy.

Figure 4 Example item and task instructions for Topic 1, Sorting and Matching with Brazilian Portuguese and English verbal instructions.



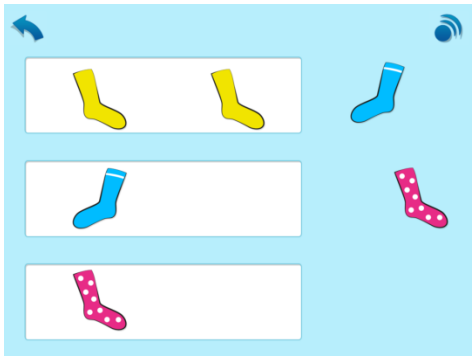
"Remember you can touch this to hear me again"

"Lembre-se, você pode tocar aqui para me escutar mais uma vez"



"Now look, yellow sock, yellow sock"

"Agora olhe, meia amarela, meia amarela"



"Your turn, drag the other socks into the boxes to match them"

"Agora é sua vez, arraste as outras meias para dentro das caixas para combinar com as que já estão lá"

Repeatable instruction:

"Drag the other socks into the boxes to match them"

"Arraste as outras meias para dentro das caixas para combinar com as que já estão lá"

Standard mathematics practice

The school curriculum followed an enquiry based pedagogical approach organised around interdisciplinary themes. Standard mathematics instruction in Kindergarten was delivered in the child's L2 (English). Mathematical concepts were integrated with other subject areas and learning activities were largely developed in whole class or small groups. For example, in an observed Kindergarten lesson from the control group (Class 3), the data-handling concept of sorting and matching was taught in a science-based activity requiring children to categorize living and non-living things. The task was demonstrated by the teacher to the whole class and then children worked in small groups to cut out printed images and then sort them into the relevant categories to create a poster to feedback to their teachers and peers. These embedded mathematics activities were typically implemented 2-3 times a week, for approximately 20 minutes a session.

Mathematics assessment

The primary outcome variable in this study was children's mathematical ability, which was assessed using the Early Grade Maths Assessment (EGMA; Brombacher, 2011). EGMA is a paper-based, age-appropriate, and internationally developed measure of mathematics attainment. It assesses core, basic, mathematical skills including number recognition, quantity discrimination, pattern completion, and numerical operations (addition and subtraction). The majority of EGMA items required a non-verbal response from the child (e.g. pointing) but for items that required a verbal

response (e.g. stating a number answer) the child could answer in Brazilian Portuguese or English, as they chose. This assessment is independent from the maths app intervention and the school curriculum so is not biased in favour of the treatment groups or the control group. Table 17 identifies which areas of the maths apps and school curriculum content are covered in EGMA. A raw score of number correct (maximum raw score = 54) was determined for each child. Reliability analysis of EGMA showed high internal consistency between pre-test and post-test scores, r = .81, Cronbach's alpha = .89.

Progression through the maths apps

The secondary outcome variable in this study was children's progression through the maths apps, which was operationalised by the number of topics completed (maximum 22 topics across the two apps for Brazilian Portuguese and 28 topics across the two apps for English). For a child to complete an app topic they had to achieve 100% correct on the end of topic quiz. Pitchford et al. (2018) showed that topics passed within the maths apps correlated significantly with learning gains measured by EGMA, demonstrating that progression through the apps is a valid indicator of mathematics learning. Monitoring data (topics passed) stored within the apps was extracted by the researcher and was available for 22 children in treatment group 1 (Brazilian Portuguese) and 18 children in treatment group 2 (English). Missing data was due to hardware corruption on one device, which was shared across the two

treatment groups and so impacted one child from each treatment group.

Language proficiency

Children's language proficiency in Brazilian Portuguese and English was assessed using a 7-item questionnaire completed by teaching staff (see Appendix D). The child's Brazilian Portuguese teacher completed the Brazilian Portuguese section of the questionnaire and the same procedure was followed by the child's English teacher (native Brazilian Portuguese speaker and speaker of English) for the English section of the questionnaire. The questionnaire was designed to measure children's abilities in speaking (4 items), reading (1 item), writing (1 item) and listening (1 item) in each language. The questionnaire was developed specifically for this study and was based on the Alberta Language and Development Questionnaire, a non-language and non-culturally specific parental questionnaire of children's language competencies in their first and second language (Paradis, Emmerzael & Sorenson Duncan, 2010). Items were scored between 0 and 3. The total score for each language (maximum score for each language = 21) was used an indication of the child's competencies in their first language, Brazilian Portuguese, and their second language, English.

Procedure

Study invitation and consent procedures.

The participating school was approached to take part in the study by our Brazilian collaborator, Maria Neves, due to the school's

location in Brazil and pre-existing access to the required touch-screen tablet devices (iPads). The school was given access to the maths apps, free of charge, by *onebillion.* Parents of Kindergarten children were sent information sheets and an information evening was held at the school by senior teaching staff and the researcher to invite and fully inform parents about the study. Opt-in parental consent was obtained for all participating children from a 92% response rate of all available children.

Support system.

Similar to the study reported in Chapter 2 (UK Proof of Concept study) the support system for intervention implementation (Domitrovich et al., 2008) consisted of pre-intervention teacher training and continued email and phone communication with the researcher throughout the intervention period. To further support intervention implementation during the initial stages of the study, the researcher was also available to provide technical support for the teachers during the first two weeks of the 10-week intervention period.

The pre-intervention teacher training for this study was delivered prior to study commencement. During this session, senior teaching staff and Kindergarten teachers discussed with the research team the practicalities of the proposed research design within their daily school timetable and tablet device availability. All teaching staff were made aware of the study procedures, including pre- and post-test assessment requirements, and were trained on

how to implement the maths app intervention by the researcher and Marc Faulder, a Kindergarten Specialist teacher from the UK who has experience of using the maths apps in his Kindergarten class (children aged 4-5 years; see Chapter 2).

Group allocation.

Children's allocation to the treatment groups (Group 1 and 2) or the control group (Group 3) was determined by senior teaching staff at the school and was based on the school timetable and tablet device availability. Although group allocation remained fixed throughout the duration of the study, it was not practically possible for the researcher to remain blind to condition.

Assessment administration.

Children's mathematics abilities were assessed using EGMA immediately before (pre-test) and immediately after (post-test) the 10-week intervention period. Language assessments were conducted at pre-test only. All assessment instructions were delivered in the child's first language, Brazilian Portuguese using audio recordings playing a standardised script. For items that required a non-verbal response, children responded by pointing. For that tasks that required a verbal response, children were allowed to respond in Brazilian Portuguese or English, as they chose. Assessments typically lasted approximately 15-20 minutes per child and were administered by the researcher on a one-to-one basis in a quiet area, free from distraction, in the child's familiar school environment. Before each assessment, the researcher explained the

task to the child, ensuring the child felt comfortable and explicitly asked if the child would like to take part in the assessment so as to gain verbal assent. Completed assessments were scored by the researcher.

Maths app intervention implementation.

Children allocated to receive the maths app intervention in Class 1 (L1 treatment group, Brazilian Portuguese) and Class 2 (L2 treatment group, English) used the maths apps in sessions lasting approximately 20 minutes, four times a week, for 10 consecutive weeks. Sessions were scheduled just before snack time and were implemented instead of the small group, embedded mathematics activities, used in standard practice.

Children used the same iPad each session, accessing their individual profile within the apps, which saved their progress. In order to aid classroom organisation, iPads were labelled with the children's name. Prior to the session, the teaching assistant set up the children's profile within the app, ensuring the correct language was selected. All children started using the apps in their allocated language on topic one in 'Maths 3-5.' Once children had completed all 10 topics, they progressed to 'Maths 4-6.' Teachers kept a register of which children had completed the 'Maths 3-5' app to track progress and support daily organisation. The maths app intervention was implemented in the children's normal classroom and children were sat in their regular seating plan of 5-7 children at one table. Children worked through the mathematics content within

their individual profile independently with headphones. Teaching staff supervising the sessions provided technical support and behaviour management when needed.

Time on task.

While the maths apps were implemented instead of the embedded mathematics activities typical of standard practice (see Table 16) the amount of time on task was not equivalent across the treatment and control groups. Specifically, standard mathematics practice was typically implemented in the control group 2-3 times a week. This equated to approximately 6-10 hours of mathematics instruction over the 10-week intervention period. In contrast, children in the two treatment groups used the maths apps 4 times a week, equalling approximately 13 hours of mathematics instruction. Therefore, children in the two treatment groups received a greater exposure to mathematics instruction over the 10-week intervention period compared to the control group (see Table 16) but time on task was equated across the two treatment groups.

Results

As noted above, one child was excluded from the data analysis due to absence for 30% of the intervention sessions and near ceiling performance (83% correct) on EGMA at pre-test. The final sample consisted of 61 children (see Table 18).

Bilingual status of sample

A paired samples t-test was conducted to compare L1 and L2 proficiency across the whole sample (see Table 18). Results showed children's L1 proficiency was significantly higher than their L2 proficiency, t (60) = 5.76, p <.001, identifying the sample as unbalanced bilinguals (Baker, 2011).

Primary outcome variable: mathematics gains

Similar to Chapter 2, a change score based analysis was deemed appropriate due to the quasi-experimental research design and differences in pre-test EGMA scores (see Table 18) across the three predetermined groups (Van Breukelen, 2006; Thomas & Zumbo, 2012). Table 18 shows mean EGMA performance at pre-test and post-test for each of the three groups. Group mean gain scores (post-test minus pre-test), percentage gains, and within-group effect sizes with 95% confidence intervals (CI) reflecting the progress of each group are also reported, as well as mean number of topics passed for each of the two treatment groups.

To examine if the maths app intervention was an effective teaching tool for supporting the acquisition of basic mathematical skills when delivered in the child's first language, Brazilian Portuguese (treatment group 1) or in the child's second language, English (treatment group 2) compared to standard mathematics practice (control group), the mean gain in EGMA score (post-test EGMA minus pre-test EGMA; see Table 18) for each of the three groups was compared using a one-way ANOVA.

Results revealed significant group differences in EGMA gain scores, $F(2, 58) = 3.78$, $p = .029$. Post hoc independent samples t-tests showed that children using the maths apps in Brazilian Portuguese (treatment group 1) made significantly greater gains than the control group, t (40) = 2.55, $p = .015$, between-group effect size $d = .79$, 95% CI = .16 – 1.42. Similarly, children using the maths apps in English (treatment group 2) made significantly greater gains than the control group, t (36) = 2.38, $p = .023$, between-group effect size $d = .77$, 95% CI = .11 – 1.43. In contrast, there were no significant difference in gains across the two maths app treatment groups, t (40) = .42, $p = .676$, between-group effect size $d = .13$, 95% CI = -.48 – .74.

Secondary outcome variable: progression through the maths apps

Table 18 reports the mean number of topics completed (number of end of topic quizzes passed) for children in treatment group 1 (Brazilian Portuguese) and 2 (English). An independent samples t-test showed children using the maths apps in Brazilian Portuguese (treatment group 1) made more progress throughout the apps compared to children using the apps in English (treatment group 2), which trended towards significance, t (38) = 1.96, $p = .057$, between-group effect size $d = .63$, 95% CI = .002 – 1.25.

Language of instruction proficiency

Spearmans' rho correlations showed a small but not statistically significant correlation between proficiency in the

language of instruction and mathematics gains (measured with EGMA) across the whole sample, $r^s = .11$, $p = .385$. However, a moderate and statistically significant relationship between proficiency in the language of instruction and progress through the maths apps was identified, $r^s = .31$, $p = .05$.

Table 18 Raw score group mean (SD) at pre-test and post-test with group mean (SD) for difference scores (post-test minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95% confidence interval) for the assessment of mathematical ability (max score 54) and group mean (SD) and range (Min-Max) for number of app topics completed.

Outcome Variable	Group 1 (Brazilian Portuguese; *n* = 23)	Group 2 (English; *n* = 19)	Group 3 (control; *n* = 19)
Mathematics Score (EGMA; max. 54)			
Pre-test Mean (SD)	15.70 (6.15)	20.53 (7.94)	16.68 (6.90)
Post-test Mean (SD)	24.83 (6.35)	30.32 (10.31)	22.53 (7.28)
Gain Mean (SD)	9.13 (4.18)	9.79 (5.94)	5.84 (4.13)
% Gain	16.9	18.1	10.8
Cohen's *d* (95% CI)	1.46 (.54 – 2.38)	1.06 (.10 – 2.03)	.83 (-.11 – 1.77)
Progress through Apps (Brazilian Portuguese max. 22; English max. 28)			
Completed topics Mean (SD)	14.18 (3.16)	12.33 (2.70)	n/a
Completed topics Range (Min-Max)	10.00- 19.00	9.00 – 18.00	n/a

Discussion

Using a quasi-experimental design this study examined the efficacy of the maths app intervention for Brazilian Portuguese-English bilingual children aged 5-6 years old attending a bilingual dual immersion private school in Brazil. Specifically, this study asked if children made greater gains in mathematics when the maths apps were delivered in the child's L1 (Brazilian Portuguese; treatment group 1) or L2 (English; treatment group 2) compared to standard instructional practice only (delivered in English; control group). The influence of proficiency in the language of instruction on children's gains in mathematics and progress within the maths apps was also considered.

This is the first study to examine app-based learning environments in a bilingual immersion context and it was conducted in a controlled setting. Specifically, only the language of instruction was manipulated across the two maths app treatment groups as the time on task and school setting remained consistent across both intervention groups. Children's access to the app mathematics content was also highly similar across the two treatment groups. In doing so, this study advances our understanding of bilingual learners' responses and the applications of bilingual educational theory (Cummins, 1976; 1981; 2000; 2008) to new app-based instructional approaches (Hirsh-Pasek et al., 2015).

Mathematics learning

This study showed children who used the maths apps in Brazilian Portuguese (treatment group 1) and in English (treatment group 2) made significantly greater learning gains in basic mathematics (16.9% and 18.1% respectively) compared to children who received standard mathematics practice only (control group; 10.8%). This suggests that in this bilingual school setting the maths apps were an effective teaching tool for supporting the acquisition of basic mathematical skills. This evidence is consistent with previous research examining the benefits of app-based mathematics instruction (see Chapter 2) and mathematics interventions targeted to bilingual learners focusing on basic mathematical skills (Barnett et al., 2007; Hemsley et al., 2014).

While, there was no significant difference in children's learning gains using EGMA when the maths apps were implemented in the child's L1 (Brazilian Portuguese) or the child's L2 (English), a larger effect size was observed for children who used the apps in Brazilian Portuguese (within-group effect size 1.46) compared to English (within-group effect size 1.06) and a small between-group effect size (Cohen's d = .13, 95% CI = -.48 – .74) across the two treatment groups was observed. This corroborates the results found for progression through the apps, as measured by topics passed. Children who received instruction with the apps in L1 (Brazilian Portuguese) made more progress through the maths apps than children who received instruction in their L2 (English), a difference

that approached significance, and was characterised by a medium effect size (Cohen's *d* = .63, 95% CI = .002 – 1.25).

Language of instruction

When examining if proficiency in the language of instruction was influential in the learning process, a weak but positive association between proficiency in the language of instruction and mathematics gains was found. However, this relationship did not reach statistical significance. This may be because the primary outcome measure, EGMA, did not align closely to the app content (see Table 17). However, a moderate positive and statistically significant relationship was identified between proficiency in the language of instruction and progress through the maths apps (number of topics completed). Specifically, children with higher proficiency in the language of instruction (L1 for treatment group 1; L2 for treatment group 2) made greater progress through the maths apps (i.e. completed more app topics) than children with less proficient language skills.

Limitations

Whilst this is the first intervention study evaluating the use of app-based mathematical instruction for young bilingual learners, there are three important limitations to be aware of when considering the current research findings and directing future studies.

First, whilst time on task was equated for the two treatment groups, the control group received less mathematics instruction

across the study than the two treatment groups (see Table 16). Specifically, children in the two treatment groups had approximately 13 hours of exposure to mathematics instruction, compared to 6-10 hours for the children in the control group. Although this procedure was based on practical reasons due to the school's capacity and routine, the additional time learning mathematics, may in part have accounted for the greater learning gains seen in the two treatment groups (Cheung & Slavin, 2013). Nevertheless, at the very least, these results indicate that incorporating additional mathematics instruction into a weekly timetable significantly improves early learning outcomes for mathematics. These results are consistent with the UK Proof of Concept study (see Chapter 4), which showed the apps significantly raised learning outcomes compared to standard mathematics practice when time on task was equated across the treatment and control groups. To address this potential limitation, future research should include time-equivalent treatment and control groups in a bilingual dual immersion setting.

Second, whilst the language of instruction was manipulated in this study when using the maths apps, an analysis of language used by children during the app sessions was not undertaken. Anecdotal observations during the app sessions revealed many of the children using the maths apps in their L2 (English; treatment group 2) utilised learning strategies, for example, counting out loud, in their L1 (Brazilian Portuguese). This code switching between languages is a very common phenomenon in the bilingual classroom (Moschkovich 2007; Kamwangamalu, 2010). It is suggested code

switching from the L2 to the L1 is a pragmatic response to difficult instructional content (Ferguson, 2003) that reflects the Interdependence Hypothesis (Cummins, 2000). Doing so may have supported children in using the maths apps in their L2 (English). This highlights the value of including a systematic examination of language use in future studies, to enable a deeper understanding of children's learning experiences in app-based learning environments and the independence and transfer between languages and skills (Cummins, 2000).

Third, although the mathematics outcome measure (EGMA; Brombacher, 2011) used in this study was chosen because it was internationally developed and largely required non-verbal responses, it primarily focused on basic mathematics skills, such as mathematical factual knowledge (for example, number recognition) and basic numerical operations (for example, simple addition and subtraction). However, basic mathematical skills only reflect one component of mathematical development. Higher order skills, such as mathematical reasoning (for example, making deductions and inferences from mathematical information) and problem solving (for example, combining and applying different areas of mathematics to solve a problem in a specific context) are also vital components of successful mathematical development (Goswami, 2006; Holmes & Dowker, 2013). Basic mathematical skills underpin these higher order abilities (Van Der Heyden & Burns, 2008; Codding et al., 2010) and bilingual mathematics interventions have been shown to benefit higher order mathematical skills (Hemsely et al., 2014).

Consequently, future research should include more comprehensive measures of mathematical development in order to examine the targeted and generalised benefits of the maths app intervention in a bilingual context.

Conclusion

In summary, Chapter 5 reported a small-scale, proof of concept study in one Brazilian Portuguese-English bilingual primary school. Results showed, in this context, the maths apps were an effective form of mathematics instruction compared to standard mathematics practice for children aged 5-6 years. No significant difference in mathematics learning gains, as measured by EGMA (primary outcome variable) were found when the apps were delivered in the child's L1 (Brazilian Portuguese) compared to when they were delivered in the child's L2 (English).

However, children who used the maths apps in Brazilian Portuguese showed a larger effect size in EGMA scores over time and made more progress through the maths apps, as reflected by the number of app topics completed (secondary outcome variable) compared to children who used the maths apps in English. Furthermore, children with higher proficiency in the language of instruction (L1 for treatment group 1; L2 for treatment group 2) made greater progress through the maths apps than children with less proficient language skills. As children were unbalanced bilinguals, whereby their L1 was significantly stronger than their L2, this may have accounted for the faster progress within the apps

observed for children who received instruction in their L1 compared to children who received instruction in their L2.

Overall, this evidence suggests that learning mathematics with these interactive apps was most effective when implemented in the child's L1. However, further research is needed to explore the influence of specific language proficiencies, particularly cognitive academic language proficiency (CALP) and basic interpersonal communication skills (BICS), as well as children's language responses during implementation. This study advances our understanding of bilingual learners' responses to app-based learning environments in a controlled, real-world, bilingual immersion setting. It is the first application of bilingual educational theory to technology-based intervention that demonstrates support for early mathematical development in bilingual children.

Chapter 6: UK Feasibility RCT

Current study

So far Chapters 4 and 5 have established the initial proof of concept evidence evaluating the maths app intervention in two quasi-experimental studies in the UK and Brazil with low-achieving and bilingual children respectively. In the further development and scaling of the intervention, the next stage is a pilot feasibility trial with a larger sample of children and more rigours quantitative RCT methods (EEF, 2018a; see Chapter 1).

To build this evidence base and address the limitations highlighted in the UK and Brazil Proof of Concept studies (see Chapters 4 and 5), the current study asked two important questions: (1) how are the maths apps most effectively implemented in a classroom setting compared to standard instructional practice? (2) Which components of mathematical developments are supported by the maths apps?

Effectiveness and implementation

First, there is a need to understand how education apps are most effectively implemented in a classroom setting. As identified in the systematic review reported in Chapter 2, app-based mathematical interventions and control conditions have been typically implemented by researchers as supplementary teaching aids. The lack of a time-equivalent, business as usual control group makes it difficult to disentangle the impact of increased instructional

time learning mathematics either in the treatment (e.g. target maths app) or control conditions (e.g. other maths or non-maths apps) from increased time learning mathematics (Foster et al., 2016; Ginsburg & Smith, 2016). It is also vital that teachers implement the intervention in their classroom, to ensure both high ecological validity and generalisability of the intervention beyond the study context (Clements et al., 2015).

To address this issue in the current study, two forms of teacher-based implemented of the new maths app intervention were compared to business as usual, standard mathematics practice. As illustrated in Table 19, children in Group 1 (treatment) used the maths apps in addition to all other standard mathematics activities, and so overall had increased exposure to mathematics instruction. In contrast, children in Group 2 (time-equivalent treatment) used the maths apps instead of a daily small group-based mathematics activity, thus time spent learning maths was equal to the children in Group 3 (control) who received standard teacher-led mathematics instruction that included a daily small group-based activity. As such, in the current study, all children continued to receive whole class mathematics instruction delivered by the teacher, which was embedded into play-based learning, consistent with the principles of guided play (Ash & Wells, 2006; Fisher et al., 2013; Weisberg et al., 2016; see Chapter 1) and is standard practice for early years classrooms in the UK.

In summary, this study asked, do children make more progress when the maths app intervention is implemented by teachers in addition to regular mathematics instruction (Group 1: treatment) or when implemented instead of a daily small-group based mathematics activity (Group 2: time-equivalent treatment) compared with children receiving standard instructional practice (Group 3: control)? Based on previous research (Outhwaite et al., 2017; Pitchford, 2015), it was predicted that children who used the maths apps (Group 1 and 2) would progress more than children receiving standard mathematical instructional practice (Group 3), and children who received the maths apps in addition to their regular mathematics instruction (Group 1) would have the strongest learning gains.

Components of mathematical development

Second, as also highlighted in previous chapters, there is a need to examine which components of mathematical development are supported by educational apps. Previous research evaluating app interventions has frequently used assessments closely aligned with the intervention content (e.g. the Curriculum Knowledge assessment, Outhwaite et al., 2017; Pitchford, 2015), which typically focuses on specific aspects of basic mathematical knowledge (e.g. Number Sense, Schacter & Jo, 2016; 2017; Schacter et al., 2015; Szkudlarek & Brannon, 2018). Studies need to take a broader view of mathematics and consider how educational apps support the acquisition of targeted components of mathematical knowledge and whether this generalises to higher-

level skills (Van der Ven et al., 2017). For example, the maths apps at the focus of this thesis primarily target factual and conceptual knowledge (see Chapter 3). This approach will help elucidate how mathematical development can be supported by interactive, individualised, educational apps.

To address these highlighted issues, the current study utilised a standardised assessment of early mathematical skills that measured the four components of early mathematical development; Fluency in Facts, Fluency in Concepts, Mathematical Reasoning, and Problem Solving (Goswami, 2006; Holmes & Dowker, 2013; see Chapter 1), before and after the 12-week intervention period. This enabled learning gains for each mathematical component, including targeted basic mathematical skills and higher-level knowledge not included in the intervention (See Chapter 3) to be compared across the three intervention groups.

Therefore, this study also asked, for each of the four components of mathematical development, do children make more progress with the apps when used in addition to regular mathematics instruction (Group 1: treatment) or when implemented instead of a daily small group-based mathematics activity (Group 2: time-equivalent treatment) compared with children receiving standard instructional practice (Group 3: control)?

Method

Design

A pupil-level randomised control trial (RCT) was conducted to evaluate the feasibility of the new maths app intervention compared to standard mathematics classroom practice. Consistent to the UK Proof of Concept study reported in Chapter 2, the maths app intervention was evaluated with children aged 4-5 years, in the first year of compulsory education in the UK, known as the EYFS. The study was conducted in the last 14 weeks of the school year before the children transitioned to the next stage of education, known in the UK as Key Stage 1.

There were 12 participating schools across Nottingham and Nottinghamshire in the East Midlands, UK, representing a range of socio-economic and multicultural backgrounds and relatively high levels of educational underachievement compared to other regions of the UK (Ofsted, 2013). Detailed information about the context of each school is reported in the implementation process evaluation in Chapter 7. Within each class of the 12 participating schools, children were randomly allocated to one of three groups outlined in Table 19. As randomisation occurred at the pupil-level within each class, this controlled against school effects influencing results.

Table 19 Description of maths intervention received and total maths-learning exposure for each group across the 12-week intervention period.

Mathematics Activities	Group 1 (treatment)	Group 2 (time-equivalent treatment)	Group 3 (control)
Maths app intervention	✓	✓	
Small group-based maths instruction	✓		✓
Whole class embedded maths activities	✓	✓	✓
Total time learning maths	Additional	Typical	Typical

This three-Group RCT design allowed the effects of maturation (Group 3: control) to be disentangled from the effects of the maths apps (Group 1 and 2) being examined. Furthermore, the inclusion of two treatment groups enabled the effects of the maths apps as a form of quality instruction (Group 2: time-equivalent treatment) to be differentiated from additional exposure to mathematics instruction (Group 1: treatment). Thus, these two forms of app implementation addressed calls from previous research

and policy (APPG, 2014; Ofsted, 2018; Pitchford et al., 2016) by considering if an increased focus on early mathematics should take the form of additional time learning mathematics (treatment) or if current allocated time can be used more efficiently (time-equivalent treatment).

Children were assessed on the Progress Test in Maths, level 5 (PTM5; Maths Assessment Resource Service, 2015), a standardised assessment of mathematical skills, immediately before (pre-test) and immediately after (post-test) an intervention period in which teachers implemented the maths apps for 30 minutes each day across 12 consecutive weeks.

The School of Psychology Ethics Committee at the University of Nottingham granted ethical approval for the study. Opt-in informed parental consent was obtained for all participating children in line with the British Psychological Society ethical guidelines. There were 85% of all available children across the 12 participating schools given parental consent to take part in the study.

Participants

The CONSORT (2010) data in Table 20 summarises the study sample at each stage of the RCT. In total, 461 children aged 4-5 years were randomly allocated to one of three groups. There were 153 children assigned to Group 1 (treatment). These children received the maths app intervention as well as all daily standard mathematics practices. There were 152 children randomly allocated to Group 2 (time-equivalent treatment). These children used the

maths app intervention instead of a daily small group-based mathematics activity that is given as part of standard mathematical practice. The remaining 156 children were assigned to Group 3 (control). These children received regular mathematics teaching only.

Table 20 CONSORT (2010) table describing the composition of the study sample at each stage of the RCT.

Stage of RCT	Total	Group 1 (treatment)	Group 2 (time-equivalent treatment)	Group 3 (control)
Randomized to group	461	153	152	156
Pre-tested	452	149	150	153
Absent at pre-test	9	4	2	3
Withdrew during study	3	0	1	2
Left school	1	0	0	1
Removed from study	2	0	1	1
Post-tested	389	126	131	132
Absent at post-test	63	23	19	21
Final sample	389	126	131	132

452 children from the 12 participating schools were pre-tested on the PTM5 (Maths Assessment Resource Service, 2015). Nine children were absent at pre-test but were still randomised to group. Of the 452 children that were pre-tested, 389 children from 11 schools were available at post-test and were given the same mathematics assessment immediately after the 12-week intervention period.

In total, 63 children who were pre-tested did not complete the post-test; one child left school, two children were removed from the study by their teachers for reasons unknown, and 60 children were absent at post-test, including 30 children from one school because of a school fieldtrip. It was not possible to follow-up on children that were absent on the day of the post-test because post-testing took place during the last week of the school year. Table 21 details descriptive data for the final sample of 389 children.

Table 21 Descriptive data, including mean age in months (SD, Min-Max) and Gender (Female: Male) for the final sample in the trial.

Descriptive Data	Total (*n* = 389)	Group 1 (treatment; *n* = 126)	Group 2 (time-equivalent treatment; *n* = 131)	Group 3 (control; *n* = 132)
Age (months)	60.64 (3.62)	60.39 (3.74)	61.00 (3.61)	60.52 (3.52)
	53.00 - 66.00	54.00 – 66.00	55.00 – 66.00	53.00 – 66.00
Gender (F:M)	196:193	60:66	73:58	63:69

Maths app intervention

The intervention consisted of two maths apps; 'Maths 3-5' and 'Maths 4-6,' developed by *onebillion.* How these maths apps correspond to the mathematical curriculum followed in the Early Years Foundation Stage 2 and Key Stage 1, as well as how the maths apps map onto Science of Learning principles is outlined in detail in Chapter 3.

As highlighted in Table 19 children in Group 1 (treatment) and children in Group 2 (time-equivalent treatment) used the maths apps for 30 minutes each day over the 12-week intervention period.

Small group mathematics instruction

The small group mathematics instruction in this study was consistent with the Number and Shape, Space and Measure content covered in the EYFS Profile (Department for Education, 2013; see Table 8, Chapter 3). Group based activities were delivered by the class teacher and focused on a particular mathematical concept from the EYFS Profile. Example activities were obtained through observations made during schools visits. For example, shape recognition was taught by the teacher drawing different shapes on the whiteboard and asking the small group of children, "What shape am I?" Children responded by calling out the answer and receiving corrective feedback from the teacher before moving onto the next item. In an activity focused on understanding the concepts of more and less, the teacher utilised a number line visual aid and physically demonstrated "1 more than 18." The teacher then asked the small

group of children, “What is 1 more than 10?” and children responded by writing their answer on an individual mini whiteboard and showing it to the teacher. The teacher would the give corrective feedback before moving onto the next item. As highlighted in Table 19, children in Group 1(treatment) and children in Group 3 (control) received instruction through daily small group mathematics activities as part of standard mathematics practice.

Whole class embedded mathematics activities

Consistent with the whole class embedded mathematics activities described in the UK Proof of Concept study (see Chapter 4); examples of whole class activities were obtained through observations during school visits. For example, children were introduced to the concept of volume in a water play activity with containers of different sizes placed in the wet play area. During free play time children had the opportunity to explore which of the different containers could hold the most water. The concept of shape recognition was introduced to children by identifying shapes in their environment. In this activity children explored their classroom environment to find examples of different shapes, which were recorded on a teacher-made worksheet. Teaching staff supervised both activities to facilitate and scaffold children’s exploratory play and thus were consistent with the principles of guided play (Ash & Wells, 2006; Fisher et al., 2013; Weisberg et al., 2016; see Chapter 1). As shown in Table 19, all children participating in this study received this form of play-based embedded instruction.

Mathematics assessment

Children's mathematical ability was assessed using the PTM5 (Maths Assessment Resource Service, 2015). The PTM5 is a paper-based, age-appropriate, standardised measure of mathematical ability, applicable for children aged 4-5 years old in the summer term of the Early Years Foundation Stage 2, when this study took place. The assessment is designed to be used by schools to track children's progress and has been used in other evaluation studies of educational interventions (Jerrim & Vignoles, 2016; Worth, Sizmur, Ager & Styles, 2015).

The PTM5 covers a concise selection of items from the Early Years Foundation Stage Profile (Department for Education, 2013) and is independent of the maths app intervention. The assessment can be delivered to groups of pupils, but this study it was administered on a one-to-one based, to maximise child engagement. Reliability analysis of the PTM5 assessment using a one-to-one administration procedure showed high internal consistency between pre-test and post-test scores, r = .67, Cronbach's alpha = .80.

The PTM5 is designed to assess the four components of mathematical proficiency outlined in Chapter 1, with increasing levels of difficulty and application of mathematical knowledge, for example: (a) Fluency in Facts and Procedures using ordinal numbers and recognising shapes (maximum raw score = 7); (b) Fluency in Conceptual Understanding measures understanding of "most" and

"least" and recognising numbers of quantity (maximum raw score = 9); (c) Mathematical Reasoning assessed drawing conclusions from mathematical information (maximum raw score = 7); (d) Problem Solving assesses making connections between different parts of mathematics to solve a problem in a particular context (maximum raw score = 3). The total maximum raw score was 26 with no discontinuation rule; children completed all questions in this assessment.

Procedure

Recruitment.

A recruitment event was held through the Nottingham Apple Education Regional Training Centre, which is organised by Marc Faulder, an Early Years Specialist and Apple Distinguished Educator (ADE). This connection to the Apple ADE network strongly facilitated school recruitment for this study. The recruitment event was organised in collaboration with the research team and Marc and was designed to inform interested schools about the study and the maths app intervention being trialled. Participating schools already had access to the required hand-held tablet device hardware (iPads). Participating schools were given access to the maths apps, free of charge, by *onebillion.* Children were recruited through schools that agreed to participate in the study and opt-in parental consent was obtained for all participating children before study commencement.

Support system.

As the study was been conducted in a larger number of schools compared to the proof of concept studies reported in Chapters 2 and 3, the support system for intervention implementation (Domitrovich et al., 2008) was much greater than previous studies. It consisted of pre-intervention teacher training, a teacher manual, school visits, and continued email communication with the researchers throughout the intervention period.

Pre-intervention teacher training is the most common form of support system (Fixsen, Naoom, Blase, & Friedman, 2005) and has been shown to support good implementation fidelity (Basen-Engquist et al., 1994), particularly with a large number of participating schools. The pre-intervention teacher training session for this study was delivered during the recruitment event mentioned above. The session aimed to give teachers the knowledge and skills they needed to use the maths apps intervention, for example, how to set up individual children's in-app profiles, demonstrations on how children progress through the app and how to deal with common difficulties. During the session teachers had the hands-on opportunity to use the maths apps and ask questions to the research team and Marc Faulder, an Early Years Specialist and Apple Distinguished Educator who has previous experience of using the maths app intervention in his class.

A teacher manual (see Appendix E) was also produced to provide further guidance on how to implement the maths app

intervention and included full details of the study protocol to maximise consistency across all participating schools. The manual did not provide instructions for when and how mathematics should be taught in standard practice. The implementation of standard practice was consistent with national guidelines provided in the Early Years Foundation Stage Profile (Department for Education, 2013) and was at the discretion and autonomy of the teaching and senior staff, as is standard in England (the Early Years Foundation Stage Profile is not statutory in Scotland and Wales).

Each school was also visited once before the trial commenced to ensure the teaching staff were fully informed of the study protocol and had successfully embedded the maths app intervention into their daily routine. Although school implementation timetables varied because of iPad availability and individual school routines, I visited all schools again during the 12-week intervention period. The purpose of this second visit was to interview the teaching staff about their experiences of the maths apps, observe the maths app intervention sessions and ensure intervention compliance. Qualitative data from these school visits is the focus of Chapter 7 (UK RCT Implementation Evaluation). The small group mathematics instruction and whole class embedded mathematical activities were also observed to gain an understanding of the standard mathematics practice.

Each school also had continued email communication support from the researcher throughout the duration of the 12-intervention

period. Support queries primarily focused on technical challenges with downloading the maths apps onto the school iPads at the start of the study, which I promptly dealt with to minimise disruption and inconvenience for the teachers.

Group allocation.

Within participating schools, children in each class were randomly assigned to one of three groups (see Table 19). This controlled for potential school and teacher effects, as the apps were implemented by the usual teaching staff in the 12 participating schools to optimise ecological validity (McNeal, Hansen, Harrington, & Giles, 2004; Spoth, Guyll, Lillehoj, Redmond, & Greenberg, 2007). As each participating class typically had one teacher and one or two teaching assistants this allowed all children from each class to be taught/supervised by their usual teaching staff that were known to them.

To guard against selection bias, each child was given a unique study identification number by class teachers as parental consent was returned. These unique study identification numbers were given to me and I then randomised children to group allocation using a random number generator. This way, the all the researchers were blind to group allocation. Once group allocation was complete, I returned a list of study identification numbers for each group to each of the class teachers taking part in the trial. Teachers were then able to match the study identification number and group allocation to individual children taking part in the study. The group

allocations remained fixed throughout the study and the researchers remained blind to the correspondence between study identification number and group allocation throughout the study.

Assessment administration.

The standardised paper-based mathematics assessment (PTM5) was administered immediately before (1 week) and immediately after (1 weeks) the 12-week intervention period. A team of trained assessors delivered the PTM5 to individual children, on a one-to-one basis, in a quiet area, free from distraction, in the child's familiar school environment. Before the assessment began, the trained assessor explained the task to the child, ensuring the child felt comfortable and explicitly asked if the child would like to take part in the assessment so as to gain assent. Because of the child's age, verbal or non-verbal assent (using a smiley face/ not smiley face response sheet) was accepted. The trained assessor then read aloud the questions from the PTM5 one at a time and the child was required to write their response in the response booklet. Each mathematics assessment lasted 10-15 minutes per child and was completed in one sitting.

The trained assessors were recruited from the University of Nottingham and were trained on administration by the research team. To guard against detection bias, all trained assessors were blind to group allocation so were unaware which children had received the maths app intervention. Completed assessment booklets were sent to GL Assessment, who independently scored all

assessments and were consequently also blind to group allocation and condition.

Maths app intervention implementation.

Children assigned to receive the maths app intervention in Group 1 (treatment) and Group 2 (time-equivalent treatment) used the maths apps for approximately 30 minutes a day, five times a week for 12 weeks. Children started using the apps on topic one in 'Maths 3-5.' Once children had completed all 10 topics, they progressed to 'Maths 4-6.' Teachers kept a register of which children had completed the 'Maths 3-5' app to track progress and support daily organisation. Children used the same iPad each day, as they had their own profile in the apps, which saved their progress. To support classroom organisation, iPads were colour coded or numbered by the teaching staff. The intervention was administered in small groups of between 10 to 15 children, depending on class size. Children worked independently through the apps, using headphones in a quiet area of their normal classroom, supervised by teaching staff who provided technical support and ensured children remained focused on the tasks.

The intervention was embedded into the daily classroom routine. As highlighted in Table 19, children assigned to Group 2 (time-equivalent treatment) used the intervention while the other children received a small group-based mathematics input activity. This ensured children in Group 2 (time-equivalent treatment) received the same amount of time on mathematics instruction as

Group 3 (control). Children assigned to Group 1 (treatment) used the maths apps at a different time during the school day, for example, during free play sessions, so as not to miss out on core subjects, such as phonics. Despite being implemented during free play time, children were supervised and instructed to use the maths apps by teaching staff; it was not an optional activity, as is sometimes conventional in free play. As such, children in Group 1 (treatment) had more exposure to mathematics instruction over the 12-intervention period. At the end of the study, the participating schools continued to have access to the maths apps, so the apps were available to all children.

Time on task.

Children allocated to receive the new maths app intervention (Group 1 and 2) used the maths apps for 30 minutes a day, five days a week, for 12 consecutive weeks, totalling 1800 minutes (30 hours) exposure to the maths app intervention.

Results

Preliminary analysis

Standardised norms for the PTM5 are based on group administration (Maths Assessment Resource Service GL Assessment, 2015) but as this study used one-to-one administration to maximise child engagement standardised scores were not considered appropriate. Instead, raw scores on the PTM5 were used as the dependent variable. A one-way Analysis of Variance (ANOVA)

showed no significant age differences across the three instruction groups, $F(2,386) = 1.03$, $p = .358$, confirming it suitable to use raw scores. There were also no observed gender differences across the three groups, $\chi(2) = 2.25$, $p = .324$. Despite making directional hypotheses that the maths apps would be more effective in supporting the development of early mathematical skills than standard classroom practice, all analyses are reported at a 2-tailed level of probability, unless otherwise stated.

Effectiveness and implementation

To establish which form of implementation of the maths app intervention was the most effective compared to standard classroom practice in supporting the acquisition of early mathematical skills and to account for minor differences in pre-test mathematical ability across the three groups (see Table 22; Van Breukelen, 2006), mean mathematical performance for each of the three groups (see Table 22) was compared using a 2 (Time: pre-test, post-test) x 3 (Group: Group 1, Group 2, Group 3) mixed ANOVA. Results showed a significant interaction between Time and Group, $F(2,386) = 3.10$, $p = .046$.

Analysis of simple main effects showed no significant differences between groups at pre-test, $F(2,386) = 1.63$, $p = .197$, or post-test, $F(2,386) = .12$, $p = .888$. Over time, significant learning gains were found for Group 1 (treatment), $t(125) = 10.54$, $p < .0001$, Group 2 (time-equivalent treatment), $t(130) = 8.97$, $p < .0001$, and Group 3 (control), $t(131) = 7.03$, $p < .0001$. The

largest within-group effect sizes (Cohen's *d* with 95% CI) reflecting the magnitude of progress were observed for Group 1 (treatment; see Table 22).

To explore the significant interaction further, planned comparisons with independent samples t-tests and between-group effect sizes (Cohen's *d* with 95% CI; Trafimov, 2015) were conducted on progress made over time (difference scores; post-test minus pre-test). Results showed that pupils in Group 1 (treatment) made significantly more progress over the 12-week intervention period than Group 3 (control), $t(256) = 2.46$, $p = .015$, between-group effect size = .31, CI = .06 - .55. Similarly, pupils in Group 2 (time-equivalent treatment) made significantly greater learning gains compared to pupils in Group 3 (control) in line with the predictions, $t(261) = 1.74$, $p = .042$, 1-tailed, between-group effect size = .21, CI = -.03 – .46. There was no significant difference in the progress made by Group 1 (treatment) and Group 2 (time-equivalent treatment), $t(255) = .64$, $p = .525$, between-group effect size = .08, CI = -.17 – .33.

Components of mathematical development

To examine how effective the maths apps were at supporting the acquisition of the four components of mathematical proficiency (Fluency in Facts, Fluency in Concepts, Mathematical Reasoning and Problem Solving) compared to normal classroom practice, mean performance on each component for the three groups (See Table 5) were compared using separate 2 (Time: pre-test, post-test) x 3

(Group: Group 1, Group 2, Group 3) mixed ANOVAs. Results showed no significant Time x Group interactions for any of the four components; Fluency in Facts $F(2, 386) = .56$, $p = .570$; Fluency in Concepts $F(2, 386) = 2.04$, $p = .132$; Mathematical Reasoning $F(2, 386) = 1.37$, $p = .256$; Problem Solving $F(2, 386) = 1.42$, $p = .244$.

Table 22 Raw score group mean (SD) at pre-test and post-test with group mean (SD) difference scores (post-test minus pre-test), percentage gains and within-group Cohen's *d* effect sizes (95% confidence interval) for the assessment of mathematical ability and each component of mathematical proficiency.

Outcome Variable	Group 1 (treatment; *n* = 126)	Group 2 (time-equivalent treatment; *n* = 131)	Group 3 (control; *n* = 132)
Total Raw Score (max. 26)			
Pre-test Mean (SD)	11.82 (4.73)	11.85 (4.74)	12.77 (4.99)
Post-test Mean (SD)	15.29 (4.16)	15.02 (5.08)	15.09 (4.82)
Gain Score Mean (SD)	3.48 (3.70)	3.17 (4.04)	2.33 (3.80)
% Gain	13.3	12.2	9.0
Cohen's *d* (95% CI)	.78 (.42 – 1.14)	.65 (.29 – 1.00)	.47 (.13 – .82)
Fluency in Facts (max. 7)			
Pre-test Mean (SD)	2.69 (1.85)	2.66 (1.78)	2.88 (1.92)
Post-test Mean (SD)	3.56 (1.82)	3.68 (2.06)	3.64 (1.95)
Gain Score Mean (SD)	.87 (1.98)	1.02 (2.00)	.76 (1.93)

Outcome Variable	Group 1 (treatment; *n* = 126)	Group 2 (time-equivalent treatment; *n* = 131)	Group 3 (control; *n* = 132)
% Gain	12.4	14.6	10.9
Cohen's *d* (95% CI)	.47 (.12 – .83)	.53 (.18 – .88)	.39 (.05 – .74)
Fluency in Concepts (max. 9)			
Pre-test Mean (SD)	5.70 (2.24)	5.80 (2.18)	6.05 (2.06)
Post-test Mean (SD)	6.90 (1.61)	6.78 (1.81)	6.76 (1.81)
Gain Score Mean (SD)	1.20 (2.03)	.98 (2.05)	.71 (1.72)
% Gain	13.3	10.9	7.9
Cohen's *d* (95% CI)	.62 (.26 – .97)	.49 (.14 – .84)	.37 (.02 – .71)
Mathematical Reasoning (max. 7)			
Pre-test Mean (SD)	2.72 (1.63)	2.69 (1.58)	3.01 (1.81)
Post-test Mean (SD)	3.74 (1.56)	3.49 (1.72)	3.65 (1.69)
Gain Score Mean (SD)	1.02 (1.86)	.79 (1.75)	.64 (1.83)
% Gain	14.6	11.3	9.1
Cohen's *d* (95% CI)	.64 (.28 – 1.00)	.48 (.14 – .83)	.37 (.02 – .71)

Outcome Variable	Group 1 (treatment; *n* = 126)	Group 2 (time-equivalent treatment; *n* = 131)	Group 3 (control; *n* = 132)
Problem Solving (max. 3)			
Pre-test Mean (SD)	.71 (.73)	.69 (.71)	.83 (.75)
Post-test Mean (SD)	1.10 (.84)	1.07 (.91)	1.05 (.82)
Gain Score Mean (SD)	.39 (.91)	.38 (1.06)	.21 (.91)
% Gain	13.0	12.7	7.0
Cohen's *d* (95% CI)	.50 (.14 – .85)	.47 (.12 – .81)	.28 (-.06 - .62)

Discussion

This chapter reports the first pupil-level RCT to be conducted evaluating the feasibility of the new maths app intervention for children aged 4-5 years old in the UK. Specifically, this RCT evaluated two forms of maths app intervention (treatment and time-equivalent treatment) compared within a standard practice control. This is the first maths app intervention study to adopt this novel and rigorous research design and it is the first study to take a componential view of mathematical development when evaluating app-based interventions.

Effectiveness and implementation

This study found that combining child-centred, curriculum based, apps with interactive touch-screen tablet technology for children aged 4-5 years old in the first year of primary school provided an effective means of delivering quality instruction that promoted that development of early mathematical skills. Specifically, the interpretation of between-group effect sizes based on normative expectations of change (Hill et al., 2008) showed at the whole sample level, children in Group 1 (treatment) who used the maths apps in addition to all normal mathematics practices were approximately 3-4 months[8] ahead of their peers in Group 3 (control)

[8] Hill et al.'s (2008) approach for translating effect sizes was applied using the following calculation: .31 (reported between-group effect size) x .095 (1 month progress for children aged 4-5 years based on nationally normed mathematics tests) = 3.26 (months progress). This approach has also been used in previous maths app intervention research (e.g. Schacter et al., 2015).

receiving standard practice only (between-group effect size .31). Children in Group 2 (time-equivalent treatment) who used the maths apps instead of one daily small group mathematics activity were shown to be approximately 2 months ahead of children in Group 3 (control) receiving standard practice only (between-group effect size .21). However, there was no significant different between implementing the maths apps as well as all standard mathematics practices or instead of one regular small group mathematics activity. This suggests the apps are a form of quality of mathematical instruction and suggests an increased focus on early mathematics at the class level (APPG, 2014; Ofsted, 2018; Pitchford et al., 2016) can take the form of efficient instructional practices without the need for extra time learning mathematics, which could potentially detract from other areas of a well-rounded curriculum.

Components of mathematical development

When examining which components of mathematical development were supported by the maths app intervention, results showed no significant Time x Group interactions for any of the four components of mathematical proficiency. This suggests all three forms of instruction supported the four areas of mathematical development.

However, when comparing within-group effect size (see Table 22), which reflect the magnitude of progress in each group, larger effect sizes were consistently observed for children who used the maths apps (Group 1 and 2) compared with children who received

standard practice (Group 3) across all four mathematical components. This suggests the maths apps strongly support learning in the areas of mathematical development targeted by the intervention (Fluency in Facts and Fluency in Concepts, see Chapter 3) that generalises to other higher-level skills not extensively or explicitly covered in the app content (Mathematical Reasoning and Problem Solving). These results suggest that the maths apps go beyond drill-based practice for fact retrieval (Salminen, Koponen, Leskinen, Poikkeus & Aro, 2015) and help build a strong foundation in basic mathematical skills that enables maths facts and simple concepts to become automated and assimilated into a higher-level conceptualisation of mathematical knowledge (Codding et al., 2010; Gersten & Chard, 1999; Karmiloff-Smith, 1994; Mayfield & Chase, 2002; Van Der Heyden & Burns, 2005). These results are consistent with cumulative learning theory (Gagné, 1968) and agree with previous intervention studies that have also targeted basic mathematical skills acquisition and demonstrated evidence of generalisation to higher-level mathematical skills (Fuchs et al., 2009; Kidd et al., 2013; van der Ven et al., 2017).

Limitations

This study contributes to the mounting evidence base evaluating this new maths app intervention developed by *onebillion* as part of a staged scaling of implementation (Outhwaite et al., 2017; Pitchford, 2015). Blind group allocation, the inclusion of a time-equivalent treatment group, and an independent standardised mathematics assessment help to address potential threats to

validity and current concerns regarding the usefulness of RCTs in educational research (Ginsburg & Smith, 2016). However, there are three important limitations to consider when interpreting the results of the current study and directing future research.

First, there were minor difference in the instructional content between the intervention treatment groups and the control group. Specifically, the maths apps included a number of topics from the UK National Primary Curriculum for Key Stage 1 (see Chapter 3). This poses a potential threat to validity (Cheung & Slavin, 2013; Ginsburg & Smith, 2016; Shadish et al., 2002) as children in the intervention groups had access to more advanced topics than children in the control group. However, this issue was addressed through the use of the independent assessment measure, which only included a concise selection of items from the EYFS Profile (Department for Education, 2013; see Chapter 3). As children were not assessed on the more advanced topics from Key Stage 1, for example, odd and even numbers, the intervention groups did not have an unfair advantage over their peers in the control group in the assessment procedures.

Second, as the intervention was implemented towards the end of the first year of school, some of the higher ability children in the two intervention treatment groups may have already mastered some of the mathematical skills covered in topics that were presented early in the apps. The revision of mathematical content may have benefited their learning (Dunlosky et al., 2013) and

boosted their confidence with the app technology. Alternatively, it may have delayed further progress within the 12-week intervention period. In contrast, children in the standard practice control group continued to work through the early years curriculum content, without this revision.

To address this issue, entry-level placement strategies should be developed consistent with the curriculum and components of mathematical development (Kadosh, Dowker, Heine, Kaufmann, & Kucian, 2013) and results from future research utilising this tool should be compared to this study to examine the impact of revision on mathematical progress.

Third, while this study demonstrates the benefits of this new maths app intervention for supporting the development of young children's mathematical skills, it is important to recognise that technology alone will not lead to success; but is dependent on how the technology is embedded into the school environment (Beach & O'Brien, 2015; Couse & Chen, 2010). Furthermore, it is critical that app-based mathematical interventions are evaluated within a systems approach.

Detailed qualitative research is needed, therefore, to explore insights into implementation factors at the teacher, classroom, and child level. Although no main effects or interactions were found in the current study when School was entered as an independent variable in the analyses reported above, understanding the school ecosystem in which the intervention is situated and how this may

be associated with children's learning with the maths apps is vital for developing recommendations for best practice in implementation. This will be addressed in the UK RCT Implementation Evaluation reported in Chapter 7.

Conclusion

In summary, Chapter 6 reported a pilot, feasibility, pupil-level RCT evaluating the new maths app intervention with 389 children aged 4-5 years across 11 EYFS classrooms in Nottingham and the surrounding areas. Consistent with the studies reported in Chapters 4 and 5, children randomly assigned to use the maths apps either as a supplementary teaching aid or instead of one daily mathematics activity made more progress in mathematics compared to control children who received standard practice. There was no significant difference between the two forms of app implementation, thus suggesting the maths apps can be embedded in a well-balanced early years curriculum. This study further drives the development and scaling of the new maths app intervention and will be supplemented by the mixed-methods Implementation Evaluation reported in Chapter 7.

Chapter 7: UK RCT Implementation Evaluation

Current study

Despite the growing evidence base demonstrating the educational benefits of the new maths app intervention for young children (see Chapters 4-6), it is important to recognise that learning gains are not isolated to the intervention alone, but may also be associated with how well the intervention is implemented (Fixsen et al., in Cook & Odom, 2013). More specifically, technology alone will not lead to success; but is dependent on how the technology is integrated into the school environment (Beach & O'Brien, 2015; Couse & Chen, 2010). Differences in implementation across the participating schools may influence the learning outcomes of the maths app intervention (Humphrey et al., 2016). As such, detailed qualitative research is needed to explore insights into the implementation of the maths app intervention within the classroom context.

In response, the study reported in this Chapter asked two key questions: (1) how is the maths app intervention implemented as part of an individual school's ecological system across the final sample of 11 participating schools in the UK Feasibility RCT reported in Chapter 6, and (2) is there a relationship between the identified implementation themes and children's learning gains with the maths apps?

Understanding implementation

First, there is a need to understand the implementation of the maths app intervention across the 11 schools that participated in the UK Feasibility RCT reported in Chapter 6. As highlighted in Chapter 1, it is argued that the primarily post-positivist approach in educational research, which places an emphasis on RCT methods (Connolly et al., 2017; Haynes et al., 2012; Torgerson & Torgerson, 2001) cannot capture the complexity of interventions (Biesta, 2010) and can be charged with removing the implementation context, which may be associated with learning outcomes (Elliot, 2004). While critiques call for pragmatic, mixed-methods approaches to address these concerns (Thomas, 2016; Shaffer, 2011), there remains a significant lack of intervention implementation evaluation within RCTs conducted in education (Connolly et al., 2018).

To address these issues, this chapter adopted a pragmatic, mixed-methods approach, utilising a multi-dimensional definition of implementation, which incorporated implementation quality, adaptation, and participant responsiveness (see Chapter 1). To achieve the insights needed, the current study conducted narrative, direct observations of the maths app intervention sessions and semi-structured, self-report interviews with participating teachers within a determinant, ecological systems approach (Bronfenbrenner, 1979; Thomas, 1992; Nilsen, 2015). This provided an understanding of how the maths apps were implemented in the 11 individual school contexts.

Implementation themes and intervention outcomes

Second, there was a need to examine the potential relationship between the identified implementation themes and the observed intervention outcomes (Elliot & Mihalic, 2004; Domitrovich et al., 2010). Previous research has shown a significant association between differences in the implementation of school based health interventions and intervention outcomes (Askell-Williams et al., 2013; Hansen et al., 2013). However, no study to date has adopted this approach when examining maths app instructional interventions. This will be an important research component for driving recommendations for best app-based teaching practice.

To achieve this the current study employed a structured judgements approach (Clarke, 2004) to consider the relationship between the identified implementation themes at the teacher, classroom, and child level and children's learning gains with the maths apps (reported in Chapter 6).

Aim 1: Understanding Implementation

Methods

Qualitative methodology

As discussed in Chapter 2, this thesis is orientated around the philosophy of pragmatism and utilises pluralistic approaches and methods that are considered the most appropriate to effectively understand the reach and utility of the maths app intervention in a real-world context (Tashakkori & Teddlie, 2010). Following this

pragmatic epistemology, as well as taking into consideration that the implementation of the maths app intervention has not been previously studied and that current determinant frameworks are not suitable for the current study (see Chapter 1), a primarily qualitative approach was adopted (Morse, 1991). More specifically, an inductive, bottom-up approach was taken, whereby implementation themes were identified from the data (Punch, 2013; Nilsen, 2015). This study utilised two sources of data to support implementation evaluation: (1) narrative, direct observations made by the researcher and (2) self-report, semi-structured interviews with the implementing teachers.

Narrative, direct observations

Research shows direct observations of implementation are more predictive of intervention outcomes than self-report methods (Domitrovich, Gest, Jones, Gill & Sanford Derousie, 2010). Consequently, direct observations collected in-situ are preferred (Hansen, 2014; Humphrey et al., 2016). Furthermore, unstructured observations are beneficial when the research topic is understudied and initial exploration is needed (Robson & McCartan, 2011).

However, direct observations only provide a momentary snapshot of intervention implementation that may or may not be consistent across all implementing sessions over the 12-week intervention period. To address this issue, multiple direct observations can be taken over time and research shows averaged ratings from multiple observations are more reliable than a single

measure (Resnicow et al., 1998). However, multiple direct observations are not always practical as they are time consuming and can also increase the risk of attrition (Humphrey et al., 2016).

Self-report interviews

On the other hand, self-report methods are often more practically feasible. Self-report methods can give detailed, summative, retrospective accounts of intervention implementation over the full length of the intervention period, instead of the limited snapshot inherent to direct observation methods (Humphrey et al., 2016). However, self-report methods are more vulnerable to sources of bias, such as demand effects, where implementers do not wish to be perceived negatively by the researchers (Humphrey, 2013). Responders' recall, attention, and understanding of different factors may also vary, for reasons unknown to the researcher. Collectively, these aspects may influence results. Research shows self-report data, when used in isolation as an implementation process evaluation (IPE) method, struggles to identify a relationship with intervention outcomes (Goldberg, Lillehoj, Griffin & Spoth, 2004).

Assimilation of narrative, direct observations and self-report interviews

To overcome these outlined methodological challenges and to ensure the qualitative methods used are compatible with the 'real-world' delivery context of the intervention a mixed-methods approach is recommended (Hansen, 2014). A mixed-methods

approach also helps to maximise internal validity (Humphrey et al., 2016). Consequently in this study, data collected from the narrative, direct observations of the maths app intervention sessions and the self-report interviews with teachers were assimilated to provide the most detailed holistic account of the maths app intervention implementation experience possible within the practical constraints of the study (Clarke, 2004).

Participants

Table 23 summarises the descriptive data for the participating teachers. Descriptive data was collected through a self-report, end of project feedback questionnaire and was available for 8 of the 11 participating schools in the final sample.

Table 23 Descriptive data for participating teachers

Characteristic	Descriptive data
Number of Teachers per EYFS Class (Min- Max)	1-3
Number of Years Teaching (Mean [SD], Min -Max)	9.92 (7.86), 2-25
Teaching Role (Total Frequency)	
Senior Leaders (e.g. Head Teacher)	1
Teachers (including EYFS Leaders)	9
Teaching Assistants	3
Technology Experience (Total Frequency)	
Experienced	10
Little Experience	3

School context

As reported in Chapter 6 (UK Feasibility RCT) there were initially 12 participating schools across Nottingham and Nottinghamshire in the East Midlands, UK, a geographical area with high levels of educational underachievement relative to other areas of the UK (Ofsted, 2013).

The schools represented a range of socio-economic and multicultural backgrounds and had a range of characteristics and

Ofsted report ratings. Ofsted inspection reports assess schools as outstanding, good, requires improvement, or inadequate (Ofsted, 2018) and can be used as a proxy measure of school quality (Gambaro, Stewart, & Waldfogel, 2015; Schagen et al., 2005). These judgement ratings are based on direct, structured observations of the school and teaching environment conducted by a team of Ofsted inspectors typically every 6 years. Although inspectors have clear and detailed guidance on inspection criteria, the ratings are based on the discretion of the inspectors, which may vary (Schagen et al., 2005). The proportion of children eligible for pupil premium or free school meals (additional government funding for children from families earning below a certain threshold) as reported in the school's most recent Ofsted report was used to tentatively indicate the SES of the school. For example, a school described here as having an SES 'below national average' was reported to have above national average proportions of children eligible for these funds in the relevant Ofsted report. Table 24 summarises the profile of each school.

Table 24 Profiles of the participating schools.

School	Ofsted rating	Indicated SES profile	Other School Characteristics
A	Good (2014)	Above national average	School A was smaller than average sized, mixed gender primary school (aged 4-11 years). The majority of children were White British and the number of children with SEN was below the national average.
B & C	Outstanding (2013)	Above national average	Schools B & C are two campus sites of the same large mixed gender primary school. Children attend one site only. The majority of children were White British, with a small minority of children from other ethnic groups. The number of children with SEN was below the national average.
D	Requires improvement (2013)	Above national average	School D was an average sized, mixed gender primary school. The majority of pupils were White British and children with SEN were below the national average.

School	Ofsted rating	Indicated SES profile	Other School Characteristics
E	Good (2013)	Below national average	School E was a smaller than average sized, mixed gender infant school (ages 4-7 years). The majority of children were White British and the number of children with SEN was below the national average.
F	Good (2012)	Below national average	School F was a larger than average sized, mixed gender Academy primary school. Children were mostly White British and the number of children with SEN was above the national average.
G	Good (2013)	Above national average	School G was an average sized, mixed gender primary school and the majority of pupils were White British. The number of children with SEN was below the national average. This school was absent during the post-test period due to a fieldtrip and were therefore not included in the final sample of schools.
H	Outstanding (2010)	Above national	School H was a larger than average, mixed gender infant school. Children came from a wide range of ethnic backgrounds, the largest

School	Ofsted rating	Indicated SES profile	Other School Characteristics
		average	proportion was Indian or Pakistani backgrounds. The number of children with EAL and SEN was above the national average.
I	Good (2011)	Above national average	School I was a larger than average sized, mixed gender Academy primary school. The majority of children were from White British, Asian or Asian British backgrounds. The number of children with EAL was above the national average and children with SEN were in line with the national average.
J	Good (2014)	Above national average	School J was a larger than average, mixed gender primary school. The number of children with SEN and children from ethnic backgrounds with EAL was above average.
K	Good (2010)	NS	School K was a larger than average, mixed gender primary school. The majority of children were of White British, Indian or Pakistani background. Approximately 25% of children spoke EAL and the number of children with SEN was below average.

School	Ofsted rating	Indicated SES profile	Other School Characteristics
L	Outstanding (2011)	Below national average	School L was an average sized, mixed gender Academy primary school and the majority of children were from ethnic minority background. The number of children with EAL and SEN was above the national average.

Note. SEN = special educational needs; EAL = English as an additional language; SES = socio-economic status; NS = not stated.

Qualitative data collection protocols

Researcher position.

Qualitative research is inherently subjective. In particular, the researcher is central to the collection and interpretation of qualitative data (Ritchie, Lewis, Nicholls, & Ormston, 2013) Therefore, it is vital that the researcher acknowledges their position within the research context and demonstrates reflexivity (Braun, Clarke, & Terry, 2014).

The researcher describes herself as a mid-twenties aged, white British female from a mixed socio-economic area. This research is part of her PhD in Psychology at the University of Nottingham examining the reach and utility of a maths app intervention to support early mathematical development in young children. After conducting the pilot research reported in the previous chapters, which drew upon primarily positivist or post-positivist paradigms, the researcher was interested in how the maths app intervention would effectively scale when used in more schools and what factors may influence children's learning gains with this technology. The adoption of more relativist approaches was therefore needed.

Narrative, direct observations.

In order to illuminate the maths app intervention implementation, informal, exploratory, direct observations were conducted in-situ by the researcher (Robson & McCartan, 2011; Hansen, 2014; Humphrey et al., 2016). The researcher conducted

the observations as a marginal participant (Robson & McCartan, 2011), whereby they were completely accepted in the classroom environment but remained passive while observing the maths app intervention session. The researcher was positioned close to the children using the maths apps, so that it was possible to observe children's interactions with the technology without disrupting the session.

The observations were semi-structured and guided by interests in implementation quality, adaptations, and participant responsiveness (Creswell & Creswell, 2018). However, as the observations were exploratory, a flexible design (Irwin & Bushnell, 1980) was followed, whereby additional observations that were not initially anticipated were also recorded. Direct observation fieldwork notes were recorded during the observations by the researcher. This pragmatic approach was taken to help develop a full and descriptive narrative of the maths app intervention implementation (Robson & McCartan, 2011). This level of detail and complexity can frequently be lost in more structured approaches to observational methods (Creswell & Creswell, 2018).

Self-report interviews.

To further understand how the maths apps were implemented in the 11 individual school contexts, free flowing, semi-structured interviews were conducted with participating teachers. The interviews were guided by the questions outlined in Table 25; however this protocol was a guiding instrument only (Cohen,

Manion & Morrison, 2007). Rather the self-report interviews were grounded as an authentic experience, whereby the researcher was keen to follow the lead of the teacher. This helped to maintain the natural, free flowing nature of the interview conservation. Fieldwork notes were recorded by the researcher during the interview. Interviews were not audio recorded and thus increased the ecological validity of the interview data. Furthermore, audio recordings of interview data were not necessary as the data collected did not require extensive interpretation and re-construction, as sometimes seen in qualitative research methods (Creswell & Creswell, 2018). Instead, the interview data collected in the current study focused on explicit and surface level meaning of teachers' practical responses (Cohen et al., 2007).

Table 25 Guiding question used in the self-report interviews

Self-report interviews guiding questions
How have you found implementing the maths apps?
What have you found challenging about implementing the maths apps?
What successes have you had with the maths apps?

Qualitative validity and reliability.

It is vital that the rigour and quality of the data collection and analysis processes are considered to ensure the validity and reliability (or authenticity and trustworthiness) of the research

findings and consequent interpretations (Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005; Creswell & Creswell, 2018). Qualitative validity encompasses checks made by the researcher to ensure the accuracy of findings from the standpoint of the researcher and the participant (Gibbs, Kealy, Willis, Green, Welch, & Daly, 2007; Creswell & Miller, 2000). On the other hand, qualitative reliability refers to measures taken to ensure the approach taken is consistent (Gibbs et al., 2007). Table 26 outlines the five qualitative validity and qualitative reliability measures taken in the current study.

Table 26 Details of the five measures taken to ensure qualitative validity and reliability

Measure	Description
Qualitative validity	
Member checking	When conducting the self-report interviews participants' responses as recorded by the researcher were checked back with the participant throughout the duration of the interview. This ensured accuracy and verification of the information.
Thick, detailed description	Where possible sufficient descriptions were recorded in the fieldwork notes to support the researcher's surface level interpretations and conclusions.
Familiarity with the participating organisation	All participating schools met with the research team at the organised recruitment event (see Chapter 6). The researcher also visited all participating teaching staff prior to study commencement and was in regular communication with the participating schools throughout the duration of the study. This also helped establish a rapport between the researcher and all of the participating teachers.

Measure	Description
Addressing demand effects	Prior to the self-report interviews participants were encouraged to be honest and were assured that there were no right or wrong answers and that we, as researchers, wanted to understand their opinions and experiences of the maths app intervention. This step is vital as self-report interviews are frequently vulnerable to bias as participants do not wish to be negatively perceived by the researcher (Humphrey, 2013).
Qualitative reliability	
Audit trial	Detailed description and justification of the research questions, design, method, procedure, and analysis are given. Furthermore, detailed records of all data collected, coded, displayed, synthesised, and interpreted were securely kept.
Researcher position	The researcher's position within the research context has been stated.

Qualitative data analysis

As discussed, this chapter has two primary aims; (1) to understand the implementation of the maths app intervention as part of an individual school's ecological system across the final sample of 11 participating schools in the UK Feasibility RCT reported in Chapter 6, and (2) to examine the relationship between the identified implementation themes and children's learning gains with the maths apps. Table 27 provides an overview summary the analytical procedures followed in the current study to address these two aims (Creswell & Creswell, 2010; Clarke, 2004).

Table 27 Qualitative data analysis process followed (Creswell & Creswell, 2018; Clarke, 2004).

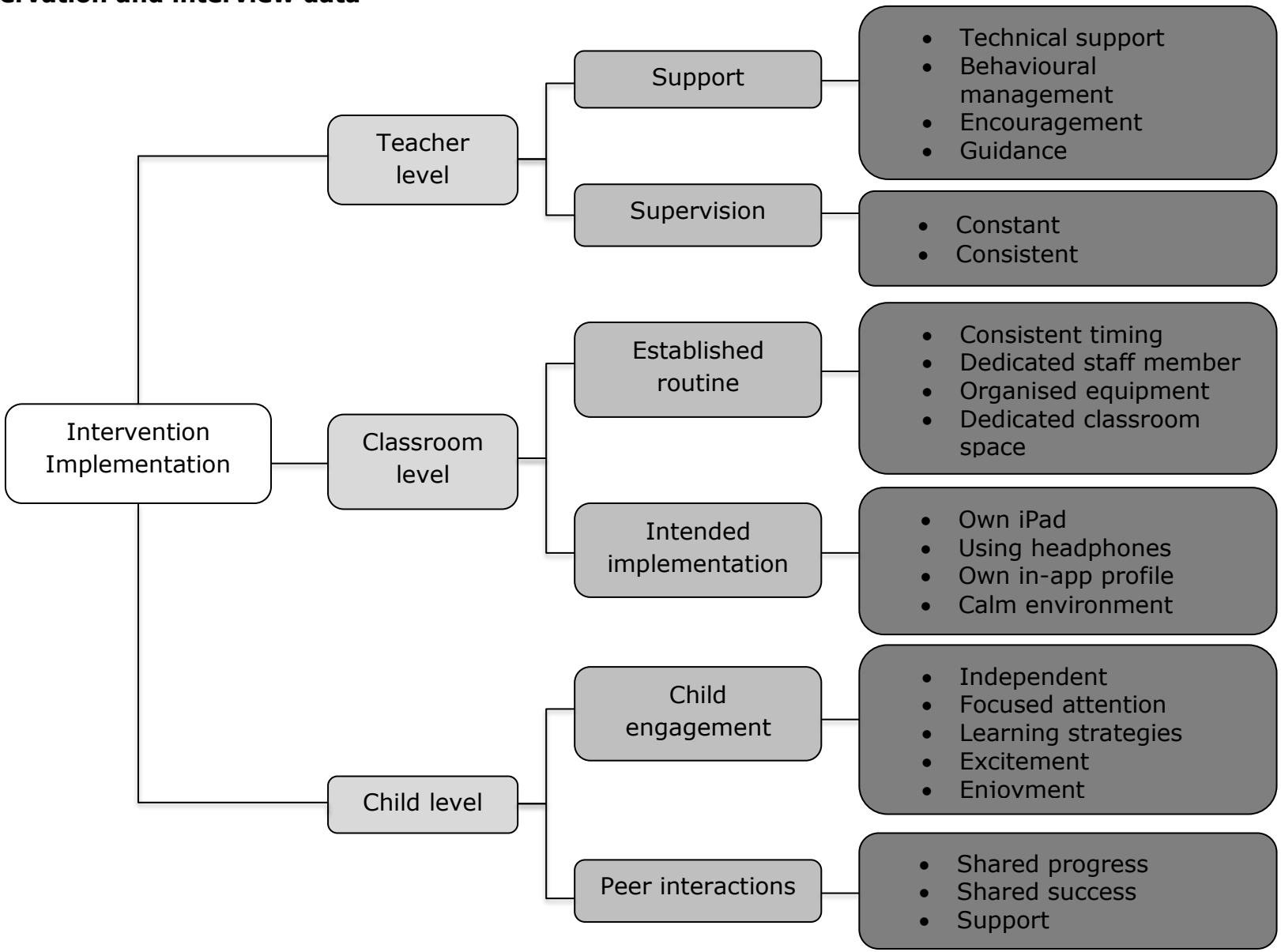
Stage	Description
Aim 1: Understanding Implementation	
Data Preparation	Original fieldwork notes photocopied and sorted securely. Original fieldwork notes anonymised (i.e. school name removed). Original fieldwork notes typed up and organised ready for analysis.
Data Familiarisation	The researcher read through the raw observation and interview datasets twice to gain a general and overarching sense of the data.
Data Coding	Inductive, bottom-up approach with open coding. A series of informal initial codes were established by looking at the datasets. Codes refined as researcher progressed through the data sets. Two cycles of data coding were conducted to ensure the data was saturated and data codes were exhaustive and consistent. A further cycle of coding was conducted to winnow data codes and ensure codes were exclusive.
Data Display	Thematic map generated incorporating the six implementation themes (see Figure 5).

Stage	Description
Aim 2: Relationship between Implementation Themes & Intervention Outcomes	
Interpretative Synthesis	Structured Judgement Methodology: Aggregated qualitative data for the six implementation themes presented as descriptive summaries for each school to convey individual experiences of implementing the maths app intervention.
Systematic Quantitative Framework	Structured Judgement Methodology: systematic quantitative framework designed and implemented (3-point Likert scale) to assess aggregated qualitative data. This step aimed to move data from individual description to general explanations (Clarke, 2004).

As outlined in Table 27, in order to address the first aim and understand the implementation of the maths app intervention across the final sample of 11 participating schools in the UK Feasibility RCT (Chapter 6), the qualitative data analysis aimed to generate a narrative account of the intervention implementation from the direct observations and self-report interviews. This was achieved by inductively identifying implementation themes from the observation and interview datasets using the same inductive, bottom-up approach (Creswell & Creswell, 2018).

When coding the observation and interview datasets, open coding was used to explore the combined data set and identify units of analysis (Cohen et al., 2007). New codes were generated until coding was complete; the dataset was saturated and all data was accounted for (Saldaña, 2015). To ensure the codes were exhaustive, exclusive, and consistent, data codes were refined through three rounds of coding (Creswell & Creswell, 2018; Saldaña, 2015). The coding process aimed to generate semantic themes by focusing on explicit and surface level meaning of the reported data (Cohen et al., 2007). Following the three rounds of coding, six implementation themes were identified across the teacher, classroom, and child levels. The six themes and associated definitions are displayed in the thematic map shown in Figure 5.

Figure 5 Thematic map summarizing the six implementation themes identified from the qualitative observation and interview data



Results & Discussion

To gain a rich, detailed qualitative understanding of the implementation of the maths app intervention across the final sample of 11 participating schools, this chapter employed narrative, direct observations and self-report interviews. Following a multi-level, ecological, determinant framework, the implementation of the maths app intervention was considered as part of an ecological system (Bronfenbrenner, 1992; Nilsen, 2015) with six themes identified at the teacher, classroom, and child level.

Teacher level

Two themes were identified at the teacher level; teacher support and teacher supervision. Teacher support encompassed four ways in which teachers supported children in using the maths apps. (1) Technical support included help in using the technology, for example, ensuring the headphones were correctly plugged into the iPad and were at a suitable volume. (2) Behavioural management included re-focusing children on the mathematics activity when they got distracted or restless. (3) Encouragement was provided to support children to persevere with a mathematics activity and/or in response to frustrated child behaviour. (4) Guidance included supporting children's understanding of the app instructions. Teacher supervision referred to the level of constant, consistent, and active presence of the teaching staff during the maths app intervention session.

Classroom level

Similarly, two themes were identified at the classroom level; established routine and intended implementation. Established routine included five actions which supported the maths app intervention to be successfully embedded into the daily classroom schedule. These actions included (1) implementing the maths app at a consistent time each day, (2) having a dedicated member of staff whose responsibility it was to implement the intervention, (3) having well organised equipment, such as colour coding the iPads so that children can easily identify which one is theirs and (4) having a dedicated space within the classroom (5) with a seating plan where children used the maths apps.

Intended implementation encompassed four ways in which the maths app intervention was implemented as intended, as outlined in the teacher manual (see Appendix E) and teacher training session (see Chapter 6). These included (1) children using their own iPad, (2) with headphones, (3) children accessing the maths app content within their own in-app profile (see Chapter 3) within (4) a calm classroom environment.

Child level

Two themes were also identified at the child level; child engagement and peer interactions. Child engagement included five ways in which children engaged with the maths apps. (1) Children were independent when using the maths apps. (2) Children presented focused attention when using the maths apps. However,

in some cases children became frustrated, restless, and unfocused. (3) Children utilised different learning strategies, such as one to one correspondence, counting aloud when completing some of the learning activities and repeating tasks until successful completion. (4) Children were excited to use the maths apps, particularly, when they completed an end of topic quiz and progressed to the next level. (5) Lastly, children were reported to enjoy using the maths apps.

Peer interactions encompassed three ways in which children interacted with other children while using the maths apps. (1) Some children shared their progress within the maths apps with their peers, for example, commenting on which level they are currently working on. (2) Similarly, some children shared their success with their peers and/or teachers, such as successfully completing one of the end of topic quizzes. (3). Some children provided support to their peers in using the maths apps, for example, help with understanding the questions and/or accessing the correct app on the main iPad interface.

Aim 2: Implementation themes and intervention outcomes

To examine the relationship between the implementation themes at the teacher, classroom, and child level and children's learning gains in response to the maths app intervention (reported in Chapter 6) the study adopted a structured judgement methodology (Clarke, 2004).

Methods

Structured judgment methodology

Central to structured judgment methodology is the problem that "the things we observe in the world, which we want to understand and explain, are the visible products of unseen processes" (Clarke, 2004, page 86). To address this problem, structured judgment methods aim to move from single, particular descriptive data to general explanations based on aggregated data. Disciplined and structured judgments on the aggregated qualitative data can be achieved by applying a systematic, quantitative framework (Clarke, 2004).

Procedure

Aggregated Qualitative Data.

To explore the relationship between the implementation of the maths app and learning outcomes it was necessary to gain further insight into the implementation conditions in each of the 11 participating schools. The individual schools' experiences were therefore reviewed against each of the six implementation themes.

In order to achieve this, the narrative, direct observation and self-report, interview data sources were assimilated into a descriptive body of data per school (see Table 27, interpretative synthesis). The descriptive summaries were structured around the six implementation themes. These descriptive summaries based on the aggregated observation and interview data aimed to convey the

individual schools' experiences of implementing the maths app intervention.

Systematic Quantitative Framework Design.

In order to make disciplined and structured judgements on the aggregated observation and interview data, a 3-point Likert rating scale of high (score 3), medium (score 2), and low (score 1) was developed. This rating scale was designed to assess the variability within the implementation themes across each school. In developing the 3-point Likert rating scale, necessary features for high (scored 3), medium (scored 2), and low (scored 1) ratings were identified based on the code definitions (see thematic map; Figure 5) and are outlined in detail in Table 28. As such, the 3-point Likert rating scale was relative across the final sample of 11 participating schools.

Systematic Quantitative Framework Implementation Procedure.

When assigning the 3-point Likert scale ratings the researcher read through all the descriptive summaries for each implementation theme in turn and allocated an initial rating. This procedure was repeated to ensure consistent and fair ratings across the whole sample. Furthermore, as the researcher collected and coded the original dataset, an additional independent researcher also followed the same rating procedure. A high inter-rater reliability was established with an average 90.6% agreement across the teacher, classroom, and child level themes (Miles & Huberman, 1994). The

additional researcher was a Masters level student with qualitative research experience and was independent from the research reported in this thesis.

Table 28 Definitions for high, medium, and low ratings with guiding questions for each of the implementation themes.

Implementation theme	Guiding question for ratings	High, medium, and low features
Teacher-level (max. 6)		
Teacher support	To what extent did teachers give *support* to children using the maths apps?	**High (scored 3):** 3-4 types of support were reported (e.g. behavioural management, technical support, and encouragement and guidance) and were consistent throughout the intervention session. **Medium (scored 2):** 2 types of support were reported, which were intermittent throughout the intervention session. **Low (scored 1):** 1(or less), type of support was reported and was intermittent throughout the intervention session.

Implementation theme	Guiding question for ratings	High, medium, and low features
Teacher supervision	To what extent did teachers *supervise* children using the maths apps?	**High (scored 3):** Constant supervision from consistent, dedicated member(s) of staff was reported throughout the intervention session. **Medium (scored 2):** Teaching staff were consistently present but not actively supervising children. Supervision from additional staff may have been reported, but this was not constant or consistent throughout the intervention session. **Low (scored 1):** Little (e.g. overarching eye from one teacher) to no supervision was reported.

Implementation theme	Guiding question for ratings	High, medium, and low features
Classroom-level (max. 6)		
Established routine	To what extent was a *daily routine* established?	**High (scored 3):** 4-5 actions for establishing a consistent, daily routine were reported (e.g. member of staff assigned responsibility of the intervention, dedicated space for children to use the apps with a seating plan, intervention timing consistent throughout the week, and well organised logistics). **Medium (scored 2):** 2-3 actions for establishing a daily routine were reported, but consistency may be varied. **Low (scored 1):** Little to no (1 or less) evidence that a consistent, daily routine was established.

Implementation theme	Guiding question for ratings	High, medium, and low features
Intended implementation	To what extent was the *intervention implemented as intended*?	**High (scored 3):** The maths app intervention was implemented as intended, as instructed by the teacher manual and teacher training session (See Chapter 6; Appendix E; e.g. children used their own iPad, children accessed app content in their own profile, correct equipment including headphones were used, and there was a calm classroom environment) and was consistent throughout the intervention session. All aspects of the definition must be present. **Medium (scored 2):** 2-3 of the implementation criteria were correctly followed, but were intermittent throughout the intervention session. **Low (scored 1):** Little to no (1 or less), implementation criteria were correctly followed.

Implementation theme	Guiding question for ratings	High, medium, and low features
Child-level (max. 3)		
Child engagement	To what extent did children *engage* with the maths apps?	**High (scored 3):** Children showed high levels of engagement with the maths apps throughout the intervention session (e.g. children worked independently with the maths apps, they were focused on the learning activity, they used learning strategies, and they showed excitement and engagement with the maths apps). In cases where there was mixed evidence, a low level of engagement was only reported in relation to one child. If mixed evidence was reported for more than one child, the 'medium' rating was allocated. **Medium (scored 2):** Mostly mixed evidence reported, including some evidence of low levels of engagement (e.g. children were reported to struggle with the apps, or became distracted and restless over the course of the intervention session).

Implementation theme	Guiding question for ratings	High, medium, and low features
		Low (scored 1): Low levels of engagement (e.g. children were reported to struggle with the apps, or became distracted and restless over the course of the intervention session) was consistently reported.
Peer interactions	To what extent did children *interact* with their peers while using the apps?	Peer interactions were not included in ratings because low peer interaction or high peer interaction may be both beneficial to implementation. The direction of relationship is not as clear as other themes.

Results

Preliminary analyses

Ofsted and SES.

A series of one way ANOVAs[9] showed no significant association between Ofsted ratings (see Table 24) and implementation rating scores at the teacher, $F(2, 10) = .738$, $p = .508$, classroom, $F(2, 10) = .633$, $p = .556$ or child level, $F(2, 10) = .189$, $p = .831$. Furthermore, independent samples t-tests showed no significant association between indicated school SES (see Table 24) and implementation rating scores at the teacher, $t(8) = .837$, $p = .427$, classroom, $t(8) = .357$, $p = .731$, or child-level, $t(8) = .980$, $p = .356$.

Implementation themes and learning gains

Within-group effect sizes (Cohen's *d*) were used as a measure of children's learning gains following the 12-week intervention period and were calculated for each treatment group (Group 1 and 2; see Chapter 6) across the final sample of 11 participating schools (see Table 29). Although there was a difference in overall exposure to learning mathematics (see Chapter 6), no significant difference was found between using the maths app intervention in addition to all regular mathematics instruction (Group 1: treatment) or instead of a daily mathematics activity (Group 2: time-equivalent; see

[9] Non-parametric tests, Kruskal-Wallis H tests (for one way ANOVAs) and Mann Whitney U tests (for independent samples t-tests) were conducted but showed the same pattern of results.

Chapter 6). Furthermore, there was a significant correlation between the learning gains observed in each group exposed to the maths apps, $r^s = .67$, $p = .024$.

Therefore, the two maths app groups (Group 1 and 2) were collapsed and within-group effect sizes (Cohen's *d*) measuring the magnitude of progress from pre- to post-test were calculated per school (see Table 29). As illustrated in Table 29, the average within-group effect size of progress across the final sample of 11 schools, was, Cohen's $d = .77$ (SD = .36, minimum- maximum range = .13 – 1.35).

Table 29 Within-group effect sizes of progress with the maths apps collapsed across Group 1 (treatment) and Group 2 (time-equivalent treatment)

School	Mean (SD)		
	Pre-test	Post-test	Cohen's *d*
Whole sample	11.77 (4.52)	15.10 (4.48)	.77 (.36)
A	13.00 (5.03)	17.59 (4.03)	1.01
B	15.14 (4.48)	15.73 (4.77)	.13
C	12.15 (3.56)	15.85 (3.39)	1.06
D	9.60 (4.26)	13.40 (5.72)	.75
E	10.69 (4.19)	16.44 (4.31)	1.35
F	10.53 (5.48)	14.45 (5.03)	.75
H	12.25 (5.80)	15.38 (5.27)	.57
I	9.74 (3.47)	12.87 (3.62)	.88
J	11.62 (4.07)	14.76 (4.51)	.73
K	13.16 (3.94)	16.89 (3.76)	.97
L	11.58 (5.39)	12.79 (4.87)	.24

Table 30 reports the mean ratings for the implementation themes at the teacher, classroom, and child level across the final sample of 11 participating schools. To examine the relationship between the implementation themes and children's learning gains in response to the maths app intervention, Spearmans' Rho correlations were conducted (see Table 30; Pitchford et al., 2018). Results showed a medium, positive, relationship between classroom level implementation and children's average learning gains with the maths apps. However, this relationship did not reach statistical significance (2-tailed *p* value = .08), which was most likely due a relative lack of power associated with the small sample size of schools (*n* = 11). No other notable relationships were observed between learning gains and teacher and child level implementation themes.

Table 30 Group mean rating (SD, Min-Max) for each implementation theme at the teacher, classroom, and child level. Spearmans' rho correlations between implementation themes and learning gains.

Variable (max score)	Mean Rating (min-max)	Correlation with Gains
Teacher level	4.18 (1.62)	.08
(max. 6)	2.00 – 6.00	
Classroom level	4.82 (1.22)	.55
(max. 6)	2.00 – 6.00	
Established routine	2.27 (.79)	**.73***
(max. 3)	1.00 – 3.00	
Intended implementation	2.55 (.69)	.02
(max. 3)	1.00 – 3.00	
Child level	2.27 (.46)	.00
(max. 3)	2.00 - 3.00	

Note. *p = .011

Despite the lack of power, the correlation between implementation rating scores at the classroom level and children's learning gains is sizable (see Table 30) and warrants further exploration as to which component of classroom level implementation was most predictive of learning gains.

As shown in Table 31, further exploratory spearman's rho correlations showed a strong, positive, significant correlation between established routine and learning gains. No significant correlation was observed for intended implementation and learning gains. When entered into an exploratory linear regression (see Table 31) established routine significantly accounted for 41% of the observed variance in children's learning gains.

Table 31 Model fits for the exploratory linear regression identifying significant predictors of learning gains.

Variables	Model		Sig.	Change		Unstd. Coefficients	Std. Coefficients	Sig.
	R	R^2	*F* (df), *p*	ΔR^2	Sig. ΔF	B, Std. Error	Beta	*t, p*
Established routine	.636	.405	6.13 (1, 9) p = .035	.405	p= .035	.29, .12	.64	2.48, p = .035

Discussion

To examine the relationship between the identified implementation themes and children's learning gains with the maths apps, this chapter employed a structured judgements approach. The qualitative data from the narrative, direct observations and self-report interviews was aggregated in light of the six identified implementation themes (see Aim 1) and this conveyed the individual maths app intervention implementation experiences in the final sample of 11 participating schools. A 3-point Likert scale, systematic quantitative framework was then developed and utilised in order to make structured judgements about the aggregated qualitative data for each school. These ratings were then correlated with children's learning gains with the maths apps.

Classroom level

Results showed a moderate, positive, but not statistically significant relationship between the classroom level implementation themes and learning gains. The lack of statistical significance was most likely due to a lack of power based on the small sample size (n = 11). Despite the lack of statistical significance, the size and direction of the correlation was of interest. When this relationship was explored further, it was found that scores on established routine was the most predictive of learning gains with the maths apps, showing a strong, positive and significant correlation. When entered into an exploratory linear regression model, established routine scores accounted for 41% of the observed variance in

learning gains. This data suggests that schools that had a well-established daily implementation routine made the most progress in mathematics over time with the maths apps intervention. This finding supports the assertion that how technology is integrated into the school environment is critical to its success (Beach & O'Brien, 2015; Couse & Chen, 2010) and supports the value of examining implementation with mixed-methods in educational research more broadly (Connolly et al., 2018).

In explaining these results, the preliminary analyses showed no relationship between Ofsted ratings, a proxy measure for school quality (Gambaro et al., 2015; Schagen et al., 2005) and the school, classroom, and child-level implementation themes. Similarly, preliminary analyses showed no association between the schools' indicated SES and the implementation themes. As such, the relationship between classroom-level, established routine and learning gains with the maths app intervention could be explained by other factors. For example, research suggests teacher's perceptions of educational technology are associated with the positive uptake and integration of technology in the classroom (Blackwell et al., 2014). Teacher's technology based perceptions and belief systems would form part of the macrosystem of the multi-level, ecological, determinant framework outlined in Chapter 1.

Teacher and child level

In contrast, no statistically significant relationships were observed between children's learning gains and the implementation themes at the teacher and child level. At the teacher level, this result may be expected due to the child-centred nature of the maths app software. In particular, the maths apps are well-designed (see Chapter 3) and are age-appropriate (Kucirkova, 2014), allowing children to access the learning activities without a specific need for direct teacher support or supervision. This finding supports the assertion that app technology can provide effective maths instruction without additional, time-consuming teaching demands (Hilton, 2016; Kucian et al., 2011).

However, the group level ratings for child engagement may not be sufficiently sensitive in order to understand the association between child engagement and learning gains with the maths apps. In the current study, teachers frequently reported high child engagement, but also reported restless, frustrated, and distracted behaviour within their class. Based on the systematic quantitative framework this mixed report was given a medium rating (see Table 28). Observations of the descriptive data showed a medium rating was the most common rating given for this implementation theme (mode = 2.00). Consequently, individual child engagement was not effectively captured with the current group-level data. This lack variation renders it difficult to detect a relationship between the variables of interest. Nevertheless, the exploratory nature of this

study has identified five important aspects of child engagement to be considered in future research.

Future directions

While this mixed-methods study adds an important first step, insight into implementation process evaluation in app-based interventions, future research on this topic should consider three issues. Firstly, in order to elucidate the relationship between established routine and learning gains observed in the current study, future studies could expand the holistic evaluation to include teacher's perceptions (e.g. Blackwell et al., 2014). Future studies could examine a potential association between teacher's perceptions of the maths app intervention and successful implementation, and if this is related to children's learning gains with the maths apps.

Secondly, future studies could examine child engagement at the individual level. For example, the researcher could conduct direct, structured observations of individual children using the maths app intervention focused on the five areas of child engagement identified in the current study. Alternatively, depending on the practical and time constraints of the particular study (Clarke, 2004), a teacher based questionnaire assessing the five areas of child engagement could be utilised. This measurement of individual child engagement could help illuminate its association with the maths app related learning gains and understand more about the universal reach and application of the maths app intervention to maximise the development of children's early mathematical skills.

Finally, the momentary nature of the narrative, direct observation and self-report interview methods used in this study afford a time limited understanding of user's experience of the maths apps. Some fluctuations in daily practice are anticipated, which poses a threat to the internal validity of the current findings. Although the assimilation of two data sources can help address this issue, it is important to consider this caveat when interpreting the current findings. Furthermore, gaining an understanding of user experience is distinct from treatment integrity measures that formally assess implementation fidelity (Humphrey et al., 2016). As such, future IPE studies should consider expanding the definition of implementation and examine how these different factors may influence the integration and success of educational interventions.

Conclusion

This chapter reports a systematic and pragmatic evaluation of the implementation of the maths app intervention reported in Chapter 6 (UK Feasibility RCT). Firstly, this chapter identified six implementation themes; support and supervision at the teacher level, established routine and intended implementation at the classroom level, and child engagement and peer interactions at the child level. These identified themes illuminated how the maths apps intervention was implemented in terms of quality, adaptations, and participant responsiveness.

Secondly, this chapter examined the relationship between the six implementation themes and children's learning gains with the

maths apps. Results identified a moderate and positive association between the classroom level theme 3-point Likert scale ratings and learning outcomes with the maths apps. Further exploration of this relationship showed established routine accounted for 41% of the observed variance in children's mathematical learning gains. In contrast, no association was found for teacher level support and supervision, as well as how well the intervention was implemented as intended and child engagement ratings. This suggests having an establish maths app intervention routine, specifically implementing the maths app at a consistent time each day, having a dedicated member of staff assigned to the intervention, having well organised equipment, such as colour coding the iPads and having a dedicated space within the classroom with a seating plan where children used the maths apps was critical for maximising children's learning outcomes.

Chapter 8: General Discussion

This thesis reports an in-depth evaluation of the reach and utility of a new maths app intervention for supporting the acquisition of early mathematical skills for young children at the start of formal education in the UK and Brazil.

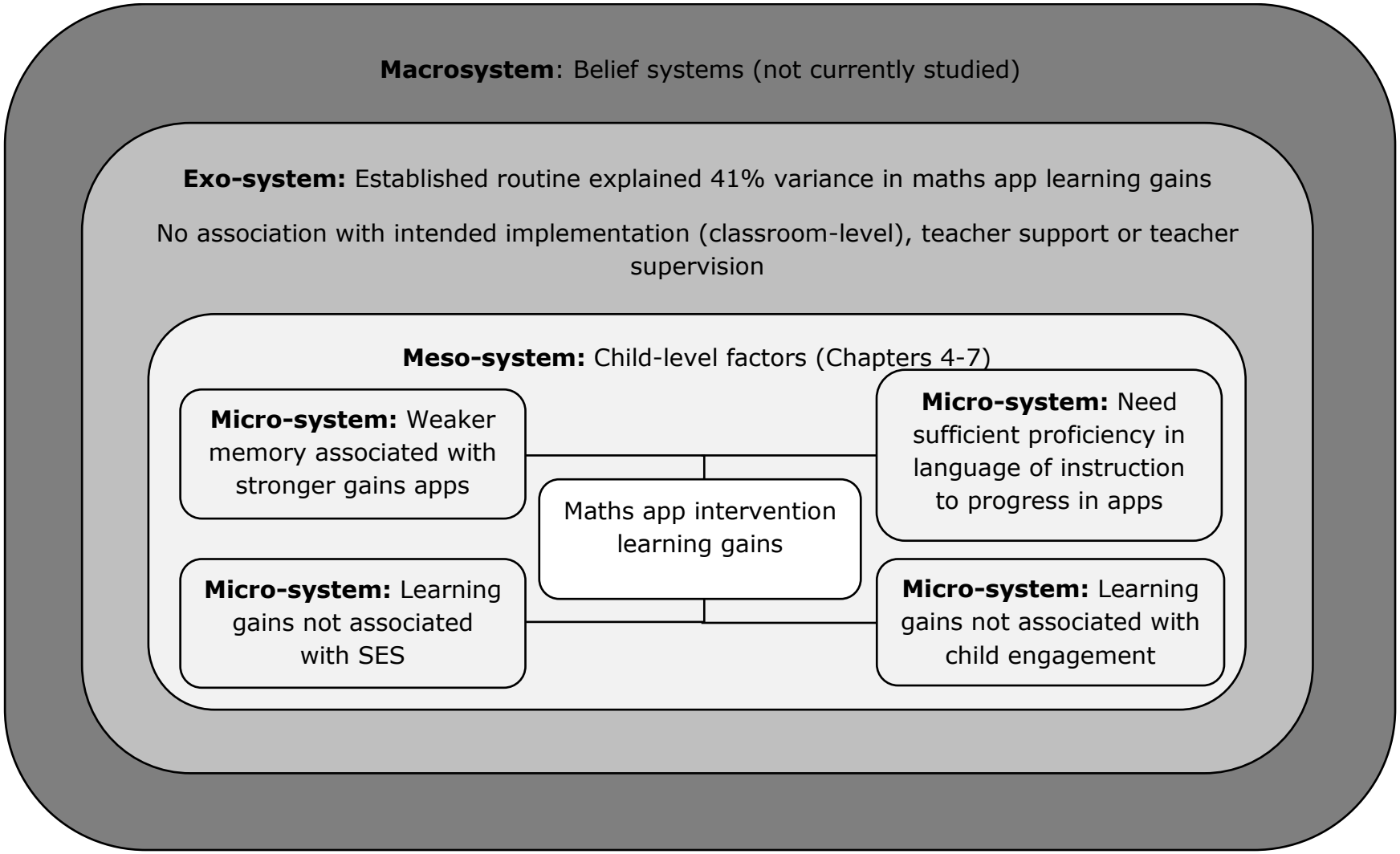
The systematic review reported in Chapter 2 identified seven eligible, previous studies that examined app-based mathematical interventions in the first three years of primary school. Results showed children who used app-based mathematical instruction made significantly stronger learning gains in comparison to a range of comparison groups, including standard practice, other educational apps, and other mathematical interventions. Increased time learning mathematics with this technology was associated with increased learning gains. However, no previous research considered factors at the teacher, classroom, and child level that may influence learning outcomes with app-based mathematics instruction.

Building on this emerging evidence base, this thesis reported a series of studies that utilised quantitative and qualitative methods and were conducted in the UK and Brazil to evaluate a new maths app intervention (see Chapter 3). The experimental research reported in Chapters 4-7 adopted a multi-level, ecological, determinant systems approach and integrated educational neuroscience and implementation science to examine a range of potential factors at the teacher, classroom, and child level that may influence children's learning outcomes with the maths app

intervention (see Figure 6). This mixed-methods approach aimed to address whether the maths apps are an effective form of mathematics instruction? Under what circumstances do the maths apps work? How do the apps work? And whom do they work for?

This chapter will critically address each of these four questions, integrating important theoretical, practical and policy implications. The quality of the quantitative data will be formally assessed and compared using the same assessment tool reported in Chapter 2 (Systematic Review). An impact evaluation of the UK Feasibility Trial will then be presented, followed by a discussion focused on the limitations of the current studies and directions for future research.

Figure 6 Summary of study results within the multi-level, ecological, determinant framework



Are the maths apps an effective form of mathematics instruction?

Consistent with the previous research presented in Chapter 2 (Systematic Review), across the three experimental, quantitative studies (Chapters 4-6), this research found that combining child-centred, curriculum based mathematics apps with interactive touch-screen tablet technology for children at the start of formal education in the UK and Brazil provided an effective means of delivering quality instruction that promoted the development of early mathematical skills. Table 32 summarises the large and significant learning gains in mathematics, including effect sizes, following the maths app intervention.

Time on task

To explore the relationship between app-based mathematics instruction learning gains and time on task, the partial correlation analysis reported in Chapter 2 (Systematic Review) was repeated to include the within-group effect sizes reported in this study. Results also showed a small, positive correlation between increased time on task and increased learning gains when controlling for the number of children receiving the app-based mathematics instruction, r (9) = .39, p = .236. While the study sample is still underpowered, the strength and direction of this exploratory correlation is consistent with previous reviews (Cheung & Slavin, 2013; see Chapter 2) and provides further support that time on task impacts learning gains.

Table 32 Summary of methodologies used and findings from each of the three quantitative studies

Study Summary	UK Proof of Concept Chapter 4	Brazil Proof of Concept Chapter 5	UK Feasibility RCT Chapter 6
Design	Between-groups (time-equivalent treatment and control) quasi-experimental design	Between-groups (Brazilian Portuguese, English, and control) quasi-experimental design	Pupil-level randomised control trial with three groups (treatment, time- equivalent treatment, and control)
Participants	27 children aged 4-5 years Including 12 low-achievers and 15 typically attaining children	61 Brazilian Portuguese-English bilingual children aged 5-6 years	389 mixed ability children aged 4-5 years
Implementation	Teaching staff in small groups instead of one daily mathematics practice activity	Teaching staff in small groups instead of typical mathematics instruction in Brazilian Portuguese or English	Teaching staff in small groups either in additional to all standard mathematical practice (treatment) or instead of one daily

Study Summary	UK Proof of Concept Chapter 4	Brazil Proof of Concept Chapter 5	UK Feasibility RCT Chapter 6
			mathematics practice activity (time-equivalent treatment)
Duration	30 minutes a day, 5 days a week for 8 weeks over a 16-week intervention period	20 minutes a day, 4 times a week for 10 weeks	30 minutes a day, 5 times a week for 12 weeks
Time on task	1200 minutes/ 20 hours	800 minutes/ 13.3 hours	1800 minutes/ 30 hours
Assessment	Paper and pen curriculum knowledge (CK) test and Mathematical Reasoning (MR; from WIAT-II)	Early Grade Mathematics Assessment (EGMA)	Progress Test in Maths Level 5
Maths App % Gain	CK = 28.8% MR = 8.1% Composite = 18.5%	Brazilian Portuguese = 16.9% English = 18.1%	Treatment = 13.4% Time-equivalent treatment = 12.2%

Study Summary	UK Proof of Concept Chapter 4	Brazil Proof of Concept Chapter 5	UK Feasibility RCT Chapter 6
Maths app within-group effect sizes (Cohen's *d*)	CK = 3.34 MR = 2.81 Composite = 3.75	Brazilian Portuguese = 1.46 English = 1.06	Treatment = .78 Time-equivalent treatment = .65
Between-group effect sizes (Cohen's *d*)	**Maths app** v Control (composite) = 1.54	**Brazilian Portuguese** v English = .13	**Maths app** v Control ≥.21 No significant difference between two maths app groups

Note. Bolded text highlights direction of effect.

Under what circumstances do the maths apps work?

In order to understand the implementation circumstances in which the maths apps work, it is important to compare different forms of intervention implementation and gain qualitative insights into how the intervention was implemented in individual participating schools.

Comparing forms of implementation

The UK Feasibility RCT (Chapter 6) builds on the emerging evidence base, which has previously implemented app-based mathematical instruction in place of a different control condition; either standard practice (Pitchford, 2015; Schacter & Jo, 2016; van der Ven et al., 2017), small group, non-technology based mathematics activities (Schacter & Jo, 2017), or other app-based instruction (Park et al., 2016; Pitchford, 2015; Schacter & Jo, 2015; Szakudlarek & Brannon, 2018). In contrast, the UK Feasibility RCT directly evaluated two forms of maths apps implementation; as a supplementary intervention (Group 1: treatment) and instead of one daily mathematics activity (Group 2: time-equivalent treatment). This was the first maths app intervention study to adopt this novel and rigorous research design.

Results showed there was no significant difference between implementing the maths apps as well as all standard mathematics practices (RCT Group 1) or instead of one regular small group, teacher-led mathematics activity (RCT Group 2). This evidence suggests the maths apps are a form of quality mathematics

instruction and is consistent with the results observed in the UK Proof of Concept study (Chapter 4), which included time-equivalent treatment and control groups.

Insights into implementation

The UK RCT Implementation Evaluation (see Chapter 7) identified six themes at the teacher, classroom, and child level as part of the multi-level, ecological, determinant framework (see Chapter 1). As highlighted in Figure 6, establishing a structured routine for implementing the maths apps was the most influential on children's learning gains with the maths apps; accounting for 41% of the observed variance. No other implementation themes were shown to be associated with learning outcomes. As discussed in Chapter 7, this suggests the high-quality, curriculum based maths apps are able to support a child-centred learning environment that does not place additional demands on teaching staff.

Educational policy and practice implications

Collectively, these results have important implications for educational policy and practice. The results suggest an increased focus on early mathematics at the class level (APPG, 2014; Ofsted, 2018; Pitchford et al., 2016) can take the form of efficient instructional practices without the need for extra time learning mathematics, which could potentially detract from other areas of a well-rounded curriculum. This is critical as critiques of educational policy have argued that the early years curriculum introduces too much time spent on formal instruction, reducing opportunities for

play-based, exploratory learning (Roberts-Holmes, 2015). This research, particularly the UK Feasibility RCT (Chapter 6) demonstrates the maths apps were an efficient form of mathematics instruction and shows the apps can be implemented as part of a well-balanced curriculum. It indicates the apps can capitalise on the benefits of play-based learning by introducing children to mathematical concepts in a game like format, which is interactive, engaging, and child-centred, but also draws on the benefits of direct instruction and targeted practice, helping to bridge the gap between formal and informal early mathematical learning opportunities (Evangelou, Sylva, Kyriacou, Wild, & Glenny, 2009).

In order to maximise children's learning gains with the maths apps, teachers should establish a strong classroom routine, consisting of a dedicated member of staff assigned the responsibility of implementing the intervention, a dedicated space with a seating plan where children are to use the maths apps, a consistent daily time in which the intervention is implemented, and well organised equipment, such as colour coded touch-screen device hardware.

How do the maths apps work?

In order to understand how the maths apps work it is important to consider mathematical development, factors known to be associated with mathematical development, such as working memory and language proficiency, as well as, Science of Learning principles and theoretical implications.

Mathematical development

As discussed in Chapter 1, it is vital to consider how the maths app intervention may support different components of mathematical development. Similar to previous studies (see Chapter 2) the Proof of Concept studies reported in Chapters 4 (UK) and 5 (Brazil) primarily measured basic mathematical skills with a researcher developed measure of Curriculum Knowledge and the Early Grade Maths Assessment (EGMA; Brombacher, 2011), respectively.

Although the Mathematical Reasoning sub-test from the WIAT-II (Wechsler, 2005) was used in the UK Proof of Concept study, it includes items covering a range of different components of mathematical development. For example, early items from this assessment require the child to identify geometric shapes and count single digit numbers, encompassing basic factual knowledge. The assessment progresses to single and multi-step word problems, reflective of higher order mathematical reasoning and problem solving skills. This combined measure renders it difficult to disentangle particular benefits of the maths app intervention to basic and higher order mathematical skills, and thus it is difficult to understand specifically how the app-based intervention works to support early mathematical development.

To address this, the Progress Test in Maths Level 5 (PTM5, Maths Assessment Resource Service, 2015) was used in the UK Feasibility RCT reported in Chapter 6. This assessment is specifically

designed to assess the four components of mathematical proficiency and thus affords a closer examination of which components of mathematical development are supported by the maths app intervention. Although results showed no significant Time X Group interaction for any of the four components of mathematical proficiency (Fluency in Facts, Fluency in Concepts, Mathematical Reasoning, and Problem Solving), larger effect sizes were consistently observed for children who used the maths apps (Group 1 and Group 2) in comparison to children who received standard teacher-led practice (Group 3) across all four mathematical components.

This evidence suggests the maths apps may work by supporting learning in the areas of mathematical development targeted by the intervention (Fluency in Facts and Fluency in Concepts; see Table 8, Chapter 3) that generalises to other higher-level skills not extensively or explicitly covered in the maths app content (Mathematical Reasoning and Problem Solving). Based on cumulative learning theory (Gagné, 1968), this evidence indicates the maths apps can support the mastery of basic mathematical skills, which enables mathematics facts and simple concepts to become automated and assimilated into a higher-level conceptualisation of mathematical knowledge (Codding et al., 2010; Gersten & Chard, 1999; Karmiloff-Smith, 1994; Mayfield & Chase, 2002; Van Der Heyden & Burns, 2005). These results are consonant with previous intervention studies that have also targeted basic mathematical skills acquisition and demonstrated evidence of

generalisation to higher-level mathematical skills (Fuchs et al., 2009; Kidd et al., 2013; van der Ven et al., 2017).

Factors that influence mathematical development

Furthermore, as highlighted in Chapter 1, some children may be vulnerable to underachievement in mathematics due to a range of factors known to be associated with early mathematical development, including low memory capacity and low levels of language proficiency in the language of instruction in the context of bilingual learners.

Memory.

As highlighted in Figure 6, results from the UK Proof of Concept study (Chapter 4) observed a small, negative relationship between short-term memory capacity and learning gains following exposure to the maths apps. Specifically, children with a weaker memory capacity made greater progress in mathematics. Although this relationship was not found to be significant in this study, most likely due to an underpowered sample, it is consistent with previous research with the same maths app intervention with older low-achieving pupils (Outhwaite et al., 2017).

This result may be attributed to some of the learning strategies built into the app, including clear, specified learning goals with simple, stepwise, repeatable task instructions, the ability to repeat learning activities as often as required, and the inclusion of interactive and multi-sensory mathematical manipulatives (Gathercole & Alloway, 2007; Rycroft-Smith et al., 2017). It is

possible that these features serve to reduce the demands on children's limited memory capacity and thus aid children's access to the mathematical learning opportunities within the maths app intervention. However, without direct manipulation and/or observation of children's interactions with these in-app features hypothesised to reduce memory demands, these suggestions remain speculative and warrant further research.

Language proficiency.

Results from the Brazil Proof of Concept study (Chapter 5) showed children who used the maths apps made significantly stronger learning gains compared to the standard practice control group regardless of whether the app was delivered in the child's first or second language. This is consistent with previous research with the same maths app intervention used with EAL children in the UK, which showed children identified as EAL through teacher reports made just as much progress with the maths apps delivered in English compared to their non-EAL peers (Study 1, Outhwaite et al., 2017).

However, the measure of language proficiency used in the Brazil Proof of Concept study affords a closer examination of the relationship between proficiency in the language of instruction and mathematical progress, beyond the dichotomous teacher EAL status reports. Results showed a moderate, positive, and significant correlation between proficiency in the language of instruction and progress through the maths apps (i.e. number of topics completed).

Specifically, children with a higher language proficiency made greater progress through the maths app intervention (see Figure 6).

This finding is consistent with the Threshold Hypothesis (Cummins, 1976), which suggests that sufficiently developed language proficiency particularly the child's CALP, is needed to enable the child to access curriculum content and to respond to instruction (Cummins, 2008). Whist the maths apps employ simple language and were identified to provide relatively high levels of contextual support, through the congruent auditory and visual information, interactive pictures, audios, and animations, combined with an on-screen teacher providing clear task demonstrations (see Chapter 3 and 5), this result could be explained in two ways. Either, (a) the vocabulary included in the maths app may have extended beyond children's CALP and/or (b) the context embedded communication in the apps falls below the level typically present in successful bilingual classroom instruction practice.

Firstly, some of the tasks in the maths apps require vocabulary that may extend beyond the children's current level of proficiency, specifically their CALP in the language of instruction. For example, in one quiz question, the on-screen teacher asks the child "how many pencils are there?" from an array of pencils, footballs, cups, bowls, and shoes. If the child cannot identify which items are pencils, their performance may reflect their low language proficiency (i.e. CALP) rather than their mathematical ability. As progress through the apps was measured by number of topics completed

(100% pass rate needed in the end of topic quiz), these vocabulary restrictions may have hindered children's progress. Language of instruction difficulties may also have been evident as children progress further through the maths apps at their own pace. More specifically, as highlighted in Table 17 in Chapter 5, many of the later maths app topics were not included in the standard Kindergarten curriculum followed by the school. As progression through the maths apps was self-paced children may have been exposed to new language beyond their current knowledge, for example, vocabulary associated with Shape and Position (Topic 1, Maths 4-6), which again may have limited the overall number of topics completed.

Secondly, the maths apps may not provide sufficiently context embedded communication, typically associated with bilingual instruction. Supportive cues, such as gestures, body language, and intonation allow the learner to draw upon their basic interpersonal communication skills (BICS) and are ubiquitous in classroom instructional practice. In standard mathematics practice these cues contribute significantly to the understanding of portrayed information in a bilingual context (Church, Ayman-Nolley, & Mahootlan, 2004; Singer & Goldin-Meadow, 2005) and may support understanding when children have not reached a sufficient threshold in their CALP. However, in order to explore these proposed explanations further, direct measures of children's L1 and L2 CALP specific vocabulary skills are needed in future research. Without these additional measures these explanations are speculative.

Science of Learning principles

Furthermore, several features of the maths apps may count for their success and underpin how these maths apps work to support early mathematical development. As highlighted in Chapter 3, the maths apps include features consistent with the principles of active (e.g. multi-sensory and direct interactions), engaged (e.g. feedback), meaningful (e.g. a staged and scaffolded curriculum), and socially interactive learning (e.g. through the on-screen teacher; Hirsh-Pasek et al., 2015). The maths apps also include high-quality curriculum based content (Blum-Ross & Livingstone, 2016; Falloon, 2013) and specific learning goals (Hirsh-Pasek et al., 2015). Additionally, the maths apps reflect a blended learning approach (Naismith et al., 2004) drawing on different instructional psychology principles, namely direct instruction through, stepped learning, feedback, repetition, and rewards (Kirschner et al., 2006) and play-based pedagogies through the opportunity for self-regulation and learner control (Gray, 2015). The combination of these app design features may account for the observed learning gains seen in the experimental intervention studies reported in Chapters 4-6.

In contrast, the standard mathematics practice seen in the UK and Brazil typically followed teacher-led or enquiry-based pedagogical approaches with a strong emphasis on learning through self-directed play and peer collaboration. For example, in the UK, the small group mathematics instruction typically consisted of teacher-directed mathematical activities and the whole class

mathematics activities were more child-directed embedding early mathematical concepts into play-based learning experiences (see Chapter 4 and 6). In Brazil, standard mathematical practice was organised around themes, which were integrated across curriculum subjects. Similar to the UK, standard practice in the private bilingual school in Brazil placed a strong emphasis on enquiry based learning through play.

While these pedagogical approaches are highly valued, particularly in the first years of formal schooling (Roberts-Holmes, 2015), critiques have argued that such constructivist approaches should be combined with cognitive activity, instructional guidance, and curricular focus (Mayer, 2004). Furthermore, curricula that contain an element of individualised and direct instruction have been shown to be highly effective in securing the acquisition of early mathematical skills (Fuchs et al., 2009; Kidd et al., 2013; van der Ven et al., 2017). The experimental intervention studies reported in Chapters 4-6 indicate that multi-sensory, individualised, direct instruction, in the form of the maths apps can be successfully embedded into the classroom routine, with significant and positive benefits for young children's mathematics achievement.

Theoretical implications

Interestingly, the maths app learning gains found in Chapters 4-6 (see Table 32) were consistent across the different educational and cultural contexts in the UK and Brazil. Furthermore, the observed learning gains were also comparable with the learning

gains found in Pitchford's (2015) study evaluating the same maths app intervention in Malawi (see Chapter 2), a low-income country with a history of poor child development (Hubber et al., 2016) and extremely low basic mathematical skills (Milner, Mulera, Banda, Matale, & Chimombo, 2011). Children in Malawi used the same maths apps, with instructions delivered in their local language for up to 1200 minutes over an 8-week intervention period. Although interventions are not typically considered universally effective (Fuchs et al., 2009), the studies reported in this thesis show comparable learning gains with the same maths apps in three radically different educational contexts. This suggests the Science of Learning principles outlined above (see Chapter 1 & 3) may be a universal mechanism underpinning learning with high-quality educational app content that transcends culture.

In sum, the interactive, multi-sensory maths apps that draw on Science of Learning, direct instruction, and play-based pedagogical principles may work by providing targeted and individualised practice in basic mathematical skills with appropriate learning strategies for a child's low memory capacity. Consistent with cumulative learning theory, this mastery of basic mathematical skills provides a strong foundation for supporting the development of higher order mathematical skills. However, in order to maximise the learning gains with the maths apps, children need to have a sufficiently developed language proficiency, particularly their CALP in order to fully access the app-based mathematical learning opportunities.

Who do the maths apps work for?

To understand who may benefit the most from the maths apps, the studies reported in this thesis further considered the relationship between a child's socio-economic status and learning gains with the maths apps and whether the intervention was particularly beneficial for children identified as low-achievers in mathematics.

Socio-economic status

Results from the UK Proof of Concept study (Chapter 4) showed a small, negative but not statistically significant correlation between children's SES and learning gains in mathematics (indicated by a difference composite mathematics score; see Figure 6). A similarly weak correlation was also observed between SES and learning gains following standard mathematical practice.

On the one hand, this evidence may suggest SES was not associated with children's mathematical progress with their allocated mathematics instruction. However, this study was conducted in one school located in a middle-class SES area of Nottingham. This may minimise the variation in children's SES backgrounds and thus limit the ability to observe a relationship between SES and mathematical progress. Nevertheless, the lack of association between SES and learning gains was also observed in previous research in lower SES areas of Nottingham and Doncaster (Outhwaite et al., 2017). Furthermore, while individual SES data was not available for all participating children in the UK Feasibility

RCT (Chapter 6), no relationship between indicated school level SES and children's learning gains with the maths apps was observed[10]. Collectively, this evidence suggests all children benefited from the maths app intervention, regardless of their SES background. This corroborates the cross-country findings where similar learning gains (see Table 32) were found in countries that differ in income status; the UK, a high-income country (see Chapter 4) and Brazil, a middle-income country (see Chapter 5).

Low-achievers

Results from the UK Proof of Concept study (Chapter 4) showed children identified by their teachers as low-achieving in mathematics made significant learning gains compared to their higher-achieving peers receiving standard practice. Furthermore, the significant gap in achievement that was evident at pre-test closed following the 16-week intervention period. At post-test, the mean curriculum knowledge performance of the higher-achieving control group was only 1.8% higher than that of the low-achieving treatment group.

These results are consistent with previous research examining the benefits of the same maths app intervention specifically for low-achieving children (Outhwaite et al., 2017). However, in this study the treatment and control groups were split by mathematical ability rendering it difficult to address whether the

[10] Independent samples test showed no significant differences between school level indicative SES ratings (above or below national average) and within-group effect sizes of progress with the maths apps (see Chapter 7), t (8) = .175, p = .865.

maths apps were particularly beneficial for low-achieving children compared to standard practice.

This question can be explored, with caution in the UK Feasibility RCT (Chapter 6). While the UK Feasibility RCT was designed to evaluate the reach of the maths app intervention at the whole class level, similar to Szkudlarek and Brannon (2018; see Chapter 2), a sub-sample of children can be statistically identified as 'low-achievers' based on their pre-test mathematics score; ≥1.5 SD below the overall group mean (Snowling, 2013). As shown in Table 33, re-analysis of the data specific to this sub-sample (n = 31) showed much stronger learning gains for low-achieving children who used the maths apps (Group 1 & 2) relative to their low-achieving peers who received standard practice only (Group 3). The strongest learning gains were seen for low achieving children who used the maths apps as well as all standard mathematics practice (Group 1). This observation raises the possibility that the maths app intervention may be particularly beneficial for low-achieving young children, who may require supplementary mathematics instruction.

Table 33 Re-analysis of 'low-achievers' sub-sample in the UK Feasibility RCT (Chapter 6)

Outcome Variable	Group 1 (treatment; *n* = 10)	Group 2 (time-equivalent treatment; *n* = 11)	Group 3 (control; *n* = 10)
Total Raw Score (max. 26)			
Pre-test Mean (SD)	3.00 (2.00)	3.27 (1.79)	3.90 (1.37)
Post-test Mean (SD)	11.40 (2.17)	8.18 (3.34)	7.20 (3.46)
Gain Score Mean (SD)	8.40 (2.59)	4.91 (3.65)	3.30 (3.37)
% Gain	32.3	18.9	12.7
Cohen's *d*	4.03	1.83	1.25
(95% CI)	(1.87 – 6.18)	(.42 – 3.24)	(-.10 – 2.61)

However, the security of these observations is relatively weak and should be considered with caution as the results are based on small sample of low-achievers (*n* = 31 across all three arms of the RCT) and so are vulnerable to the threats to internal validity that are associated with identifying low-achievers on baseline performance (Barnett et al., 2004; see Chapter 2). Nevertheless, these additional findings are indicative for future research, including a current large scale, EEF funded, cluster RCT efficacy trial

evaluating the maths app intervention as a targeted intervention for low-achieving children (identified through teacher assessments) aged 5-6 years old across 113 schools in the UK.

Collectively, the UK research reported in this study suggests the learning gains associated with the maths apps were not associated with the child's SES background, indicating the wide reach of the intervention. Furthermore, the maths apps may be particularly beneficial for children identified as struggling with mathematics and can help bridge the divide between higher- and low-achieving young children (Burns et al., 2010; Tan & Tan, 2015).

Educational policy and practice implications

These findings have important implications for educational policy and practice. Traditionally mathematical interventions are implemented with children most at risk, either based on ability or SES, for example, children severely struggling to learn mathematics and those from low-income households are more likely to receive targeted interventions (Cozard & Riccomini, 2016; Schacter & Jo, 2016; Schacter et al., 2015)

However, the research presented in this thesis suggests this maths app intervention can be implemented relatively easily and efficiently in everyday classroom practice and is effective for supporting all children's mathematical development. This suggests the maths app intervention is a suitable class-wide teaching resource. This a particularly critical issue as research shows all children need a boost in early mathematical skills (APPG, 2014;

Ofsted, 2018; Pitchford et al., 2016) and based on other more traditional models of mathematical intervention children who may not typically be identified as struggling with mathematics may not receive the required extra support. In contrast, this research shows the maths apps can reach and benefit all children.

Furthermore, as many schools already have access to the hardware, this intervention places few additional financial demands on schools and requires minimal teacher training beyond how to use the touch-screen tablet device hardware. As outlined above, only an established routine was shown to be associated with children's learning gains with the maths apps (see Chapter 7). As such, the emphasis should be on ensuring teachers feel confident using the technology and support them to successfully embed the technology into their daily routine.

Study quality assessment

Overall, the reported studies all add to the growing evidence base outlined in Chapter 2 (Systematic Review). To assess and compare the potential bias across the three quantitative studies (Chapters 4-6) the same quality assessment tool designed for the Systematic Review (see Chapter 2) was utilised. The high validity and reliability of qualitative data was reported in Chapter 7 (UK RCT Implementation Evaluation).

Table 34 Study quality assessment results for the experimental, quantitative studies reported in Chapters 4-6.

Study	Randomisation (Allocation Bias)			Group Comparability (Confounding Bias)			Blinding (Performance & Detection Bias)		Withdrawals (Attrition Bias)	
	Random allocation	Detailed methods	Number stated	Comparability details	Initial equivalence		Implementers blinded	Assessors blinded	>80% in final analysis	Reasons stated
UK Proof of Concept	N	Y	Y	Y	N		N	N	Y	NA
Brazil Proof of Concept	N	Y	Y	Y	P		N	N	Y	Y
UK Feasibility RCT	Y	Y	Y	Y	P		N	Y	Y	Y

Note. Y = Yes; N = No; P = Partially; NS = Not stated; NA = Not applicable

Table 34, summaries the quality of the three quantitative studies (Chapter 4-6). Overall, the UK and Brazil Proof of Concept studies were considered to have a moderate risk of bias. In contrast, the risk of bias in the UK Feasibility RCT was rated as low. Specifically, the Proof of Concept studies did not have random group allocation, instead achievement levels based on teacher reports (UK) and predetermined class groups (Brazil) were utilised. Although these research designs address the key research questions, it does increase the risk of allocation bias, which may influence the study findings. This issue was addressed through pupil-level random allocation procedures used in the UK Feasibility RCT. Nevertheless, unlike previous research (see Chapter 2), all studies provided sufficient details of group allocation procedures, including clearly stating the number of children per group.

The risk of confounding bias is moderate for all three quantitative studies. Although, clear details on the comparability between groups, based on gender, age, and pre-test mathematical ability was given, none of the studies had initial equivalence in terms of baseline mathematical performance. To address this issue, all studies utilised a change score based statistical analysis (Thomas & Zumbo, 2016; Van Breukelen, 2006).

Similar to previous intervention studies (see Chapter 2), the risk of performance and detection bias was high in both Proof of Concept studies. This was because both the participating teachers implementing the maths apps and the researcher assessing the

children's mathematical abilities were aware of children's group allocation. This may have led to an unconscious bias, in favour of the maths app intervention groups.

To address this threat to internal validity and reduce the risk of performance and detection bias, the UK Feasibility RCT implemented several blinding procedures throughout the duration of the study. Specifically, as outlined in Chapter 6, the researchers who assessed children's mathematical performance before and after the 12-week intervention period and the company who independently scored the mathematical assessments were unaware of group allocation.

Finally, the risk of withdrawal bias was considered low across the three studies. Although the UK Feasibility RCT had an attrition rate of 15.6%, which was increased due to one school being on a school fieldtrip on the day of post-testing (n = 30), more than 80% of the children initially enrolled on the study were included in the final sample and reasons for withdrawal were clearly stated.

Impact evaluation

In all of the studies reported in this thesis teachers implemented the maths app intervention. As previously discussed, teacher led implementation is critical for ensuring ecological validity and generalisability of the intervention outside of the research context (Clements et al., 2015). To assess the uptake of the new maths app intervention beyond the study an immediate and 2-year

follow-up impact evaluation was conducted with the 11 schools participating in the UK Feasibility RCT (Chapter 6).

Immediate follow-up evaluation

Immediately following the 12-week intervention period, 13 participating teachers from 8 of the 11 schools completed and returned the end of study feedback questionnaire (see Appendix F). Participant demographics are reported in Chapter 7 (see Table 23). 100% of teachers reported feeling confident or very confident in using the maths apps in their classroom indicating the feasibility of teacher-led intervention implementation.

All participating teachers responded that they intend on using the maths apps in the next academic year. Six teachers indicated that they would like to implement the maths apps in small groups as a time-limited, targeted intervention tool for children struggling in mathematics. Two teachers also indicated a preference for implementing the maths apps in small groups as an independent daily mathematics activity. The child's independent use of the maths apps was also highlighted by four teachers who said they would like to implement the maths apps as either an independent free choice activity or as part of an independent challenges routine. The remaining two teachers said they would like to integrate the maths apps within the early years curriculum as a way of providing additional practice for concepts introduced in teacher-led standard practice. These responses highlight the different potential avenues

for implementing the maths apps based on the perceived needs of the class and individual children.

Two-year follow-up evaluation

During the two years after the UK Feasibility Trial (Chapter 6) was completed, communication was maintained between the research team and the participating schools. For example, the research findings were disseminated through a teacher workshop, individual school reports, and the journal article published in April 2018. In July 2018, all participating schools were invited to complete a further feedback questionnaire (see Appendix G); 2 of the 11 participating schools responded. Both teachers indicated they have continued to the use the maths apps in their classrooms either as a whole class mathematics activity in EYFS or as targeted intervention with children struggling in mathematics, including older children aged 5-7 years. In total, across the two schools, approximately 145 children have used the maths apps beyond the research context. Teachers said that they valued the self-paced nature of the maths apps, as it allows children to take ownership of their learning and increases access to learning opportunities for children who find the pace of whole class maths sessions challenging. Teachers reported the maths apps have supported a greater mastery of mathematics.

One teacher also highlighted the importance of establishing a structured routine for encouraging children's independence and engagement with the maths apps. This anecdotal evidence supports

the findings reported in the UK RCT implementation evaluation reported in Chapter 7. However, teachers did report they have faced challenges with the technology hardware, such as tablet devices updating and clearing children's progress and finding reliable headphones. Furthermore, one teacher indicated that sufficient teacher training is required to maximise the continued use of the maths app intervention.

Limitations and future research

Although this thesis research makes a valuable contribution to the evidence base on the use of app-based mathematics instruction in the early years of education, there are four key limitations, which future studies should address for the advancement of this research area.

Long-term follow-up

All of the studies reported in thesis focused on immediate learning gains following the maths app intervention; sustained learning gains assessed with a delayed post-test were not included. Only two studies to date have included such measures (Study 1, Outhwaite et al., 2017; Van der Ven et al., 2017). While results show positive sustained benefits of app-based mathematics instruction, conducting long-term follow up assessments are practically challenging, in terms of time, funding, and potentially high attrition rates.

Nevertheless, critiques argue long-term longitudinal studies are needed to consider potential novelty effects associated with app-based mathematics instruction, which may occur in studies of a short duration (Lieberman, Bates, & So, 2009). Although the research presented in this thesis partially addresses this issue by implementing the maths apps for a minimum of 8 weeks (see Table 32) and through examining the relationship between time on task and children's learning gains, further research is required. Specifically, longitudinal research including follow up assessments is needed to understand the long-term benefits of the maths app intervention, including the sustainability of the reduced achievement gap between lower and typically attaining children. This is currently being examined in the Educational Endowment Foundation funded, national trial of the new maths app intervention with low-achieving children aged 5-6 years from low-income geographical areas (EEF, 2018b).

Additional benefits

The current research has primarily focused on children's mathematical learning gains in response to the maths app intervention. However, the design of the maths apps (see Chapter 3) may also afford additional benefits, such as increased enjoyment, independence, and confidence. The immediate impact evaluation of the UK Feasibility RCT (see above) aimed to provide some indicative evidence.

The end of study feedback questionnaire (see Appendix F) aimed to assess teacher's perceived benefits of the maths app intervention for both mathematical skills and other areas of child development and the school curriculum. Results showed 92% of teachers perceived the maths apps to be beneficial for mathematical skill development. 83% and 66% of teachers saw the maths apps to be beneficial or very beneficial for problem solving and mathematical language, respectively. In terms of additional anecdotal benefits of the intervention, 100% of teachers thought the maths apps supported children's independent learning and 92% of teachers reported the apps increased children's confidence, attention, and engagement in mathematics. 85% of teachers said that children enjoyed using the maths apps. However, limited perceived benefits for confidence, enjoyment, attention, and engagement were reported in other areas of the early years school curriculum.

While this teacher-reported anecdotal evidence supports the empirical findings reported in Chapter 6 (UK Feasibility RCT), the security of the anecdotal findings is limited. Specifically, the questionnaire is vulnerable to subjectivity in the interpretation of questions and potential social desirability bias (Creswell & Creswell, 2010). Further research examining additional benefits should include more objective, empirical assessments, such as measures of sustained visual attention (Pitchford & Outhwaite, 2016). This will enhance the holistic understanding of the reach and utility of the new maths app intervention.

Home learning environment

Educational maths apps also have the potential to support access to mathematical learning at home, which was highlighted in recent UK educational policy calling for "new technology to spearhead classroom revolution" (Department for Education, 2018). While experimental research in this area is in its infancy, studies show promising results. For example, Berkowitz et al. (2015) found the use of educational maths apps at home supported young children's early mathematical achievement in school, particularly for parents who experienced mathematical anxiety.

Furthermore, anecdotal evidence from the immediate impact evaluation showed 85% of teachers who responded to the questionnaire said they would recommend the new maths apps to parents. The majority of these teachers said that the maths apps could be a valuable tool to bridge the gap between home and schools as the apps are easy to use, self-paced, and enjoyable.

The potential utility of the maths app intervention at home is also supported by the UK RCT Implementation Evaluation study (Chapter 7), as strong learning gains with the apps were not associated with high levels of support, supervision, or intended implementation. This suggests the apps could be easily implemented by parents who may not have extensive teaching knowledge and experience. However, further research is required to solidify the initial evidence base and identify recommendations for best practice. Addressing parental and educator concerns

surrounding the use of app technology at home is particularly critical amidst the current screen time debate (Greenfield, 2015; Palmer, 2007; Sigman, 2012).

Teacher's belief system

As illustrated in Figure 6, teacher's beliefs comprise the macro-level system of the multi-level, ecological, determinant framework. Although this thesis research has attempted to capture teacher's informal perceptions of the maths app intervention through the impact evaluation (see above) more formal evaluation is required. In particular, future longitudinal research should examine the relationship between teacher's perceptions, children's learning gains, and the continued use of the maths app intervention (Blackwell et al., 2014).

This expansion of the multi-level, ecological, determinant framework with the research evidence base would help further identify the determinants associated with the maths app intervention (Nilsen, 2015). As such, these potential findings could make important contributions to educational policy and practice, such as the critical role of teacher training opportunities for implementing technology in their classroom.

Conclusion

In conclusion, this thesis has examined the reach and utility of a new maths app intervention for supporting early mathematical development in young children in the UK and Brazil. Following a

pragmatic, mixed-methods approach, utilising a multi-level, ecological, determinant framework, this research has demonstrated the maths apps are a form of effective mathematical instruction that can be implemented as part of a well-balanced early years curriculum. The maths apps can be easily implemented by teachers and do not require additional time-consuming teaching demands. However, learning gains with the maths apps can be maximised through successful embedding the intervention within an established routine.

The maths apps have been shown to support early mathematical development for all children, regardless of baseline ability. However, low-achieving children may particularly benefit from this form of self-paced instruction. In particular, features of the interactive maths apps may also reduce the demands on the child's memory capacity. However, in a bilingual context, children need to have a sufficiently developed proficiency in the language of instruction to progress through the apps. A child's socio-economic status was not found to be related to learning gains with the apps.

It is suggested that the maths apps which map highly on the universal principles of active, engaged, meaningful, and socially interactive learning work by combining aspects of targeted practice, direct instruction, and play-based pedagogies, enabling children to master basic mathematical factual and conceptual skills. This provides a strong foundation for higher-order mathematical reasoning and problem solving skills.

Collectively, this evidence suggests the interactive maths apps can provide an accessible, inclusive, and child-centred learning environment, supplemental to classroom instruction. This thesis demonstrates the potential for app-based mathematics instruction to contribute to the long-term journey towards ensuring quality education for all.

References

Abedi, J. (2003). *Impact of Student Language Background on Content-Based Performance: Analyses of Extant Data*. Los Angeles, USA: CRESST/ University of California.

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology*, *3*(301), 1-9.

Adams, G., & Carnine, D. (2003). Direct instruction. In H. Swanson, K. Harris & S. Graham (Eds.), *Handbook of Learning Disabilities*. New York, USA: Guliford.

Alloway, T. P. (2006). How does working memory work in the classroom?. *Educational Research and Reviews*, *1*(4), 134-139.

Alloway, T. P. (2011). *Improving Working Memory: Supporting Students' Learning*. London, UK: Sage.

Alloway, T. P., Alloway, R. G., & Wootan, S. (2014). Home sweet home: Does where you live matter to working memory and other cognitive skills?. *Journal of Experimental Child Psychology*, *124*, 124-131.

Anders, Y., Rossbach, H. G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly*, *27*(2), 231-244.

APPG. (2014). *Maths and numeracy in the early years*. London, UK: All Party Parliamentary Group for Maths & Numeracy.

Ash, D., & Wells, G. (2006). Dialogic inquiry in classroom and museum. In Z. Bekerman, N. Burbules, & D. Silberman-Keller (Eds.), *Learning in Places* (pp. 35–54). New York, USA: Peter Lang.

Askell-Williams, H., Dix, K. L., Lawson, M. J., & Slee, P. T. (2013). Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies. *School Effectiveness and School Improvement*, *24*(3), 357-381.

Baker, C., & Hornberger, N. H. (2001). *An Introductory Reader to the Writings of Jim Cummins.* Clevedon, UK: Multilingual Matters.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215-220.

Barnett, W. S., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly*, *22*(3), 277-293.

Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development*, *20*(1), 80-128.

Basen-Engquist, K., O'Hara-Tompkins, N., Lovato, C. Y., Lewis, M. J., Parcel, G. S., & Gingiss, P. (1994). The effect of two types of teacher training on implementation of Smart Choices: a tobacco prevention curriculum. *Journal of School Health*, *64*(8), 334-339.

Beach, R., & O'Brien, D. (2015). *Using apps for learning across the curriculum. A literacy-based framework and guide.* New York, USA: Routledge.

Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on Educational Communications and Technology* (pp. 505-518). New York, USA: Springer.

Benassi, V., Overson, C. E., & Hakala, C. (2014). *Applying science of learning in education: Infusing psychological science into the curriculum*. Retrieved from http://teachpsych.org/ebooks/asle2014/index.php

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, *21*(4), 755-765.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, *12*(1), 23-33.

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, C. S., & Beilock, S. L. (2015). Maths at home adds up to achievement in schools. *Science, 350*(6257), 196-198.

Biesta, G. J. (2010). Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, *29*(5), 491-503.

Björklund, C. (2008). Toddlers' opportunities to learn mathematics. *International Journal of Early Childhood*, *40*(1), 81-95.

Björn, P. M., Aunola, K., & Nurmi, J. E. (2016). Primary school text comprehension predicts mathematical word problem-solving skills in secondary school. *Educational Psychology*, *36*(2), 362-377.

Blackwell, C. K., Lauricella, A. R., & Wartella, E. (2014). Factors influencing digital technology use in early childhood education. *Computers & Education*, *77*, 82-90.

Blum-Ross, A., & Livingstone, S. (2016). *Families and screen time: Current advice and emerging research*. Media Policy Brief 17. London, UK: Media Policy Project, London School of Economics and Political Science.

Brabeck, M., Jeffrey, J., & Fry, S. (2011). *Practice for knowledge acquisition (not drill and kill*). Washington DC, USA: American Psychological Association.

Bransford, J. D., Brown, A., & Cocking, R. (1999). *How people learn: Mind, brain, experience, and school*. Washington DC, USA: National Research Council.

Brantlinger, E., Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional children*, *71*(2), 195-207.

Braun, V., Clarke, V., & Terry, G. (2014). Thematic analysis. In P. Rohleder & A. C. Lyons (Eds.). Qualitative Research in Clinical and Health Psychology. Hampshire, UK: Palgrave Macmillan.

Brombacher, A. (2010). *Malawi Early Grade Mathematics Assessment (EGMA): National Baseline Report.* Malawi: USAID/Malawi and the Ministry of Education, Science and Technology.

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Massachusetts, USA: Harvard University Press.

Brown, M., & Richardson, M. (2017). Understanding and Synthesizing Numerical Data from Intervention Studies. In A. Boland, M. G. Cherry, & R. Dickson (Eds.), *Doing a Systematic Review a Student's Guide* (pp. 131-155). London, UK: Sage.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36-41.

Burns, M. K., Kanive, R., & DeGrande, M. (2012). Effect of a computer-delivered math fact intervention as a supplemental intervention for math in third and fourth grades. *Remedial and Special Education*, *33*(3), 184-191.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, *332*(6033), 1049-1053.

Calder, N. (2016). 'Makes learning easier- they're active' using apps in early years mathematics. In N. Kucirkova & G. Falloon (Eds.), *Apps, technology and younger learners. International evidence for teaching* (pp. 135-146). London, UK: Routledge.

Calvert, S. L., Strong, B. L., Jacobs, E. L., & Conger, E. E. (2007). Interaction and participation for young Hispanic and Caucasian girls' and boys' learning of media content. *Media Psychology*, *9*(2), 431-445.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago, USA: RandMcNally.

Carnine, D., & Gersten, R. (2000). The nature and roles of research in improving achievement in mathematics. *Journal for Research in Mathematics Education*, *31*(2), 138-143.

Carr, J. (2012). Does Maths achievement h'APP'en when iPads and game-based learning are incorparated into fifth-grade Mathematics instruction? *Journal of Information Technology Education, 11,* 269-286.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing Maths achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105.

Chodura, S., Kuhn, J. T., & Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. *Zeitschrift für Psychologie*, *223*(2), 129-144.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning?. *International Journal of Bilingual Education and Bilingualism*, *7*(4), 303-319.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford, UK: Oxford University Press.

Clarke, B. (2014). *The Use of Tablets in UK Schools- Stage 4.* London, UK: Family, Kids and Youth Research Group.

Clarke. (2004). "Structured judgement methods"- the best of both worlds? In Z. Todd, B. Nerlich, S. McKeown, & D. Clarke (Eds.). *Mixing Methods in Psychology: The Intergration of Qualitative and Quantitative Practice.* Hove, UK: Psychology Press.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, *38*(2)136-163.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, USA: Routledge.

Clements, D. H., Baroody, A. J., & Sarama, J. (2013). *Background research on early mathematics.* Washington DC, USA: National Governors Association.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2015). Sustainability of a scale-up intervention in early mathematics: A longitudinal evaluation of implementation fidelity. *Early Education and Development*, *26*(3), 427-449.

Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioural Education, 19*(1), 93-105.

Coelho Liberali, F., & Megale, A. (2016). Elite bilingual education in Brazil: an applied linguist's perspective. *Colombian Applied Linguistics Journal*, *18*(2), 95-108.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New Jersey, USA: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York, USA: Routledge.

Connolly, P., Biggart, A., Miller, S., O'Hare, L., & Thurston, A. (2017). *Using Randomised Controlled Trials in Education*. London, UK: SAGE.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, 60(3), 276-291.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional children*, *79*(2), 135-144.

Cooper, L. Z. (2005). Developmentally appropriate digital environments for young children. *Library Trends*, *54*(2), 286-302.

Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education*, *43*(1), 75-96.

Cozad, L. E., & Riccomini, P. J. (2016). Effects of Digital-Based Math Fluency Interventions on Learners with Math Difficulties: A Review of the Literature. *Journal of Special Education Apprenticeship*, *5*(2), 1-19.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12-26.

Creswell, J. W., & Creswell, D. (2018). *Research Design. Qualitative, Quantitative, & Mixed Methods Approaches.* London, UK: Sage.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124-130.

Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education*, *110*, 51-63.

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1992). *Optimal experience: Psychological Studies of Flow in Consciousness*. Cambridge, UK: Cambridge University Press.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, *9*, 1-43.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Eds.), *Schooling and Language Minority Students. A Theoretical Framework*. California, USA: California State Department of Education.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.

Cummins, J. (2008). BICS and CALP: Empirical and Theoretical Status of the Distinction. In B.Street & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education.* New York, USA: Springer.

Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review*, *36*(4), 562-581.

Delazer, M. (2003). Neuropsychological findings on conceptual knowledge of arithmetic. In A. Baroody & A. D. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 385-407). New Jersey, USA: Erlbaum.

Denvir, B., & Brown, M. (1986). Understanding of number concepts in low attaining 7–9 year olds: Part I. Development of descriptive framework and diagnostic instrument. *Educational Studies in Mathematics*, *17*(1), 15-36.

Department for Education. (2010). *Phonics Core Criteria and the Self-assessment Process*. London, UK: Department for Education.

Department for Education. (2013). *Early years outcomes.* London, UK: Department for Education.

Department for Education. (2014). *National curriculum in England: mathematics programmes of study.* London, UK: Department for Education.

Department for Education. (2016). *Phonics screening check and Key Stage 1 assessments in England, 2016*. London, UK: Department for Education.

Department for Education. (2018). *New technology to spearhead classroom revolution.* London, UK: Department for Education.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., ... & Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, *1*(3), 6-28.

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*(3), 284-298.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience and education*. Hove, UK: Psychology Press.

Drinkwater, D. (2013). *Brazilian government spends big on nearly 500,000 tablets for local teachers.* Retrieved from http://tabtimes.com/news/ education/2013/11/27/brazilian-government-spends-big-nearly-500000-tabletslocal-teachers

Duijzer, C. A. C. G., Shayan, S., Bakker, A., Van der Schaaf, M. F., & Abrahamson, D. (2017). Touchscreen Tablets: Cooridnating Action and Perception for Mathsematical Cognition. *Frontiers in Psychology, 8*(144), doi:10.3389/fpsyg.2017.00144.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japal, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446.

Dundar, Y., & Fleeman, N. (2017). Developing My Search Strategy. In A. Boland, M. G. Cherry, & R. Dickson (Eds.), *Doing a Systematic Review a Student's Guide* (pp. 61-79). London, UK: Sage.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327-350.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237-256.

Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology*, *58*(2), 107-112.

Eden, J., Wheatley, B., McNeil, B. & Sox, H. (2008). *Knowing What Works in Health Care. A Roadmap for the Nation.* Washington, DC, USA: National Academies Press.

Education Endowment Foundation (2018a). *Pipeline of EEF trials.* Retrieved from https://educationendowmentfoundation.org.uk/projects-and-

evaluation/evaluating-projects/evaluation-information-for-grantees/pipeline-of-eef-trials/

Education Endowment Foundation (2018b). *onebillion: app-based maths learning.* Retrieved from https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/onebillion-app-based-maths-learning/

Education Endowment Foundation. (2017). *The Attainment Gap.* Retrieved from https://educationendowmentfoundation.org.uk/public/files/Annual_Reports/EEF_Attainment_Gap_Report_2018.pdf

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, *5*(1), 47-53.

Elliott, J. (2004). Making Evidence-based Practice Educational. In G. Thomas & R. Pring (Eds). *Evidence-Based Practice in Education*. Maidenhead, UK: Open University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406.

Evangelou, M., Sylva, K., Kyriacou, M., Wild, M., & Glenny, G. (2009). *Early years learning and development literature review*. London, UK: Department for Children, Schools and Families.

Falloon, G. (2013). Young students using iPads: App design and content influences on their learning pathways. *Computers & Education*, *68*, 505-521.

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology*, *108*(4), 528-546.

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, *84*(6), 1872-1878.

Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). *Implementation research: a synthesis of the literature.*Florida, USA: University of South Florida.

Flewitt, R., Messer, D., & Kucirkova, N. (2015). New directions for early literacy in a digital age: The iPad. *Journal of Early Childhood Literacy*, *15*(3), 289-310.

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: developers' views of implementation barriers and facilitators. *School Mental Health*, *1*(1), 26-36.

Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. M. (2016). Improving mathematics learning of

kindergarten students through computer-assisted instruction. *Journal for Research in Mathematics Education, 47,* 206–232.

Fritz, J. M., & Cleland, J. (2003). Effectiveness versus efficacy: more than a debate over language. *Journal of Orthopaedic & Sports Physical Therapy, 33*(4), 163-165.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. *Reading research quarterly*, *41*(1), 93-99.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*(1), 29-43.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., ... & Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology*, *101*(3), 561-576.

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly*, *27*(2), 198-209.

Gagné, R. M. (1968). Contributions of learning to human development. *Psychological Review*, *75*(3), 177-191.

Gambaro, L., Stewart, K., & Waldfogel, J. (2015). A question of quality: do children from disadvantaged backgrounds receive lower quality early childhood education and care?. *British Educational Research Journal*, *41*(4), 553-574.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441-467.

Gathercole, S. E., & Alloway, T. P. (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry*, *47*(1), 4-15.

Gathercole, S. E., & Alloway, T. P. (2007). *Understanding working memory: A classroom guide.* London, UK: Harcourt Assessment.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In S. J. Pickering (Ed.). *Working Memory and Education* (pp. 219-240). London, UK: Elsevier.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177-190.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345-362.

Geary, D. C. (2011a). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539-1552.

Geary, D. C. (2011b). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics*, *32*(3), 250-263.

Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 253-267). Hove, UK: Psychology Press.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, *33*(1), 18-28.

Gibbs, L., Kealy, M., Willis, K., Green, J., Welch, N., & Daly, J. (2007). What have sampling and data collection got to do with good qualitative research?. *Australian and New Zealand journal of public health*, *31*(6), 540-544.

Ginsburg, A., & Smith, M. S. (2016). *Do randomized controlled trials meet the "Gold Standard"?* Washington, DC, USA: American Enterprise Institute.

Ginsburg, H. P., and Baroody, A. J. (2003). *Test of Early Mathematics Ability (3rd edition*). Texas, USA: Pro-Ed.

Ginsburg, H. P., Pappas, S., & Seo, K. H. (2001). Everyday mathematical knowledge. In S. L. Golbeck (Ed.), *PsychoTaking Shape 1877 Logical Perspectives on Early Childhood Education* (pp. 181–219). New Jersey, USA: Erlbaum.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, *93*(8), 1261-1267.

Glutting, J., & Jordan, N. C. (2012). *Number Sense Screener*. Baltimore, USA:Brookes Publishing.

Golbeck, S. L. (2001). *Psychological Perspectives on Early Childhood Education*. New Jersey, USA: Erlbaum.

Gomez, L., Freeman, D., & Freeman, Y. (2005). Dual language education: A promising 50–50 model. *Bilingual Research Journal*, *29*(1), 145-164.

Goswami, U. (2006). Neuroscience and education: from research to practice? *Nature Reviews Neuroscience*, *7*(5), 406-413.

Grabinger, R. S., & Dunlap, J. C. (1995). Rich environments for active learning: A definition. *Association for Learning Technology Journal*, *3*(2), 5-34.

Grant, M. M., Tamim, S., Brown, D. B., Sweeney, J. P., Ferguson, F. K., & Jones, L. B. (2015). Teaching and learning with mobile computing devices: Case study in K-12 classrooms. *TechTrends*, *59*(4), 32-45.

Gray, P. (2015). *Free to learn: Why unleashing the instinct to play will make our children happier, more self-reliant, and better students for life*. New York, USA: Basic Books.

Greenfield, S. (2015). *Mind Change: How digital technologies are leaving their mark on our brains*. London, UK: Random House.

Greenhalgh, J., & Brown, T. (2017). Quality Assessment: Where Do I Begin? In A. Boland, M. G. Cherry, & R. Dickson (Eds.), *Doing a Systematic Review a Student's Guide* (pp. 107-131). London, UK: Sage.

Greenhalgh, T., Robert, G., Bate, P., Macfarlane, F., & Kyriakidou, O. (2008). *Diffusion of innovations in health service organisations: a systematic literature review*. Chichester, UK: Wiley.

Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology, 106,* 58–68.

Haden, C. A., Cohen, T., Uttal, D. H., & Marcus, M. (2016). Building learning: Narrating experiences in a children's museum. In D. M. Sobel, & J. L. Jipson (Eds.), *Cognitive Development in Museum Settings: Relating Research and Practice* (pp. 84-103). New York, USA: Routledge.

Hagena, M., Leiss, D., & Schwippert, K. (2017). Using Reading Strategy Training to Foster Students' Mathematical Modelling Competencies: Results of a Quasi-Experimental Control Trial. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7b), 4057-4085.

Hall, A. (2007). Vygotsky goes online: Learning design from a socio-cultural perspective. In *Learning and Socio-Cultural Theory: Exploring Modern Vygotskian Perspectives International Workshop 2007*, *1*(6), 94-107.

Hammersley, M. (2007). *Educational Research and Evidence-Based Practice*. London, UK: Sage.

Hansen, W. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.). *Defining Prevention Science* (pp. 335-359). New York, USA: Springer.

Hansen, W. B., Pankratz, M. M., Dusenbury, L., Giles, S. M., Bishop, D. C., Albritton, J., ... & Strack, J. (2013). Styles of adaptation: The impact of frequency and valence of adaptation on preventing substance use. *Health Education*, *113*(4), 345-363.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, *15*(1), 75-90.

Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London, UK: Cabinet Office.

Haβler, B., Major, L., & Hennessy, S. (2015). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139-156.

Hemsley, G., Holm, A., & Dodd, B. (2014). Better in both? Bilingual intervention in an Australian school context. *Australian Journal of Education*, *58*(2), 139-152.

Herodotou, C. (2018). Young children and tablets: A systematic review of effects on learning and development. *Journal of Computer Assisted Learning*, *34*(1), 1-9.

Higgins, J. P. T., & Green, S. (2006). *Cochrane handbook for systematic reviews of interventions 4.2.6.* Chichester, UK: Wiley.

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation Full Report. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Pu

blications/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Hilton, A. (2016). Engaging Primary School Students in Mathematics: Can iPads Make a Difference? *International Journal of Science and Mathematics Education*, doi:10.1007/s10763-016-9771-5.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons from the Science of Learning. *Psychological Science, 16*(1), 3-34.

Hodkinson, P., & Smith, J. (2004). "The Relationship between Research, Policy and Practice." In G. Thomas & R. Pring (Eds). *Evidence-Based Practice in Education*. Maidenhead, UK: Open University Press.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9-F15.

Holmes, W., & Dowker, A. (2013). Catch up numeracy: a targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education*, *15*(3), 249-265.

Hubber, P., Outhwaite, L. A., Chigeda, A., McGrath, S., Hodgen, J., & Pitchford, N. J. (2016). Should touch screen tablets be used to improve educational outcomes in primary school children in developing countries? *Frontiers in Psychology, 7*(839), doi:10.3389/fpsyg.2016.00839.

Humphrey, N. (2013). *Social and emotional learning: a critical appraisal.* London, UK: Sage.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in educational settings: A synthesis of the literature.* London, UK: Education Endowment Foundation.

Hwang, W. Y., Shadiev, R., Tseng, C. W., & Huang, Y. M. (2015). Exploring Effects of Multi-Touch Tabletop on Collaborative Fraction Learning and the Relationship of Learning Behavior and Interaction with Learning Achievement. *Educational Technology & Society*, *18*(4), 459-473.

Imbo, I., & Vandierendonck, A. (2008). Practice effects on strategy selection and strategy efficiency in simple mental arithmetic. *Psychological Research*, *72*(5), 528-541.

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology*, *38*(3), 455-464.

Irwin, D. M., & Bushnell, M. M. (1980). *Observational strategies for child study*. New York, USA: Holt, Rinehart, and Winston.

James, K. H., & Swain, S. N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental Science*, *14*(4), 673-678.

Jerrim, J., & Vignoles, A. (2016). The link between East Asian 'mastery' teaching methods and English children's Mathsematics skills. *Economics of Education Review, 50,* 29-44.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., & Meira, A. (2012). *Technology outlook for Brazilian primary and secondary education 2012–2017:* An NMC Horizon Project regional analysis. Texas, USA: The New Media Consortium.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Johnson, W., McGue, M., & Iacono, W. G. (2007). Socioeconomic status and school grades: Placing their association in broader context in a sample of biological and adoptive families. *Intelligence*, *35*(6), 526-541.

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews*, *15*(1), 60-68.

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized

controlled study. *Journal of Educational Psychology*, *104*(3), 647-660.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181-195.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*(3), 834-850.

Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*(1), 153-175.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850-867.

Kalaycioğlu, D. B. (2015). The Influence of Socioeconomic Status, Self-Efficacy, and Anxiety on Mathematics Achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and

the USA. *Educational Sciences: Theory and Practice*, *15*(5), 1391-1401.

Kamwangamalu, N. M. (2010). Multilingualism and codeswitching in education. In N. H. Hornberger & S. Lee McKay (Eds.), *Sociolinguistics and Language Education*. Bristol, UK: Multilingual Matters.

Karmiloff-Smith, A. (1994). Precis of Beyond modularity: A developmental perspective on cognitive science. *Behavioural and Brain Sciences, 17*(4), 693-745.

Kaufman, A. S., & Kaufman, N. L. (2004). *KABC-II: Kaufman assessment battery for children*. Texas, USA: AGS Publishing/

Kempert, S., Saalbach, H., & Hardy, I. (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology*, *103*(3), 54-561.

Kersey, A. J., & James, K. H. (2013). Brain activation patterns resulting from learning letter forms through active self-production and passive observation in young children. *Frontiers in Psychology*, *4*(567), 1-15.

Kidd, J. K., Carlson, A. G., Gadzichowski, K. M., Boyer, C. E., Gallington, D. A., & Pasnak, R. (2013). Effects of patterning instruction on the academic achievement of 1st-grade children. *Journal of Research in Childhood Education*, *27*(2), 224-238.

Kim, Y., & Steiner, P. (2016). Quasi-experimental designs for causal inference. *Educational Psychologist*, *51*(3-4), 395-405.

Kirschner, P. A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Korat, O., Shamir, A., & Segal-Drori, O. (2014). E-books as a support for young children's language and literacy: the case of Hebrew-speaking children. *Early Child Development and Care*, *184*(7), 998-1016.

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education*, *24*(2), 97-114.

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., … von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage*, *57*(3), 782-795.

Kucirkova, N. (2014). iPads in early education: separating assumptions and evidence. *Frontiers in Psychology, 5*(715), doi:10.3389/fpsyg.2014.00715.

Kucirkova, N., Messer, D., Sheehy, K., & Panadero, C. F. (2014). Children's engagement with educational iPad apps: Insights

from a Spanish classroom. *Computers & Education*, *71*, 175-184.

Kucirkova, N., Sheehy, K., & Messer, D. (2015). A Vygotskian perspective on parent–child talk during iPad story sharing. *Journal of Research in Reading*, *38*(4), 428-441.

Laher, S., & Cockcroft, K. (2017). Moving from culturally biased to culturally responsive assessment practices in low-resource, multicultural settings. *Professional Psychology: Research and Practice*, *48*(2), 115-121.

LeFevre, J., DeStefano, D., Coleman, B., Shanahan, T., & Campbell, J. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 361-378). New York, USA: Psychology Press.

Lendrum, A., Humphrey, N., & Greenberg, M. (2016). Implementing for success in school-based mental health promotion: The role of quality in resolving the tension between fidelity and adaptation. In R. Shute & P. Slee (Eds.), *Mental health and wellbeing through schools: The way forward* (pp. 53-63). London, UK: Taylor and Francis.

Lieberman, D. A., Bates, C. H., & So, J. (2009). Young children's learning with digital media. *Computers in the Schools*, *26*(4), 271-283.

Lillard, A. S. (2005). *Montessori: The science behind the genius.*Oxford, UK: Oxford University Press.

Lillard, A. S. (2011). Mindfulness practices in education: Montessori's approach. *Mindfulness*, *2*(2), 78-85.

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior*, *31*(2), 242-257.

Lindahl, M. G., & Folkesson, A. M. (2012). ICT in preschool: friend or foe? The significance of norms in a changing practice. *International Journal of Early Years Education*, *20*(4), 422-436.

Lowyck, J. (2014). Bridging learning theories and technology-enhanced environments: A critical appraisal of its history. In In J. M. Spector, M. D. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 3-20). New York, USA: Springer.

Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisted: A key model for instructional technology. *Educational Technology Research and Development, 53*(4), 41-55.

Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research and Development*, *53*(4), 41-55.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333-369.

Maths Assessment Resource Service. (2015). *Progress Test in Maths Level 5*. London, UK: GL Assessment.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14-19.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*(2), 105-123.

Mayo, M. J. (2009). Video games: A route to large-scale STEM education? *Science*, *323*(5910), 79-82.

McDonagh, M., Peterson, K., Raina, P., Chang, S., & Shekelle, P. (2013). *Avoiding bias in selecting studies.* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK126701/

McDonald, S. K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, *35*(3), 15-24.

McLennan, D., Barnes, H., Noble, M., Davies, J., Garratt, E., & Dibben, C. (2011). *The English Indices of Deprivation 2010: Technical report*. Retrieved from https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010-technical-report

McNeal Jr, R. B., Hansen, W. B., Harrington, N. G., & Giles, S. M. (2004). How All Stars works: An examination of program effects on mediating variables. *Health Education & Behavior*, *31*(2), 165-178.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270-291.

Mertens, D. (2015). Mixed Methods and Wicked Problems. *Journal of Mixed Methods Research, 9*(1), 3-6.

Messer, D., Thomas, L., Holliman, A., & Kucirkova, N. (2018). Evaluating the effectiveness of an educational programming intervention on children's mathematics skills, spatial awareness and working memory. *Education and Information Technologies*, doi:10.1007/s10639-018-9747-x.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London, UK: Sage.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*(3), 272-277.

Milner, G., Mulera, D., Banda, T. C., Matale, E., and Chimombo, J. (2011). *Trends in Achievement Levels of Grade 6 Learners in Malawi.* Gaborone: The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, UK: Cambridge University Press.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264-269.

Montessori, M. (1967). *The discovery of the child*. (M. J. Costello, Trans.). New York, USA: Ballantine.

Moore, J. E., Bumbarger, B. K., & Cooper, B. R. (2013). Examining adaptations of evidence-based programs in natural contexts. *The Journal of Primary Prevention*, *34*(3), 147-161.

Morgan, C. (2013). Language and mathematics: a field without boundaries. In B. Ubuz, C. Haser, & A. Mariotti (Eds.), *Proceedings of CERME 8*. Ankara, Turkey: METU.

Morrison, G. R., Ross, S. M., & Baldwin, W. (1992). Learner control of context and instructional support in learning elementary school mathematics. *Educational Technology Research and Development*, *40*(1), 5-13.

Morrison, K. (2001). Randomised controlled trials for evidence-based education: some problems in judging'what works'. *Evaluation & Research in Education*, *15*(2), 69-83.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, *40*(2), 120-123.

Moschkovich, J. (2007). Using two languages when learning mathematics. *Educational studies in Mathematics*, *64*(2), 121-144.

Muis, K. R., Ranellucci, J., Trevors, G., & Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction*, *38*, 1-13.

Nacher, V., Jaen, J., Navarro, E., Catala, A., & González, P. (2015). Multi-touch gestures for pre-kindergarten children. *International Journal of Human-Computer Studies*, *73*, 37-51.

Naismith, L., Lonsdale, P., Vavoula, G. N., & Sharples, M. (2004). *Mobile technologies and learning in futurelab literature review series.* Report No. 11, Futurelab.

Naylor, P. J., Nettlefold, L., Race, D., Hoy, C., Ashe, M. C., Higgins, J. W., & McKay, H. A. (2015). Implementation of school based physical activity interventions: a systematic review. *Preventive Medicine*, *72*, 95-115.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention

fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*(4), 374-396.

Nielsen, J. (2014). *Response Times: The 3 Important Limits.* Retrieved from https://www.nngroup.com/articles/response-times-3-important-limits/

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*(53), 1-13.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., ... & Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly*, *25*(3), 314-328.

Ofsted. (2018). *Bold Beginnings: The Reception curriculum in a sample of good and outstanding primary schools.* London, UK: Ofsted.

Ofsted. (n.d.). *Inspection Reports.* Retrieved from https://reports.beta.ofsted.gov.uk for each participating school.

Organisation for Economic Co-Operation and Development. (2016). *PISA 2015 results.* Retrieved from https://www.oecd.org/pisa/keyfindings/

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2018). Raising early achievement in math with interactive apps: a randomized control trial. *Journal of Educational Psychology*. Doi:10.1037/edu0000286.

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers & Education*, *108*, 43-58.

Palmer, S. (2007). *Toxic childhood: How the modern world is damaging our children and what we can do about it*. London, UK: Orion.

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278-293.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348-367.

Pavio, A. (1986). *Mental Representations: A dual coding approach.* Oxford, UK: Oxford University Press.

Pawson, R., & Tilley, N. (1997). Realistic evaluation. In S. Matthieson (Ed.), *Encyclopedia of Evaluation* (pp. 359-367). London, UK: Sage.

Peng, P., Wang, C., & Namkung, J. (2018). Understanding the Cognition Related to Mathematics Difficulties: A Meta-Analysis on the Cognitive Deficit Profiles and the Bottleneck Theory. *Review of Educational Research*, *88*(3), 434-476.

Piaget, J. (1977). *Psychology and epistemology. Towards a theory of knowledge*. Harmondsworth, UK: Penguin.

Pitchford, N. J. (2015). Development of early Mathematical skills with a tablet intervention: a randomised control trial in Malawi. *Frontiers in Psychology, 6*(485), doi:10.3389/fpsyg.2015.00485.

Pitchford, N. J., & Outhwaite, L. A. (2016). Can touch screen tablets be used to assess cognitive and motor skills in early years primary school children? A cross-cultural study. *Frontiers in Psychology*, *7*(1666), doi: 10.3389/fpsyg.2016.01666.

Pitchford, N. J., Papini, C., Outhwaite, L. A., & Gulliford, A. (2016). Fine Motor Skills Predict Maths Ability Better than They Predict Reading Ability in the Early Primary School Years. *Frontiers in Psychology, 7*(783), doi:10.3389/fpsyg.2016.00783.

Plowman, L., & Stephen, C. (2005). Children, play, and computers in pre-school education. *British Journal of Educational Technology*, *36*(2), 145-157.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154-166.

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. London, UK: Sage.

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of experimental child psychology*, *110*(4), 647-658.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110-122.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375-394.

Ravitch, D. (2010). *The life and death of the great American school system: How testing and choice are undermining education.* New York, USA: Basic Books.

Razel, M., & Eylon, B. S. (1990). Development of visual cognition: Transfer effects of the Agam program. *Journal of Applied Developmental Psychology*, *11*(4), 459-485.

Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., ... & Wang, D. T. (1998). How best to measure implementation of school health curricula: a comparison of three measures. *Health Education Research*, *13*(2), 239-250.

Revicki, D. A., & Frank, L. (1999). Pharmacoeconomic evaluation in the real world. *Pharmacoeconomics*, *15*(5), 423-434.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943-973.

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*, *123*(3), A12-A12.

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (2013). *Qualitative research practice: A guide for social science students and researchers*. London, UK: Sage.

Rittle-Johnson, B., Fyfe, E. R., Hofer, K. G., & Farran, D. C. (2016). Early Math Trajectories: Low-Income Children's Mathematics Knowledge From Ages 4 to 11. *Child Development*. doi: 10.1111/cdev.12662.

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., ... & Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: a randomized clinical trial. *JAMA Pediatrics*, *170*(5), e154568-e154568.

Roberts-Holmes, G. (2015). The 'datafication' of early years pedagogy: 'If the teaching is good, the data should be good and if there's bad teaching, there is bad data'. *Journal of Education Policy*, *30*(3), 302-315.

Robson, C., & McCartan, K. (2011). *Real World Research.* Chichester, UK: Wiley.

Rogers, E. M. (2003). *The diffusion of innovations.* New York, USA: Free Press.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural Studies of Mind* (pp. 139–164). Cambridge, UK: Cambridge University Press.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641.

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. *The California School Psychologist*, *14*(1), 23-34.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67.

Rycroft-Smith, L., Gilmore, C., & Cragg, L. (2017). *Why is working memory important for mathematics learning?* Cambridge, UK: Cambridge Mathematics.

Sabanathan, S., Wills, B., & Gladstone, M. (2015). Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately?. *Archives of Disease in Childhood*, *100*(5), 482-488.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. London, UK: Sage.

Sale, J., Lohfeld, L., & Brazil, K. (2002). Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research. *Quality & Quantity, 36,* 43-53.

Sammons, P. (2010). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioural research.* London, UK: Sage.

Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development*

Perspectives, *3,* 145–150, doi:10.1111/j.1750-8606.2009.00095.

Sarama, J., Lange, A. A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly*, *27*(3), 489-502.

Sawyer, R. K. (2006). *The Cambridge Handbook of the Learning Sciences*. Cambridge, UK: Cambridge University Press.

See, B. H., Gorard, S., & Siddiqui, N. (2015). Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme. *Studies in Educational Evaluation*, *47*, 83-92.

Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior*, *55*, 223-229.

Schacter, J., & Jo, B. (2017). Improving preschoolers' mathematics achievement with tablets: a randomized controlled trial. *Mathematics Education Research Journal*, *29*(3), 313-327.

Schacter, J., Shih, J., Allen, C. M., DeVaul, L., Adkins, A. B., Ito, T., & Jo, B. (2015). Math Shelf: A Randomized Trial of a Prekindergarten Tablet Number Sense Curriculum. *Early Education and Development*, *27*(1), 74-88.

Schagen, S., Blenkinsop, S., Schagen, I., Scott, E., Eggers, M., Warwick, I., ... & Aggleton, P. (2005). Evaluating the impact of the National Healthy School Standard: using national datasets. *Health Education Research*, *20*(6), 688-696.

Sedgwick, P. (2013). Selection bias versus allocation bias. *BMJ*, *346*, f3345.

Shadish, D. T., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, USA: Mifflin and Company.

Shaffer, P. (2011). Against excessive rhetoric in impact assessment: overstating the case for randomised controlled experiments. *Journal of Development Studies*, *47*(11), 1619-1635.

Shapiro, L. (2010). *Embodied Cognition*. London, UK: Routledge.

Shuell, T. J. (1990). Phases of meaningful learning. *Review of Educational Research*, *60,* 531–547.

Sigman, A. (2012). Time for a view on screen time*. Archives of Disease in Childhood*, doi:10.1136/archdischild-2012-302196.

Singal, A. G., Higgins, P. D., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, *5*(1), e45. doi:10.1038/ctg.2013.13.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85-89.

Slavin, R. (2012). Foreward. In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (p. xv). Cambridge, UK: Cambridge University Press.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, *78*(3), 427-515.

Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, *4*(4), 370-380.

Snow, C. E., & Uccelli, P. (2009). The Challenge of Academic Language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge Handbook of Literacy* (pp. 112–133). Cambridge, UK: Cambridge University Press.

Snowling, M. J. (2013). Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs*, *13*(1), 7-14.

Spector, J. M. (2008). Theoretical foundations. In J. M. Spector, M. D. Merrill, J. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and*

Technology (pp. 21–28). New York, USA: Routledge Taylor & Francis.

Spoth, R., Guyll, M., Lillehoj, C. J., Redmond, C., & Greenberg, M. (2007). Prosper study of evidence-based intervention implementation quality by community–university partnerships. *Journal of Community Psychology*, *35*(8), 981-999.

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203-219.

Stacy, S. T., Cartwright, M., Arwood, Z., Canfield, J. P., & Kloos, H. (2017). Addressing the Math-Practice Gap in Elementary School: Are tablets a feasible tool for informal math practice? *Frontiers in Psychology*, *8*(179), doi:10.3389/fpsyg.2017.00179.

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, *6*(2), 103-114.

Sutton-Smith, B. (2001). *The Ambiguity of Play*. Massachusetts, USA: Harvard University Press.

Swanson, H. L., & Hoskyn, M. (1999). Definition x treatment interactions for students with learning disabilities. *School Psychology Review*, *28*(4), 644-659.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2010). *Early childhood matters: Evidence from the effective pre-school and primary education project*. London, UK: Routledge.

Szkudlarek, E., & Brannon, E. M. (2018). Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers. *Frontiers in Psychology*, *9*(606), 1-19.

Tamim, R. M., Borokhovski, E., Pickup, D., & Bernard, R. M. (2015). *Large-scale, government-supported educational tablet initiatives*. British Columbia, Canada: Commonwealth of Learning.

Tan, C. K., & Tan, C. P. (2015). Effects of the handheld technology instructional approach on performances of students of different achievement levels. *Computers & Education*, *82*, 306-314.

Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research*. London, UK: Sage.

Tauson, M., & Stannard, L. (2018). *Edtech for Learning in Emergencies and Displaced Settings. A Rigourous Review and Narrative Synthesis*. London, UK: Save the Children.

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In

defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37-43.

Thomas, G. (1992). Ecological Interventions. In S. Wolfendale, T. Bryans, M. Fox, A. Labram, A. Sigston (Eds.), *The Profession and Practice of Educational Psychology: Future Directions* (pp. 49-68). London, UK: Continuum International Publishing.

Thomas, G. (2016). After the gold rush: Questioning the "gold standard" and reappraising the status of experiment and randomized controlled trials in education. *Harvard Educational Review*, *86*(3), 390-411.

Thomas, G., & Pring. R. (2004). *Evidence-Based Practice in Education.* Maidenhead, UK: Open University Press.

Thomas, R. E., Crutcher, R., & Lorenzetti, D. (2007). A systematic review of the methodological quality and outcomes of RCTs to teach medical undergraduates surgical and emergency procedures. *Canadian Journal of Surgery*, *50*(4), 278-290.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review*, *31*(4), 498-514.

Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, *49*(3), 316-328.

Torgerson, D. J. (2001). Contamination in trials: is cluster randomisation the answer?. *BMJ*, *322*(7282), 355-357.

Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British Journal of Educational Studies*, *51*(1), 36-45.

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, *77*(3), 786-799.

Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. *Reading Research Quarterly*, *28,* 219-233.

UNESCO Global Monitoring Report. (2014). *Teaching and Learning: Achieving quality for all.* Retrieved from http://unesdoc.unesco.org/images/0022/002266/226662e.pdf

UNESCO. (2011). *State of Education in Brazil.* Paris, France: UNESCO.

United Nations Convention on the Rights of the Child. (1989). *United Nations Convention on the Rights of the Child.* Retrieved from https://downloads.unicef.org.uk/wp-content/uploads/2010/05/UNCRC_united_nations_convention_on_the_rights_of_the_child.pdf

United Nations. (2015). *Sustainable Development Goal 4.* Retrieved from https://sustainabledevelopment.un.org/sdg4

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104-132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038-1066.

Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of clinical epidemiology*, *59*(9), 920-925.

Van de Walle, J. A., Lovin, L. H., Karp, K. S., & Bay-Williams, J. M. (2014). *Teaching Student-Centered Mathematics: Developmentally Appropriate Instruction for Grades PreK-2 (volume 1).* Boston, USA: Pearson.

Van de Walle, J., Lovin, L. H., Karp, K. S., & Bay-Williams, J. M. (2014). *Teaching student-centered mathematics: developmentally appropriate instruction for grades pre-K-2.* Boston, USA: Pearson.

Van Der Heyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to

guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention*, *30*(3), 15-31.

van der Ven, F., Segers, E., Takashima, A., & Verhoeven, L. (2017). Effects of a tablet game intervention on simple addition and subtraction fluency in first graders. *Computers in Human Behavior*, *72*, 200-207.

Vatavu, R. D., Cramariuc, G., & Schipor, D. M. (2015). Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies*, *74*, 54-76.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2014). Finding the missing piece: Blocks, puzzles, and shapes fuel school readiness. *Trends in Neuroscience and Education*, *3*(1), 7-13.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, USA: Harvard University Press.

Walker, D., Mickes, L., Bajic, D., Nailon, C. R., & Rickard, T. C. (2013). A test of two methods of arithmetic fluency training and implications for educational practice. *Journal of Applied Research in Memory and Cognition*, *2*(1), 25-32.

Wang, F., Xie, H., Wang, Y., Hao, Y., & An, J. (2016). Using touchscreen tablets to help young children learn to tell

time. *Frontiers in Psychology*, *7*(1800), doi:10.3389/fpsyg.2016.01800.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352-360.

Wechsler, D. (2005). *Wechsler individual achievement test 2nd edition (WIAT II).* London: The Psychological Corp.

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, *25*(3), 177-182.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*(2), 575-587.

Wood, E., Gottardo, A., Grant, A., Evans, M. A., Phillips, L., & Savage, R. (2012). Developing tools for assessing and using commercially available reading software programs to promote the development of early reading skills in children. *NHSA Dialog*, *15*(4), 350-354.

Wood, E., Grant, A. K., Gottardo, A., Savage, R., & Evans, M. A. (2017). Software to Promote Young Children's Growth in

Literacy: A Comparison of Online and Offline Formats. *Early Childhood Education Journal*, *45*(2), 207-217.

Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015). *Improving Numeracy and Literacy Evaulation report and Executive Summary*. Berkshire, UK: National Foundation for Educational Research.

Wouters, P., Paas, F., & van Merriënboer, J. J. (2008). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research*, *78*(3), 645-675.

Yudkin, P. L., & Stratton, I. M. (1996). How to deal with regression to the mean in intervention studies. *Lancet*, *347*, 241-243.

Zimmerman, F. J., Christakis, D. A., & Meltzoff, A. N. (2007). Television and DVD/video viewing in children younger than 2 years. *Archives of Pediatrics & Adolescent Medicine*, *161*(5), 473-479.

Zimmermann, L., Moser, A., Lee, H., Gerhardstein, P., & Barr, R. (2017). The ghost in the touchscreen: Social scaffolds promote learning by toddlers. *Child Development*, *88*(6), 2013-2025.

Zosh, J. M., Lytle, S. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2017). Putting the education back in educational apps: How content and context interact to promote learning. In R. Barr & D.

Linebarger (Eds.), *Media Exposure During Infancy and Early Childhood* (pp. 259-282). New York, USA: Springer.

Appendix

Appendix A Search strategy code used in the Systematic Review reported in Chapter 2.

((((("young children" OR kindergarten OR preschool OR elementary OR primary)) AND (tablet OR "tablet technology" OR "educational apps" OR "interactive apps")) AND (mathematics OR maths OR math)) AND (achievement OR development OR learning)

Appendix B Descriptive data extracted from the seven included studies reported in the Systematic Review in Chapter 2.

Study	Design	Assessment Measure	Country	Age Range	Enrolled *N*	Final *N*	Attrition %
Schacter & Jo (2016)	Clustered quasi-experimental study with non-randomised groups	Researcher developed assessment of number sense ability. Tablet based delivery.	USA	4-5 years	227	162	71.4
Schacter & Jo (2017)	Cluster RCT with full randomisation	Researcher developed assessment of number sense ability. Tablet based delivery.	USA	4-5 years	433	378	87.3
Schacter et al. (2015)	Pupil RCT with full randomisation	Researcher developed assessment of number sense ability. Tablet based delivery.	USA	4-5 years	100	86	86.0

Study	Design	Assessment Measure	Country	Age Range	Enrolled *N*	Final *N*	Attrition %
Park et al. (2016)	Pupil RCT with pseudo-randomisation	TEMA-3	USA	3-5 years	109	103	94.5
Szkudlarek & Brannon (2018)	Pupil RCT with pseudo-randomisation	Modified number sense screener.	USA	3-4 years	158	157	99.4
Van der Van et al. (2017)	Cluster RCT with pseudo-randomisation	Researcher developed assessment of arithmetic skills.	Netherla-nds	6-7 years	103	102	99.0
Pitchford (2015)	Pupil-level RCT with full randomisation	Researcher developed assessment of curriculum knowledge and mathematical concepts.	Malawi	6-13 years	318	283	89.0

Appendix C Paper-and-pen assessment of Mathematical Curriculum Knowledge used in the UK Proof of Concept Study reported in Chapter 4.

Version A: Questions

The test is administered on a one to one basis with the researcher reading the question clearly to the child. The child responds either verbally (e.g. saying their answer aloud) or non-verbally (e.g. pointing to their answer), the researcher than records the child's response on the stimuli response sheet (see below). Some items (denoted by asterisk) require the child to draw/mark their answer on the stimuli response sheet.

Item	Question	Answer
1	Which is the odd one out?	Big triangle
2	Which is the matching pair?	2 owls
3	Show me 2 stars	Identifies 2 stars
4	How many frogs are there?	4
5	What comes next in the pattern?	Circle
6	Where are the triangle and the square?	Identifies triangle and square *1 point for both. No points if only one shape is identified*

Item	Question	Answer
7*	Draw 4 tomatoes on the plate	Draws 4 'tomatoes' *Award 1 point for 4 distinct drawings e.g. tomatoes, circles*
8	Which box has 5 strawberries?	Bottom right
9	Which child is in front of the fence?	Left child
10	Which pencil is on the left?	Left pencil
11	Which set of circles matches the number?	Left box
12	Which set of circles does not match the number?	Bottom domino
13	Which is the odd one out?	Long rectangle
14	What comes next in the pattern?	Circle
15	What are the missing numbers?	4 and 8 *1 point for both. No points for only one correct answer*
16	How many frogs are there?	8
17	Which number is 1 more than 4?	5
18	Which number is 1 less than 3?	2

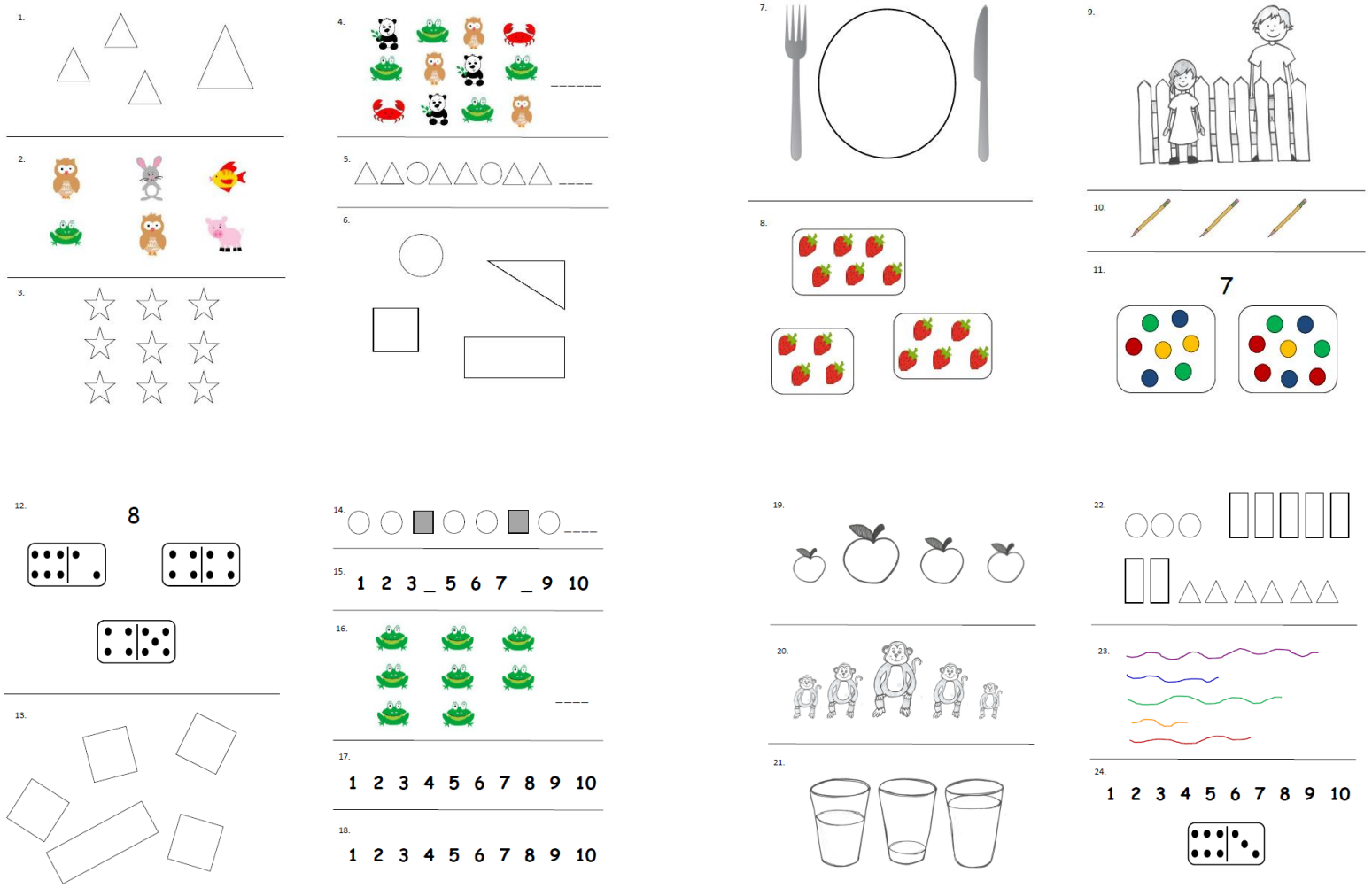
Item	Question	Answer
19	Which is the largest apple?	Second left
20	Which 2 monkeys match?	2 monkeys either side of the middle
21	Which cup has the most water?	Right cup
22	Which shape is there fewest of?	Circles
23	Which line is the longest?	First line
24	How many dots are on the domino?	9
25	There are 4 fish. I add 3 more *(researcher draws 3 fish)*. How many fish are there now?	7
26	There are 8 pandas. I take away 4 *(researcher crosses out 4 pandas)*. How many pandas are there now?	4
27	There are 5 cats. I add 2 more *(researcher draws 2 cats).* How many cats are there now?	7

Item	Question	Answer
28	There are 9 dogs. I take away 4 *(researcher crosses out 4 dogs).* How many dogs are there now?	5
29*	Draw a line of symmetry	Draws straight line down the middle of the item
30*	Split the circle into quarters	Divides circle into 4 equal parts
31	How many carrots are there?	18
32*	Draw 12 apples on the tree	Draws 12 'apples' *Award 1 point for 12 distinct drawings e.g. apples, circles*
33	Which is the number 16?	Identifies 16
34	Which number is 1 more than 17?	18
35	Which are the numbers 11, 15 and 20?	Identifies 11, 15, 20 *1 point for all. No points for only one/ some correct answer(s).*
36	There are 2 plates and 8 cakes. How many cakes go on each plate?	4

Item	Question	Answer
37	There are 2 children and 8 sweets. How many sweets can each child have?	4
38	How many fish are there in each set?	3
39	There are 6 cows. How many groups of 2 can I make?	3
40	There are 4 pigs. How many groups of 2 can I make?	2
41	Which is the 7th circle?	Identifies 7th circle
42	What are the missing numbers?	8 and 16 *1 point for both. No points for only one correct answer.*
43*	Circle 12 cherries	12 individual cherries or 6 pairs
44	Which child is 3rd line for the bus?	3rd child
45	Which numbers are even?	2, 8, 12, 16 *1 point for all. No points for only one correct answer.*
46	Which clock shows 6 o'clock?	Top left

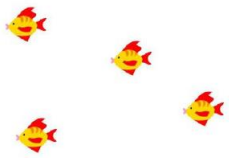
Item	Question	Answer
47	Which clock shows 3 o’clock?	Bottom
48	Which clock shows half past 2?	Right
49	Which clock shows a different time?	Left
50	Number the pictures in the right order	4, 3, 1, 2

Version A: Stimuli Response Sheet



25.

1 2 3 4 5 6 7 8 9 10



26.

1 2 3 4 5 6 7 8 9 10



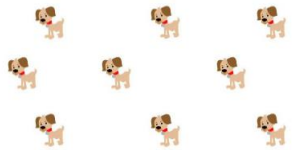
27.

1 2 3 4 5 6 7 8 9 10



28.

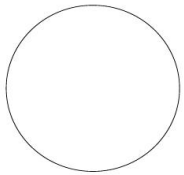
1 2 3 4 5 6 7 8 9 10



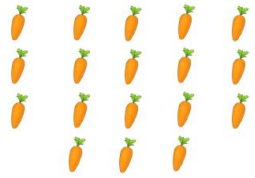
29.



30.



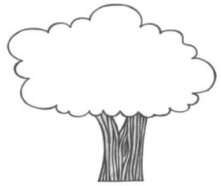
31.



1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

32.



33.

1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

34.

11 12 13 14 15 16 17 18 19 20

35.

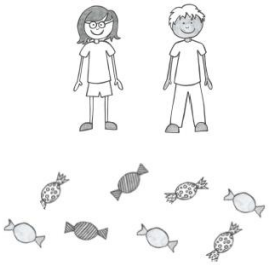
1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

36.

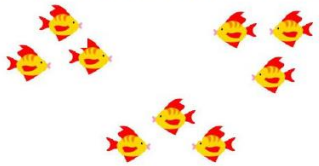


37.



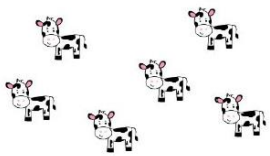
38.

1 2 3 4 5 6 7 8 9 10



39.

1 2 3 4 5

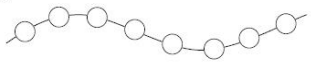


40.

1 2 3 4 5



41.

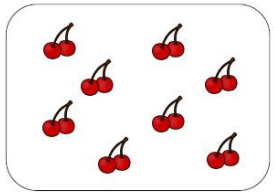


42.

2 4 6 _ 10 12 14 _ 18 20

43.

12



44.



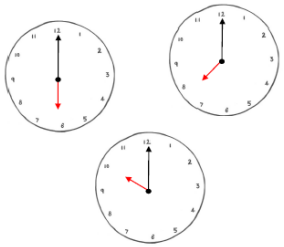
45.

15 11 16

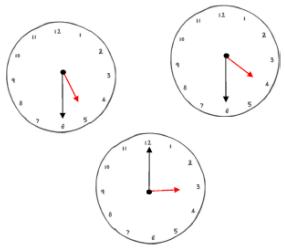
12 2

1 8 3

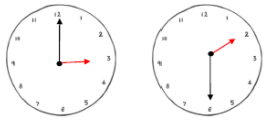
46.



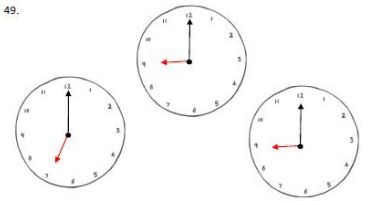
47.



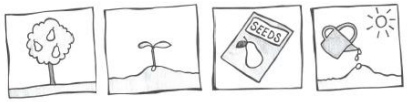
48.



49.



50.



Participant number:

Date:

Gender:

Score:

Version B: Questions

Same administration procedure used as described in Version A.

Item	Question	Answer
1	Which is the odd one out?	Big circle
2	Which is the matching pair?	2 pigs
3	Show me 3 stars	Identifies 3 stars
4	How many frogs are there?	5
5	What comes next in the pattern?	Square
6	Where are the circle and the rectangle?	Identifies circle and the rectangle *1 point for both. No points if only one shape is identified*
7*	Draw 5 tomatoes on the plate	Draws 5 'tomatoes' *Award 1 point for 4 distinct drawings e.g. tomatoes, circles*
8	Which box has 6 strawberries?	Bottom
9	Which child is in behind of the fence?	Right child
10	Which carrot is on the right?	Right carrot
11	Which set of circles matches the number?	Left box

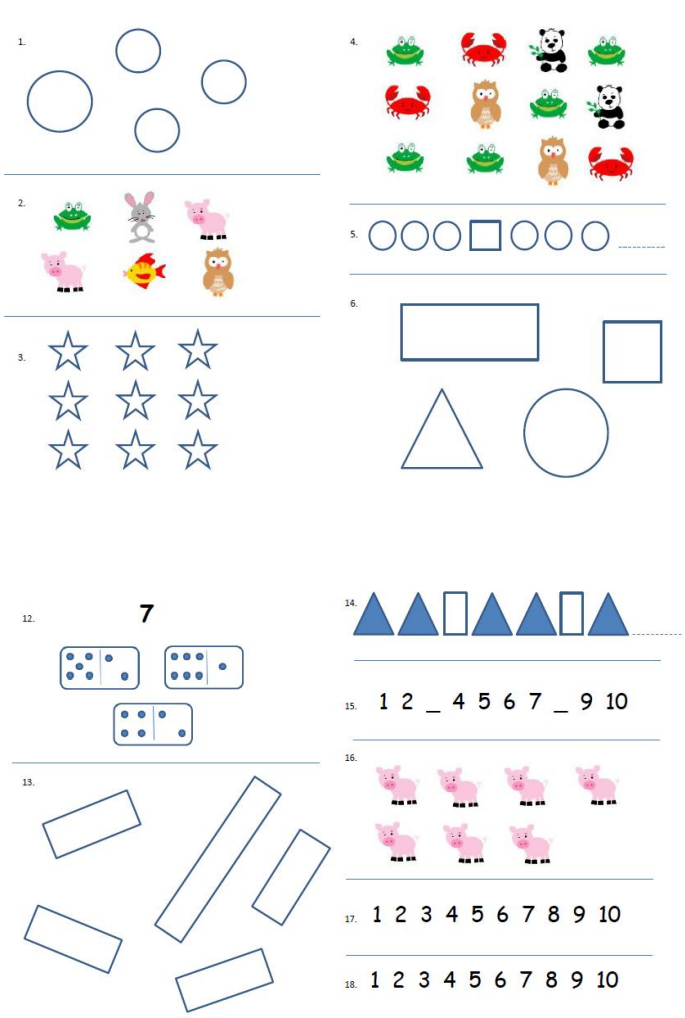
Item	Question	Answer
12	Which set of circles does not match the number?	Bottom domino
13	Which is the odd one out?	Long rectangle
14	What comes next in the pattern?	Triangle
15	What are the missing numbers?	3 and 8 *1 point for both. No points for only one correct answer*
16	How many pigs are there?	7
17	Which number is 1 more than 7?	8
18	Which number is 2 less than 5?	3
19	Which is the largest apple?	Left
20	Which 2 monkeys match?	Far left and far right
21	Which cup has the most water?	Middle cup
22	Which shape is there fewest of?	Circles
23	Which line is the longest?	Last line
24	How many dots are on the domino?	6

Item	Question	Answer
25	There are 3 frogs. I add 2 more *(researcher draws 2 frogs)*. How many frogs are there now?	5
26	There are 8 cats. I take away 3 *(researcher crosses out 3 cats)*. How many cats are there now?	5
27	There are 4 fish. I add 4 more *(researcher draws 4 fish).* How many fish are there now?	8
28	There are 10 owls. I take away 5 *(researcher crosses out 5 owls).* How many owls are there now?	5
29*	Draw a line of symmetry	Draws straight line down the middle of the item
30*	Split the circle in half	Divides circle into 2 equal parts
31	How many pencils are there?	15

Item	Question	Answer
32*	Draw 14 apples on the tree	Draws 14 'apples' *Award 1 point for 14 distinct drawings e.g. apples, circles*
33	Which is the number 18?	Identifies 18
34	Which number is 1 more than 19?	20
35	Which are the numbers 10, 16, and 19?	Identifies 10, 16, 19 *1 point for all. No points for only one/ some correct answer(s).*
36	There are 2 plates and 6 cakes. How many cakes go on each plate?	3
37	There are 2 children and 6 sweets. How many sweets can each child have?	3
38	How many frogs are there in each set?	4
39	There are 8 pandas. How many groups of 2 can I make?	4
40	There are 6 crabs. How many groups of 2	3

Item	Question	Answer
	can I make?	
41	Which is the 5th circle?	Identifies 5th circle
42	What are the missing numbers?	4 and 18
		1 point for both. No points for only one correct answer.
43*	Circle 10 cherries	10 individual cherries or 5 pairs
44	Which child is 4th line for the bus?	4th child
45	Which numbers are even?	6, 8, 12, 14
		1 point for all. No points for only one correct answer.
46	Which clock shows 4 o'clock?	Top left
47	Which clock shows 9 o'clock?	Top left
48	Which clock shows half past 4?	Left
49	Which clock shows the same time?	Bottom two
50	Number the pictures in the right order	2, 1, 3, 4

Version B: Stimuli Response Sheet



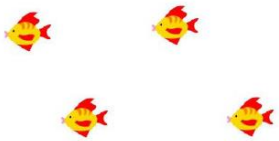
25. 1 2 3 4 5 6 7 8 9 10



26. 1 2 3 4 5 6 7 8 9 10



27. 1 2 3 4 5 6 7 8 9 10



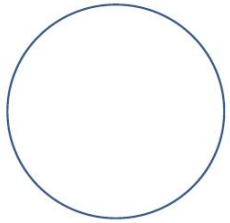
28. 1 2 3 4 5 6 7 8 9 10



29.



30.



31.



1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

32.



33.

1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

34.

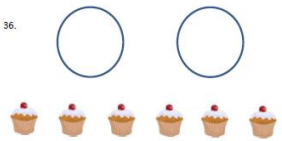
11 12 13 14 15 16 17 18 19 20

35.

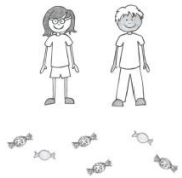
1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

36.



37.

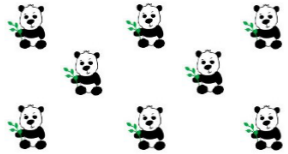


38. 1 2 3 4 5 6 7 8 9 10



39.

1 2 3 4 5 6

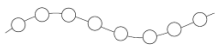


40.

1 2 3 4 5 6



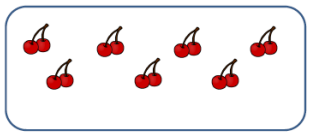
41.



42. 2 __ 6 8 10 12 14 16 __ 20

43.

10



44.



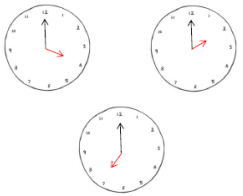
45.

3 6 12

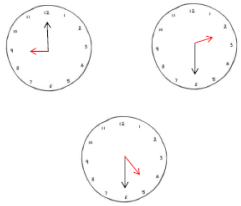
8 15

14 7 5

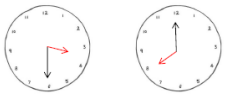
46.



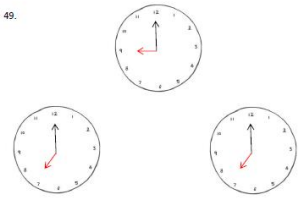
47.



48.



49.



Participant number:

Date:

Gender:

Score:

50.



Appendix D Assessment of child's language proficiency used in the Brazil Proof of Concept Study reported in Chapter 5.

The 7-item questionnaire was completed by the child's Brazilian Portuguese and English teachers. The questionnaire was translated by a native Brazilian Portuguese speaker for non-English speaking teachers.

Child's name:

Current Language Competency	**In Brazilian Portuguese**	**In English**
Compared with other children of the same age, how do you think this child expresses him/herself? 0 = not very well; 1 = a little less well; 2 = the same; 3 = very good/better/one of the best		
Compared to other children of the same age, how do you think this child pronounces words? 0 = not very well; 1 = a little less well; 2 = the same; 3 = very good/better/one of the best		

Current Language Competency	**In Brazilian Portuguese**	**In English**
Is it easy for you and other teachers to have a conversation with this child? 3 = very easy; 2 = easy enough; 1 = sometimes not easy; 0 = no, very hard		
Compared to other children of the same age, does this child have difficulty producing correct sentences? *e.g. have appropriate vocabulary, correct grammar, long enough sentences to get the idea across* 3 = no difficulties, maybe better; 2 = same; 1 = some difficulties; 0 = a lot of difficulties		
How does this child read compared with other children his/her age? 0 = noticeably worse than other children; 1 = not as well as other children, 2 = same as other children, 3 = very well, maybe better		

Current Language Competency	**In Brazilian Portuguese**	**In English**
How does this child write compared with other children his/her age? 0 = noticeably worse than other children; 1 = not as well as other children, 2 = same as other children, 3 = very well, maybe better		
Compared to other children of the same age, how do you think this child understands spoken instructions in each language? 0 = not very well; 1 = a little less well; 2 = the same; 3 = very good/better/one of the best		
Total	**/21**	**/21**

Appendix E Teacher Manual provided to all schools participating in the UK Feasibility RCT reported in Chapter 6.

Unlocking Talent UK Teacher Manual

XXXXXX Primary School





Contents

Useful contacts 388

Study timeline 389

Parents' consent 390

Tips for gaining parents' consent: 390

Children's study ID 390

Allocating children to group 391

Maths assessments 391

Installing the onebillion maths apps 393

Implementing the Unlocking Talent intervention 395

Daily routine 395

Individual iPad 396

General tips for implementation 397

What to do if a member of teaching staff is away 397

School visits 397

Observations and comments 398

Appendix 399

Useful contacts
If at any point during the study you are unsure what to do, please contact Laura and she will be able to help you.

Name	Role	Email
Laura Outhwaite	PhD student running the project	lpxlo4@nottingham.ac.uk
Marc Faulder	F2 teacher and Apple Distinguished Educator	marc.faulder@burtonjoyce.notts.sch.uk
Dr Nicola Pitchford	Associate professor supervising the project	nicola.pitchford@nottingham.ac.uk
Anthea Gulliford	Educational psychologist supervising the project	anthea.gulliford@nottingham.ac.uk
Professor Stephen Jackson	Head of University of Nottingham ethics committee	stephen.jackson@nottingham.ac.uk

Study timeline

Teachers send out parent information sheets and consent forms	From 01 March
Teachers fill in children's study ID table (included in the appendix)	By 15 March
Teachers email Laura number of participating children	By 15 March
Pre-test assessments by trained assessors	Weeks beginning 21 and 28 March Times to be individually arranged
Laura allocates participating children to group	By 01 April
Teachers install onebillion maths apps onto school iPads with help from Laura if needed	By 01 April
Intervention to begin	On 18 April
Visits from Laura	Between 18 April and 15 July Times to be individually arranged
Post-test assessments by trained assessors	Week beginning 18 July Times to be individually arranged
Study and maths reports	August 2016

Parents' consent
In order for children to take part in the study a parent or guardian needs to complete and sign the consent form.
Consent forms and the parent information letter should be sent home with a covering letter from the school showing your support for the project.

You will be emailed electronic copies of the consent form and parent information letter. An additional hard copy is included in the Appendix.

Tips for gaining parents' consent:

- Have spare consent forms for parents/guardians to sign when dropping off or picking up their children from school
- If you have an upcoming parents evening have consent forms available
- Send parents/guardians reminders via text message

Children's study ID
Children are to be given a unique study ID. This ensures confidentiality and ensures the researchers and assessors are not biased because they will not know which children are using the Unlocking Talent intervention.

As children's consent forms are returned please write their names in table 1 included in the Appendix.

Study ID	Child's name
1A	Joe Jones
2A	Amy Smith
3A	Katie Brown
4A	

Keep this list safe. You will need to do 2 things with the list:

1. Refer to this list when the trained assessors come to run the maths assessments before and after the intervention. This will allow the trained assessors to collect the right child for assessment and record the correct study ID without knowing the child's group allocation.
2. Make a duplicate copy and place in the sealed envelope included with this manual. Laura will collect the sealed envelope when visiting your school.

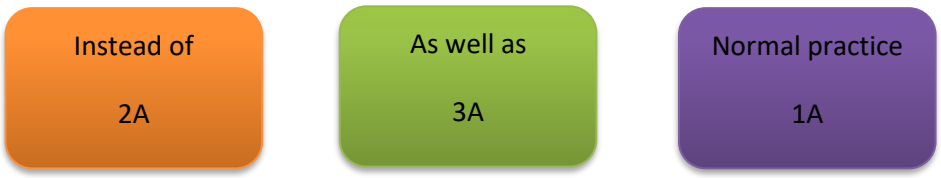
Allocating children to group
When consent forms have been collected and table 1 completed please email Laura the total number of participating children by **15 March**.

You do not need to give Laura the names of the participating children – just the number of children taking part in the study in your class.

Laura will randomly allocate children to one of the three groups on the basis of their study ID. She will know the study IDs from the number of children taking part in the study.

Group allocations will be sent back to you by 01 April, before the Easter holidays. You can then set up the iPads with the onebillion maths apps ready to start the intervention on 18 April when you return back to school.

It's vital that children remain in these groups. This is the information you will be sent back:



Using your children's study ID list you will be able to identify the names of children in each group, like so:



Maths assessments
All participating children will be assessed on the Progress Test in Maths, Level 5 (GL assessments) before and after the 12-week intervention.

Depending on the number of participating children in your school, assessments will take between 1-3 days. Assessment days will be organised through communication with Laura and the school's lead contact. The dates can be flexible to meet the needs of your school.

Pre-test assessment time frame	**Post-test assessment time frame**
Week beginning 21 March	Week beginning 18 July
Week beginning 28 March	

Trained assessors will carry out all of assessments on a one-to-one basis. Assessments need to be conducted in a quiet space, for example, a shared space area. It is expected each assessment should take between 30-40 minutes to administer.

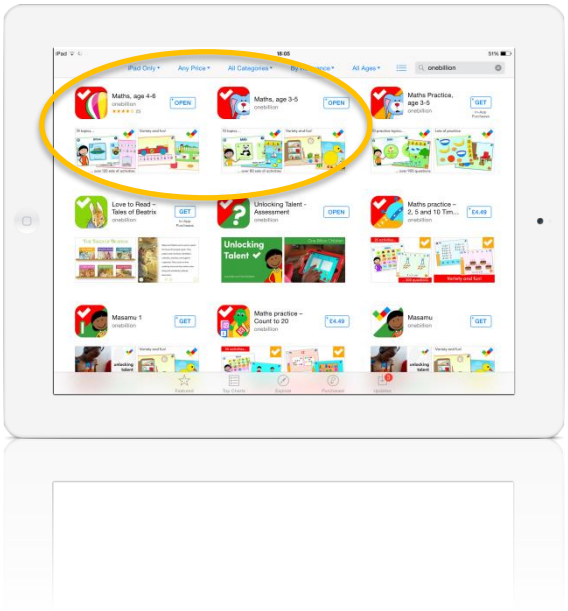
It is important the trained assessors don't know which group the child is in.

Trained assessors will be in groups of 3 in order to minimise disruption and assess children as quickly as possible. All of the assessors will have DBS clearance.

It would be helpful if you could set up a table with 3 chairs for the assessors and 3 chairs for the children in a quiet area on the days that your assessments will take place.

Installing the onebillion maths apps
The onebillion maths apps 'Maths 3-5' and 'Maths 4-6' need to be installed onto the school iPads.

Step 1: Download the 'Maths 3-5' and 'Maths 4-6' apps from the app store onto each of your school iPads.



Step 2: In each app the first topic will be unlocked. The rest will be locked.



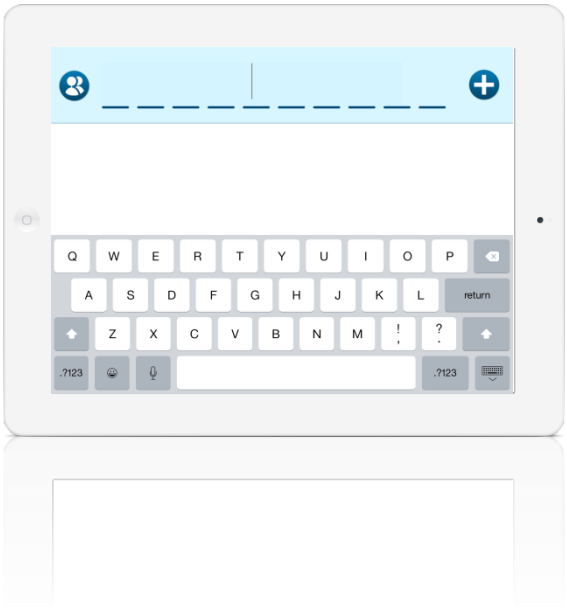


In order to access all topics in each app you need to use the unlock codes listed in the table below.

Your unlock codes are:

	Number of pupils	Number of iPads	Maths 3-5 code	Maths 4-6 code
XXXXXX Primary School	29	**10**	XXXXXX	XXXXXX

Step 3: Type the unlock code in the name section of the app. Repeat for both 'Maths 3-5' and 'Math 4-6' for each iPad.



Step 4: All of the 'Maths 3-5' and 'Maths 4-6' content is now unlocked.



If you have any problems downloading and accessing the maths apps, please email Laura and she will be able to provide assistance.

Implementing the Unlocking Talent intervention
The Unlocking Intervention will be implemented for 12 weeks starting the week beginning 18 April.

Daily routine
Children allocated to receive the Unlocking Talent intervention should use the maths apps for approximately 30 minutes a day. This can be build up over time to meet the needs of the individual child or school routine.

Individual schools will have different maths routines and so this section provides general implementation guidance for each group. Examples from Marc Faulder at Burton Joyce Primary School are provided. It is important that you are able to implement the intervention in a way that fits into your school routine. Please include any comments in the observations and comments section from page 12.

It is also vital that all children continue to receive maths input and maths target groups based on their next steps regardless of what group they are in.

'Instead of'

Children in the 'instead of' group should use the maths apps in place of one daily maths activity.

At Burton Joyce they have a daily maths problem-solving group which is separate to their maths planning, maths input and maths groups. It is based on a Nottingham City Maths Project scheme. Instead of attending this group, children access the onebillion software instead. This session ran at 11:30am, just before lunch.

Children should continue to take part in other maths activities and instruction throughout the week, such as, guided maths input and independent maths tasks.

'As well as'

Children in the 'as well as' group should use the maths apps in addition to their daily and weekly maths activities.

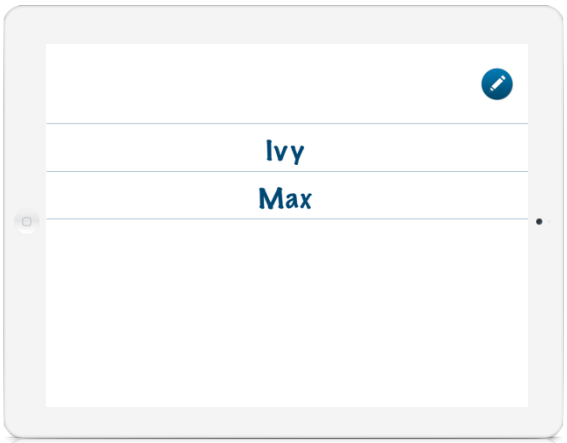
At Burton Joyce this group of children used the maths apps during the first session of the day.

'Normal practice'

Children in the 'normal practice' group should continue to have all maths activities standard to your school routine.

Individual iPad

Children allocated to receive the Unlocking Talent intervention should use the maths apps independently within their name on the same iPad each day (see picture below).



As shown in the picture you can have more than one child on each iPad. It would be beneficial to allocate children an iPad based on their group. For example, Ivy would be in the 'instead of' group and Max in the 'as well as' group. That way children will not need to use the iPad at the same time.

The apps are largely self-led and children should start each session from where they left off from the previous day, as shown in the picture below. Initially children may require supervision to help them get used to the routine of the intervention. However, you will notice most children will become more independent as the intervention progresses.



General tips for implementation
Classroom organisation is key. Here are some tips from Marc Faulder, F2 teacher and Apple Distinguished Educator, who has been using the Unlocking Talent intervention since June 2015.

Find regular times to run the intervention that fits in with your classroom and school routine.

At Burton Joyce we have found it is best to run the 'as well as' group at the very beginning of the school day and the 'instead of' group at the end of the morning session. These are the times we are able to access the school iPads on a daily basis. These are also good times to make use of the iPads as other classes are introducing learning or giving feedback before lunchtime.

Keep your own set of headphones and teach your children how to set them up.

For your iPad groups to run smoothly you need to have headphones and their cables arranged carefully; you won't want to be untangling them every day. Keep a set aside for yourself to use every day! Also, train your children to get them out and plug them in themselves. That will save a lot of time too!

Ensuring the 'instead of' group don't miss out on maths lessons.

The onebillion maths apps instead of the Nottingham City Maths Project scheme works well because the nature of the apps and the nature of the scheme are very similar. They are both repetitive, they are both instructional and they both focus on maths vocabulary. Children who are in our 'instead-of' group still access our maths input and maths target groups that we plan for based on their next steps.

What to do if a member of teaching staff is away
If you have a supply teacher ideally the intervention should continue running for that day. All the normal teaching staff will be familiar with the routine. Similarly, children will become more familiar and independent in the intervention routine as the weeks progress.

If you do need to miss a session for whatever reason, for example, a school trip, please make a note of it in the observation and comments section from page 12.

School visits
Over the 12 intervention weeks Laura will visit your school to see how you are getting on with the projects, hear your views about the intervention and will be able to provide any additional support.

These visits are flexible to meet the needs of your school. Laura will make arrangements with the school's lead contact.

Observations and comments
Please use this space to record any observations or comments you would to share with us as they happen. For example, but not limited to:

- If something worked particularly well
- How you overcame any challenges
- How you have found implementing the intervention in your daily practice
- How the children are interacting with the apps
- Any changes you have noticed in the children

Appendix
Children's study ID table

Keep this list safe. You will need to do 2 things with the list:

1. Refer to this list when the trained assessors come to run the maths assessments before and after the intervention. This will allow the trained assessors to collect the right child for assessment and record the correct study ID without knowing the child's group allocation.
2. Make a duplicate copy and place in the included sealed envelope. Laura will collect the sealed envelope when visiting your school.

Class 1

Study ID	Child's name
A1	
A2	
A3	
A4	
A5	
A6	
A7	
A8	
A9	
A10	
A11	
A12	
A13	
A14	
A15	
A16	
A17	

A18	
A19	
A20	
A21	
A22	
A23	
A24	
A25	
A26	
A27	
A28	
A29	
A30	

School of Psychology
University Park
University of Nottingham
Nottingham
NG7 2RD.

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

March 2016

Parent information sheet
Research project: Unlocking Talent UK

Researchers: Laura Outhwaite, Dr Nikki Pitchford & Anthea Gulliford
Contact details: lpxlo4@nottingham.ac.uk,
nicola.pitchford@nottingham.ac.uk, anthea.gulliford@nottingham.ac.uk

Dear parent / guardian,

Why I am being contacted?
This is an invitation for your child to take part in a study looking at the potential benefits of a new intervention, Unlocking Talent, which is designed to support the development of early maths skills. The Unlocking Talent intervention includes the maths iPad apps 'Maths 3-5' and 'Maths 4-6.'

The maths apps have been especially developed for Foundation year children by onebillion, a not-for-profit organisation. We would like to know the extent to which the maths apps can support learning compared to normal maths lessons that take place in Foundation year.

Your child's school has agreed to participate in the study. We are contacting you to ask your permission for your child to take part.

What will participation in the study involve?
Before you decide if you wish for your child to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

If your child participates, they will be randomly allocated to either receive the Unlocking Talent intervention as well as to their normal maths lessons, or to have their normal maths lessons only. It is important to note, all children will continue to have their normal maths lessons aligned with the national curriculum.

If you child is allocated to receive the Unlocking Talent intervention, they will use the maths apps individually on an iPad for approximately 30 minutes a day for 12 weeks. They will work independently in small groups under the supervision of teaching staff. Activities within the apps include, sorting, matching, counting and understanding patterns, which are aligned to the activities they are being taught in class.

All children that take part in the study will be given a short maths task in March, before the intervention start, and again in July, after the

intervention has finished. This task should take approximately 30 minutes.

Participation in this study is completely voluntary and your child is under no obligation to take part. You are free to withdraw your child at any point before or during the study. All data collected will be kept confidential and used for research purpose only. Results from the maths task will be released to your child's school, as these will be helpful to their teacher. Should we decide to publish the results of the study individual children will not be identified, as only group results will be published.

What should I do if I want my child to take part in the study?
If you would like your child to take part in this study please sign the study consent form and hand it in to your child's teacher. Only children with parental consent can take part in the study.

Who should I contact if I have any further questions?
If you have any questions or concerns please do not hesitate to contact us using the contact details above. If you have any concerns about the study, please contact: Prof Stephen Jackson (Chair of Ethics Committee): stephen.jackson@nottingham.ac.uk

We thank you for considering this request.

Laura Outhwaite	**Dr Nikki Pitchford**	**Anthea Gulliford**
Doctoral Researcher	*Associate Professor*	*Educational Psychologist*
University of Nottingham	*University of Nottingham*	*University of Nottingham*

PARENT CONSENT FORM

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Study: Unlocking Talent UK
Investigators: Laura Outhwaite, Dr Nikki Pitchford, Anthea Gulliford
School of Psychology, University of Nottingham
Email: lpxlo4@nottingham.ac.uk / nicola.pitchford@nottingham.ac.uk / anthea.gulliford@nottingham.ac.uk

		Please tick box
1.	I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.	☐
2.	I understand that my child's participation is voluntary and that I am free to withdraw the data at any time before it is processed, without giving a reason.	☐
3.	I agree to my child taking part in the study.	☐

_____	_____	_____
Name of Parent	Date	Signature

_____	_____	_____
Name of Child	Date of birth	Class teacher

PLEASE RETURN THIS FORM TO YOUR CHILD'S TEACHER AS SOON AS POSSIBLE.

Appendix F Teacher Feedback Questionnaire (see Chapter 8)

Unlocking Talent RCT Teacher Feedback Questionnaire

How confident do you feel using the onebillion maths apps in your classroom?	**1** Very confident	**2** Confident	**3** Neutral	**4** A little under confident	**5** Very under confident
Any comments?					

How useful do you find the following information and training received about the onebillion maths apps?					
Recruitment event at Burton Joyce Primary School	**1** Very useful	**2** Useful	**3** Neutral	**4** A little unhelpful	**5** Very unhelpful
Teacher manual	**1** Very useful	**2** Useful	**3** Neutral	**4** A little unhelpful	**5** Very unhelpful
Laura's school visits	**1** Very useful	**2** Useful	**3** Neutral	**4** A little unhelpful	**5** Very unhelpful
Any comments?					

Do you think the onebillion maths apps have benefitted the following:					
Children's maths skills	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Other areas of children's learning e.g. problem solving	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's use of maths language	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's independent learning	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's self confidence in maths	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's self confidence in other areas of the curriculum	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's enjoyment of maths	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's enjoyment of other areas of the curriculum	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit

Children's attention and engagement while using the maths apps	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Children's attention and engagement in other areas of the curriculum	**1** Very beneficial	**2** Beneficial	**3** Neutral	**4** Limited observed benefit	**5** No observed benefit
Any comments?					

Do you see any other advantages to the onebillion maths apps?

What disadvantages do you see with the onebillion maths apps?

Do you intent on using the onebillion maths apps next academic year?

How do you intend to use the apps in your classroom next year?

Would you recommend the onebillion maths apps?			
To other teachers	Yes	No	Unsure
Why?			
To parents	Yes	No	Unsure
Why?			

What level of experience do you have using iPads in your classroom?	**1** Very experienced	**2** Experienced	**3** Neutral	**4** A little unexperienced	**5** Very unexperienced
Any comments?					

How many years have you been a teacher?	
What is your teaching role?	
School name	

Would you like to take part in further research on this project or others we are conducting in the future?			
This project	Yes	No	Maybe
Other projects	Yes	No	Maybe
Are there any particular areas or age groups you would like exploring?			

Do you have any further comments or feedback about the project?

Thank you for taking part in the Unlocking Talent RCT. We really value your feedback and perspective about the project and the onebillion maths apps.

Appendix G Two year follow-up evaluation Teacher Feedback Questionnaire (see Chapter 8)

Unlocking Talent UK Study Feedback

It has been almost 2 years since we finished the Unlocking Talent UK study in summer 2016. As you all continued to receive access to the onebillion maths apps, we are interested to hear whether you are still using the apps and if so, how you are getting on with them. Please could you fill in this brief Google questionnaire to let us know your thoughts.

Have you continued to use the *onebillion* maths apps after taking part in the Unlocking Talent UK study in 2016? *Tick one answer*	Yes	No
If yes, how are you using the onebillion maths apps at your school now? *Check all that apply* ○ With the whole class ○ As a targeted intervention with children in need of extra support learning maths ○ Other: _____		
If yes, what age group do you use the apps with?		
If yes, how many teachers or teaching assistants use the onebillion maths apps?		

If yes, how many children have used the apps since the study in 2016?

If yes, how useful did you find the following information received about the outcomes of the Unlocking Talent UK study? *Mark one answer for each*

Results event at Burton Joyce Primary School (October 2016)	**1** Very unhelpful	**2**	**3**	**4**	**5** Very useful
Individual school results (received Autumn 2016)	**1** Very unhelpful	**2**	**3**	**4**	**5** Very useful
Journal article (received July 2018)	**1** Very unhelpful	**2**	**3**	**4**	**5** Very useful

Any further comments?

What advantages, if any, have you seen from using the onebillion maths apps in your school?

What disadvantages, if any, have you seen from using the onebillion maths apps in your school?

Would you like to take part in further research on this project or others we are conducting in the future? *Mark one answer for each*			
Research similar to this project	Yes	No	Maybe
Other projects	Yes	No	Maybe
Are there any particular areas or age groups you would like exploring?			

Do you have any further comments or feedback about the project?